



**EUROPEAN COMMISSION**  
**DIRECTORATE-GENERAL RESEARCH & INNOVATION**  
Open Innovation and Open Science  
**Research infrastructure**



## GRANT AGREEMENT

**NUMBER — 731015 — ELEXIS**

This **Agreement** ('the Agreement') is **between** the following parties:

**on the one part,**

the **European Union** ('the EU'), represented by the European Commission ('the Commission'), represented for the purposes of signature of this Agreement by **DIRECTORATE-GENERAL RESEARCH & INNOVATION**, Open Innovation and Open Science, Administration and finance, Pascale CID,

**and**

**on the other part,**

1. 'the coordinator'<sup>3</sup> :

**INSTITUT JOZEF STEFAN (JSI)**, established in Jamova 39, LJUBLJANA 1000, Slovenia, VAT number: SI55560822, represented for the purposes of signing the Agreement by Director, Jadran LENARČIČ

and the following other beneficiaries, if they sign their 'Accession Form' (see Annex 3 and Article 56):

2. **LEXICAL COMPUTING CZ SRO (LC)**, established in BOTANICKA 554/68A, BRNO 602 00, Czech Republic, VAT number: CZ29295491,

3. **STICHTING INSTITUUT VOOR DE NEDERLANDSE TAAL (IVDNT)**, established in MATTHIAS DE VRIESHOFF 2-3, LEIDEN 2311 BZ, Netherlands, VAT number: NL002818267B01,

4. **UNIVERSITA DEGLI STUDI DI ROMA LA SAPIENZA (UNIROMA1)**, established in Piazzale Aldo Moro 5, ROMA 00185, Italy, VAT number: IT02133771002,

5. **NATIONAL UNIVERSITY OF IRELAND GALWAY (NUI GALWAY)**, established in UNIVERSITY ROAD, GALWAY, Ireland, VAT number: IE0022578J,

6. **OESTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN (OEAW)**, established in DR. IGNAZ SEIPEL-PLATZ 2, WIEN 1010, Austria, VAT number: ATU37116303,

7. **CENTAR ZA DIGITALNE HUMANISTICKE NAUKE (BCDH)**, established in DZORDZA VASINGTONA 28A, BEOGRAD 11000, Serbia, VAT number: RS105604641,

---

<sup>3</sup> The coordinator shall be the ITD/IADP/TA technical coordinator.

8. **MAGYAR TUDOMANYOS AKADEMIA, NYELVTUDOMANYI INTEZET (MTANYTI)**, established in BENCZUR UTCA 33, BUDAPEST 1068, Hungary, VAT number: HU15300571,

9. **INSTITUTE FOR BULGARIAN LANGUAGE PROF LYUBOMIR ANDREYCHIN (IBL)**, established in UL. SHIPCHENSKI PROHOD 52 BL 17, Sofia 1113, Bulgaria, VAT number: BG000665498,

10. **UNIVERSIDADE NOVA DE LISBOA (FCSH-UNL)**, established in CAMPUS DE CAMPOLIDE, LISBOA 1099 085, Portugal, VAT number: PT501559094,

11. **K DICTIONARIES LTD (K Dictionaries)**, established in 8 NAHUM HANAVI STREET, TEL AVIV 6350310, Israel, VAT number: IL511783672,

12. **CONSIGLIO NAZIONALE DELLE RICERCHE (CNR)**, established in PIAZZALE ALDO MORO 7, ROMA 00185, Italy, VAT number: IT02118311006,

13. **DET DANSKE SPROG- OG LITTERATURSELSKAB (DSL)**, established in CHRISTIANS BRYGGE 1, KOBENHAVN 1219, Denmark, VAT number: DK59188917,

14. **KOBENHAVNS UNIVERSITET (UCPH)**, established in NORREGADE 10, KOBENHAVN 1165, Denmark, VAT number: DK29979812,

15. **UNIVERSITAT TRIER (UT)**, established in UNIVERSITATSRING 15, TRIER 54296, Germany, VAT number: DE149881695,

16. **EESTI KEELE INSTITUUT (EKI)**, established in ROOSIKRANTSI 6, TALLINN 10119, Estonia, VAT number: EE100252305,

17. **REAL ACADEMIA ESPANOLA (RAE)**, established in CALLE FELIPE IV 4, MADRID 28014, Spain, VAT number: ESQ2868010F,

Unless otherwise specified, references to ‘beneficiary’ or ‘beneficiaries’ include the coordinator.

The parties referred to above have agreed to enter into the Agreement under the terms and conditions below.

By signing the Agreement or the Accession Form, the beneficiaries accept the grant and agree to implement it under their own responsibility and in accordance with the Agreement, with all the obligations and conditions it sets out.

The Agreement is composed of:

Terms and Conditions

- Annex 1 Description of the action
- Annex 2 Estimated budget for the action
  - 2a Additional information on the estimated budget
- Annex 3 Accession Forms
- Annex 4 Model for the financial statements
- Annex 5 Model for the certificate on the financial statements (CFS)
- Annex 6 Model for the certificate on the methodology

## TERMS AND CONDITIONS

### TABLE OF CONTENTS

<b>CHAPTER 1 GENERAL .....</b>	<b>12</b>
ARTICLE 1 — SUBJECT OF THE AGREEMENT.....	12
<b>CHAPTER 2 ACTION .....</b>	<b>12</b>
ARTICLE 2 — ACTION TO BE IMPLEMENTED .....	12
ARTICLE 3 — DURATION AND STARTING DATE OF THE ACTION.....	12
ARTICLE 4 — ESTIMATED BUDGET AND BUDGET TRANSFERS.....	12
4.1 Estimated budget.....	12
4.2 Budget transfers.....	12
<b>CHAPTER 3 GRANT .....</b>	<b>12</b>
ARTICLE 5 — GRANT AMOUNT, FORM OF GRANT, REIMBURSEMENT RATES AND FORMS OF COSTS.....	12
5.1 Maximum grant amount.....	12
5.2 Form of grant, reimbursement rates and forms of costs.....	12
5.3 Final grant amount — Calculation.....	13
5.4 Revised final grant amount — Calculation.....	15
ARTICLE 6 — ELIGIBLE AND INELIGIBLE COSTS.....	15
6.1 General conditions for costs to be eligible.....	15
6.2 Specific conditions for costs to be eligible.....	16
6.3 Conditions for costs of linked third parties to be eligible.....	22
6.4 Conditions for in-kind contributions provided by third parties free of charge to be eligible.....	22
6.5 Ineligible costs.....	22
6.6 Consequences of declaration of ineligible costs.....	23
<b>CHAPTER 4 RIGHTS AND OBLIGATIONS OF THE PARTIES .....</b>	<b>23</b>
<b>SECTION 1 RIGHTS AND OBLIGATIONS RELATED TO IMPLEMENTING THE ACTION .....</b>	<b>23</b>
ARTICLE 7 — GENERAL OBLIGATION TO PROPERLY IMPLEMENT THE ACTION.....	23
7.1 General obligation to properly implement the action.....	23
7.2 Consequences of non-compliance.....	23
ARTICLE 8 — RESOURCES TO IMPLEMENT THE ACTION — THIRD PARTIES INVOLVED IN THE ACTION.....	23
ARTICLE 9 — IMPLEMENTATION OF ACTION TASKS BY BENEFICIARIES NOT RECEIVING EU FUNDING.....	24
ARTICLE 10 — PURCHASE OF GOODS, WORKS OR SERVICES.....	24

10.1 Rules for purchasing goods, works or services.....	24
10.2 Consequences of non-compliance.....	24
ARTICLE 11 — USE OF IN-KIND CONTRIBUTIONS PROVIDED BY THIRD PARTIES AGAINST PAYMENT.....	24
11.1 Rules for the use of in-kind contributions against payment.....	24
11.2 Consequences of non-compliance.....	25
ARTICLE 12 — USE OF IN-KIND CONTRIBUTIONS PROVIDED BY THIRD PARTIES FREE OF CHARGE.....	25
12.1 Rules for the use of in-kind contributions free of charge.....	25
12.2 Consequences of non-compliance.....	25
ARTICLE 13 — IMPLEMENTATION OF ACTION TASKS BY SUBCONTRACTORS.....	26
13.1 Rules for subcontracting action tasks.....	26
13.2 Consequences of non-compliance.....	26
ARTICLE 14 — IMPLEMENTATION OF ACTION TASKS BY LINKED THIRD PARTIES.....	26
14.1 Rules for calling upon linked third parties to implement part of the action.....	26
14.2 Consequences of non-compliance.....	27
ARTICLE 15 — FINANCIAL SUPPORT TO THIRD PARTIES.....	27
15.1 Rules for providing financial support to third parties.....	27
15.2 Financial support in the form of prizes.....	27
15.3 Consequences of non-compliance.....	27
ARTICLE 16 — PROVISION OF TRANS-NATIONAL OR VIRTUAL ACCESS TO RESEARCH INFRASTRUCTURE.....	28
16.1 Rules for providing trans-national access to research infrastructure.....	28
16.2 Rules for providing virtual access to research infrastructure.....	29
16.3 Consequences of non-compliance.....	29
<b>SECTION 2 RIGHTS AND OBLIGATIONS RELATED TO THE GRANT ADMINISTRATION .....</b>	<b>30</b>
ARTICLE 17 — GENERAL OBLIGATION TO INFORM.....	30
17.1 General obligation to provide information upon request.....	30
17.2 Obligation to keep information up to date and to inform about events and circumstances likely to affect the Agreement.....	30
17.3 Consequences of non-compliance.....	30
ARTICLE 18 — KEEPING RECORDS — SUPPORTING DOCUMENTATION.....	30
18.1 Obligation to keep records and other supporting documentation.....	31
18.2 Consequences of non-compliance.....	32
ARTICLE 19 — SUBMISSION OF DELIVERABLES.....	32
19.1 Obligation to submit deliverables.....	32

19.2	Consequences of non-compliance.....	32
ARTICLE 20 — REPORTING — PAYMENT REQUESTS.....		32
20.1	Obligation to submit reports.....	32
20.2	Reporting periods.....	32
20.3	Periodic reports — Requests for interim payments.....	33
20.4	Final report — Request for payment of the balance.....	34
20.5	Information on cumulative expenditure incurred.....	35
20.6	Currency for financial statements and conversion into euro.....	35
20.7	Language of reports.....	35
20.8	Consequences of non-compliance.....	35
ARTICLE 21 — PAYMENTS AND PAYMENT ARRANGEMENTS.....		35
21.1	Payments to be made.....	35
21.2	Pre-financing payment — Amount — Amount retained for the Guarantee Fund.....	36
21.3	Interim payments — Amount — Calculation.....	36
21.4	Payment of the balance — Amount — Calculation — Release of the amount retained for the Guarantee Fund.....	37
21.5	Notification of amounts due.....	37
21.6	Currency for payments.....	38
21.7	Payments to the coordinator — Distribution to the beneficiaries.....	38
21.8	Bank account for payments.....	38
21.9	Costs of payment transfers.....	38
21.10	Date of payment.....	38
21.11	Consequences of non-compliance.....	38
ARTICLE 22 — CHECKS, REVIEWS, AUDITS AND INVESTIGATIONS — EXTENSION OF FINDINGS.....		39
22.1	Checks, reviews and audits by the Commission.....	39
22.2	Investigations by the European Anti-Fraud Office (OLAF).....	41
22.3	Checks and audits by the European Court of Auditors (ECA).....	41
22.4	Checks, reviews, audits and investigations for international organisations.....	41
22.5	Consequences of findings in checks, reviews, audits and investigations — Extension of findings.....	42
22.6	Consequences of non-compliance.....	43
ARTICLE 23 — EVALUATION OF THE IMPACT OF THE ACTION.....		43
23.1	Right to evaluate the impact of the action.....	43
23.2	Consequences of non-compliance.....	44
<b>SECTION 3 RIGHTS AND OBLIGATIONS RELATED TO BACKGROUND AND RESULTS .....</b>		<b>44</b>

<b>SUBSECTION 1 GENERAL .....</b>	<b>44</b>
ARTICLE 23a — MANAGEMENT OF INTELLECTUAL PROPERTY.....	44
23a.1 Obligation to take measures to implement the Commission Recommendation on the management of intellectual property in knowledge transfer activities.....	44
23a.2 Consequences of non-compliance.....	44
<b>SUBSECTION 2 RIGHTS AND OBLIGATIONS RELATED TO BACKGROUND .....</b>	<b>44</b>
ARTICLE 24 — AGREEMENT ON BACKGROUND.....	44
24.1 Agreement on background.....	45
24.2 Consequences of non-compliance.....	45
ARTICLE 25 — ACCESS RIGHTS TO BACKGROUND.....	45
25.1 Exercise of access rights — Waiving of access rights — No sub-licensing.....	45
25.2 Access rights for other beneficiaries, for implementing their own tasks under the action.....	45
25.3 Access rights for other beneficiaries, for exploiting their own results.....	45
25.4 Access rights for affiliated entities.....	46
25.5 Access rights for third parties.....	46
25.6 Consequences of non-compliance.....	46
<b>SUBSECTION 3 RIGHTS AND OBLIGATIONS RELATED TO RESULTS .....</b>	<b>46</b>
ARTICLE 26 — OWNERSHIP OF RESULTS.....	46
26.1 Ownership by the beneficiary that generates the results.....	46
26.2 Joint ownership by several beneficiaries.....	47
26.3 Rights of third parties (including personnel).....	47
26.4 EU ownership, to protect results.....	47
26.5 Consequences of non-compliance.....	48
ARTICLE 27 — PROTECTION OF RESULTS — VISIBILITY OF EU FUNDING.....	48
27.1 Obligation to protect the results.....	48
27.2 EU ownership, to protect the results.....	49
27.3 Information on EU funding.....	49
27.4 Consequences of non-compliance.....	49
ARTICLE 28 — EXPLOITATION OF RESULTS.....	49
28.1 Obligation to exploit the results.....	49
28.2 Results that could contribute to European or international standards — Information on EU funding.....	49
28.3 Consequences of non-compliance.....	50
ARTICLE 29 — DISSEMINATION OF RESULTS — OPEN ACCESS — VISIBILITY OF EU FUNDING.....	50
29.1 Obligation to disseminate results.....	50

29.2	Open access to scientific publications.....	50
29.3	Open access to research data.....	51
29.4	Information on EU funding — Obligation and right to use the EU emblem.....	51
29.5	Disclaimer excluding Commission responsibility.....	52
29.6	Consequences of non-compliance.....	52
<b>ARTICLE 30 — TRANSFER AND LICENSING OF RESULTS.....</b>		<b>52</b>
30.1	Transfer of ownership.....	52
30.2	Granting licenses.....	52
30.3	Commission right to object to transfers or licensing.....	53
30.4	Consequences of non-compliance.....	53
<b>ARTICLE 31 — ACCESS RIGHTS TO RESULTS.....</b>		<b>53</b>
31.1	Exercise of access rights — Waiving of access rights — No sub-licensing.....	53
31.2	Access rights for other beneficiaries, for implementing their own tasks under the action.....	53
31.3	Access rights for other beneficiaries, for exploiting their own results.....	53
31.4	Access rights of affiliated entities.....	53
31.5	Access rights for the EU institutions, bodies, offices or agencies and EU Member States.....	54
31.6	Access rights for third parties.....	54
31.7	Consequences of non-compliance.....	54
<b>SECTION 4 OTHER RIGHTS AND OBLIGATIONS .....</b>		<b>54</b>
<b>ARTICLE 32 — RECRUITMENT AND WORKING CONDITIONS FOR RESEARCHERS.....</b>		<b>54</b>
32.1	Obligation to take measures to implement the European Charter for Researchers and Code of Conduct for the Recruitment of Researchers.....	54
32.2	Consequences of non-compliance.....	54
<b>ARTICLE 33 — GENDER EQUALITY.....</b>		<b>55</b>
33.1	Obligation to aim for gender equality.....	55
33.2	Consequences of non-compliance.....	55
<b>ARTICLE 34 — ETHICS AND RESEARCH INTEGRITY.....</b>		<b>55</b>
34.1	Obligation to comply with ethical and research integrity principles.....	55
34.2	Activities raising ethical issues.....	56
34.3	Activities involving human embryos or human embryonic stem cells.....	57
34.4	Consequences of non-compliance.....	57
<b>ARTICLE 35 — CONFLICT OF INTERESTS.....</b>		<b>57</b>
35.1	Obligation to avoid a conflict of interests.....	57
35.2	Consequences of non-compliance.....	57
<b>ARTICLE 36 — CONFIDENTIALITY.....</b>		<b>57</b>



36.1	General obligation to maintain confidentiality.....	57
36.2	Consequences of non-compliance.....	58
<b>ARTICLE 37 — SECURITY-RELATED OBLIGATIONS.....</b>		<b>59</b>
37.1	Results with a security recommendation.....	59
37.2	Classified information.....	59
37.3	Activities involving dual-use goods or dangerous materials and substances.....	59
37.4	Consequences of non-compliance.....	59
<b>ARTICLE 38 — PROMOTING THE ACTION — VISIBILITY OF EU FUNDING.....</b>		<b>59</b>
38.1	Communication activities by beneficiaries.....	59
38.2	Communication activities by the Commission.....	60
38.3	Consequences of non-compliance.....	61
<b>ARTICLE 39 — PROCESSING OF PERSONAL DATA.....</b>		<b>61</b>
39.1	Processing of personal data by the Commission.....	61
39.2	Processing of personal data by the beneficiaries.....	62
39.3	Consequences of non-compliance.....	62
<b>ARTICLE 40 — ASSIGNMENTS OF CLAIMS FOR PAYMENT AGAINST THE COMMISSION.....</b>		<b>62</b>
<b>CHAPTER 5 DIVISION OF BENEFICIARIES’ ROLES AND RESPONSIBILITIES — RELATIONSHIP WITH COMPLEMENTARY BENEFICIARIES — RELATIONSHIP WITH PARTNERS OF A JOINT ACTION .....</b>		<b>62</b>
<b>ARTICLE 41 — DIVISION OF BENEFICIARIES’ ROLES AND RESPONSIBILITIES — RELATIONSHIP WITH COMPLEMENTARY BENEFICIARIES — RELATIONSHIP WITH PARTNERS OF A JOINT ACTION.....</b>		<b>62</b>
41.1	Roles and responsibility towards the Commission.....	62
41.2	Internal division of roles and responsibilities.....	63
41.3	Internal arrangements between beneficiaries — Consortium agreement.....	63
41.4	Relationship with complementary beneficiaries — Collaboration agreement.....	64
41.5	Relationship with partners of a joint action — Coordination agreement.....	64
<b>CHAPTER 6 REJECTION OF COSTS — REDUCTION OF THE GRANT — RECOVERY — SANCTIONS — DAMAGES — SUSPENSION — TERMINATION — FORCE MAJEURE .....</b>		<b>64</b>
<b>SECTION 1 REJECTION OF COSTS — REDUCTION OF THE GRANT — RECOVERY — SANCTIONS .....</b>		<b>64</b>
<b>ARTICLE 42 — REJECTION OF INELIGIBLE COSTS.....</b>		<b>64</b>
42.1	Conditions.....	64
42.2	Ineligible costs to be rejected — Calculation — Procedure.....	64
42.3	Effects.....	65
<b>ARTICLE 43 — REDUCTION OF THE GRANT.....</b>		<b>65</b>
43.1	Conditions.....	65

43.2 Amount to be reduced — Calculation — Procedure.....	66
43.3 Effects.....	66
ARTICLE 44 — RECOVERY OF UNDUE AMOUNTS.....	66
44.1 Amount to be recovered — Calculation — Procedure.....	66
ARTICLE 45 — ADMINISTRATIVE SANCTIONS.....	70
<b>SECTION 2 LIABILITY FOR DAMAGES .....</b>	<b>70</b>
ARTICLE 46 — LIABILITY FOR DAMAGES.....	70
46.1 Liability of the Commission.....	70
46.2 Liability of the beneficiaries.....	70
<b>SECTION 3 SUSPENSION AND TERMINATION .....</b>	<b>71</b>
ARTICLE 47 — SUSPENSION OF PAYMENT DEADLINE.....	71
47.1 Conditions.....	71
47.2 Procedure.....	71
ARTICLE 48 — SUSPENSION OF PAYMENTS.....	71
48.1 Conditions.....	71
48.2 Procedure.....	72
ARTICLE 49 — SUSPENSION OF THE ACTION IMPLEMENTATION.....	72
49.1 Suspension of the action implementation, by the beneficiaries.....	72
49.2 Suspension of the action implementation, by the Commission.....	73
ARTICLE 50 — TERMINATION OF THE AGREEMENT OR OF THE PARTICIPATION OF ONE OR MORE BENEFICIARIES.....	74
50.1 Termination of the Agreement, by the beneficiaries.....	74
50.2 Termination of the participation of one or more beneficiaries, by the beneficiaries.....	75
50.3 Termination of the Agreement or the participation of one or more beneficiaries, by the Commission.....	77
<b>SECTION 4 FORCE MAJEURE .....</b>	<b>81</b>
ARTICLE 51 — FORCE MAJEURE.....	81
<b>CHAPTER 7 FINAL PROVISIONS .....</b>	<b>82</b>
ARTICLE 52 — COMMUNICATION BETWEEN THE PARTIES.....	82
52.1 Form and means of communication.....	82
52.2 Date of communication.....	83
52.3 Addresses for communication.....	83
ARTICLE 53 — INTERPRETATION OF THE AGREEMENT.....	83
53.1 Precedence of the Terms and Conditions over the Annexes.....	83
53.2 Privileges and immunities.....	83
ARTICLE 54 — CALCULATION OF PERIODS, DATES AND DEADLINES.....	84

ARTICLE 55 — AMENDMENTS TO THE AGREEMENT.....	84
55.1 Conditions.....	84
55.2 Procedure.....	84
ARTICLE 56 — ACCESSION TO THE AGREEMENT.....	85
56.1 Accession of the beneficiaries mentioned in the Preamble.....	85
56.2 Addition of new beneficiaries.....	85
ARTICLE 57 — APPLICABLE LAW AND SETTLEMENT OF DISPUTES.....	85
57.1 Applicable law.....	85
57.2 Dispute settlement.....	85
ARTICLE 58 — ENTRY INTO FORCE OF THE AGREEMENT.....	86

## **CHAPTER 1 GENERAL**

### **ARTICLE 1 — SUBJECT OF THE AGREEMENT**

This Agreement sets out the rights and obligations and the terms and conditions applicable to the grant awarded to the beneficiaries for implementing the action set out in Chapter 2.

## **CHAPTER 2 ACTION**

### **ARTICLE 2 — ACTION TO BE IMPLEMENTED**

The grant is awarded for the action entitled ‘**European Lexicographic Infrastructure — ELEXIS**’ (**‘action’**), as described in Annex 1.

### **ARTICLE 3 — DURATION AND STARTING DATE OF THE ACTION**

The duration of the action will be **48 months** as of 1 February 2018 (**‘starting date of the action’**).

### **ARTICLE 4 — ESTIMATED BUDGET AND BUDGET TRANSFERS**

#### **4.1 Estimated budget**

The **‘estimated budget’** for the action is set out in Annex 2.

It contains the estimated eligible costs and the forms of costs, broken down by beneficiary (and linked third party) and budget category (see Articles 5, 6, and 14).

#### **4.2 Budget transfers**

The estimated budget breakdown indicated in Annex 2 may be adjusted — without an amendment (see Article 55) — by transfers of amounts between beneficiaries, budget categories and/or forms of costs set out in Annex 2, if the action is implemented as described in Annex 1.

However, the beneficiaries may not add costs relating to subcontracts not provided for in Annex 1, unless such additional subcontracts are approved by an amendment or in accordance with Article 13.

## **CHAPTER 3 GRANT**

### **ARTICLE 5 — GRANT AMOUNT, FORM OF GRANT, REIMBURSEMENT RATES AND FORMS OF COSTS**

#### **5.1 Maximum grant amount**

The **‘maximum grant amount’** is **EUR 4,999,967.50** (four million nine hundred and ninety nine thousand nine hundred and sixty seven EURO and fifty eurocents).

#### **5.2 Form of grant, reimbursement rates and forms of costs**

The grant reimburses **100% of the action's eligible costs** (see Article 6) (**'reimbursement of eligible costs grant'**) (see Annex 2).

The estimated eligible costs of the action are EUR **4,999,967.50** (four million nine hundred and ninety nine thousand nine hundred and sixty seven EURO and fifty eurocents).

Eligible costs (see Article 6) must be declared under the following forms (**'forms of costs'**):

(a) for **direct personnel costs**:

- as actually incurred costs (**'actual costs'**) or
- on the basis of an amount per unit calculated by the beneficiary in accordance with its usual cost accounting practices (**'unit costs'**).

Personnel **costs for SME owners or beneficiaries that are natural persons** not receiving a salary (see Article 6.2, Points A.4 and A.5) must be declared on the basis of the amount per unit set out in Annex 2a (**unit costs**);

(b) for **direct costs of subcontracting**: as actually incurred costs (**actual costs**);

(c) for **direct costs of providing financial support to third parties**: not applicable;

(d) for **other direct costs**: as actually incurred costs (**actual costs**);

(e) for **indirect costs**: on the basis of a flat-rate applied as set out in Article 6.2, Point E (**'flat-rate costs'**);

(f) **specific cost category(ies)**: not applicable.

### 5.3 Final grant amount — Calculation

The **'final grant amount'** depends on the actual extent to which the action is implemented in accordance with the Agreement's terms and conditions.

This amount is calculated by the Commission — when the payment of the balance is made (see Article 21.4) — in the following steps:

Step 1 – Application of the reimbursement rates to the eligible costs

Step 2 – Limit to the maximum grant amount

Step 3 – Reduction due to the no-profit rule

Step 4 – Reduction due to substantial errors, irregularities or fraud or serious breach of obligations

#### 5.3.1 Step 1 — Application of the reimbursement rates to the eligible costs

The reimbursement rate(s) (see Article 5.2) are applied to the eligible costs (actual costs, unit costs and flat-rate costs; see Article 6) declared by the beneficiaries and linked third parties (see Article 20) and approved by the Commission (see Article 21).

#### 5.3.2 Step 2 — Limit to the maximum grant amount

If the amount obtained following Step 1 is higher than the maximum grant amount set out in Article 5.1, it will be limited to the latter.

### 5.3.3 Step 3 — Reduction due to the no-profit rule

The grant must not produce a profit.

‘**Profit**’ means the surplus of the amount obtained following Steps 1 and 2 plus the action’s total receipts, over the action’s total eligible costs.

The ‘**action’s total eligible costs**’ are the consolidated total eligible costs approved by the Commission.

The ‘**action’s total receipts**’ are the consolidated total receipts generated during its duration (see Article 3).

The following are considered **receipts**:

- (a) income generated by the action; if the income is generated from selling equipment or other assets purchased under the Agreement, the receipt is up to the amount declared as eligible under the Agreement;
- (b) financial contributions given by third parties to the beneficiary or to a linked third party specifically to be used for the action, and
- (c) in-kind contributions provided by third parties free of charge and specifically to be used for the action, if they have been declared as eligible costs.

The following are however not considered receipts:

- (a) income generated by exploiting the action’s results (see Article 28);
- (b) financial contributions by third parties, if they may be used to cover costs other than the eligible costs (see Article 6);
- (c) financial contributions by third parties with no obligation to repay any amount unused at the end of the period set out in Article 3.

If there is a profit, it will be deducted from the amount obtained following Steps 1 and 2.

### 5.3.4 Step 4 — Reduction due to substantial errors, irregularities or fraud or serious breach of obligations — Reduced grant amount — Calculation

If the grant is reduced (see Article 43), the Commission will calculate the reduced grant amount by deducting the amount of the reduction (calculated in proportion to the seriousness of the errors, irregularities or fraud or breach of obligations, in accordance with Article 43.2) from the maximum grant amount set out in Article 5.1.

The final grant amount will be the lower of the following two:

- the amount obtained following Steps 1 to 3 or
- the reduced grant amount following Step 4.

## 5.4 Revised final grant amount — Calculation

If — after the payment of the balance (in particular, after checks, reviews, audits or investigations; see Article 22) — the Commission rejects costs (see Article 42) or reduces the grant (see Article 43), it will calculate the ‘**revised final grant amount**’ for the beneficiary concerned by the findings.

This amount is calculated by the Commission on the basis of the findings, as follows:

- in case of **rejection of costs**: by applying the reimbursement rate to the revised eligible costs approved by the Commission for the beneficiary concerned;
- in case of **reduction of the grant**: by calculating the concerned beneficiary’s share in the grant amount reduced in proportion to the seriousness of the errors, irregularities or fraud or breach of obligations (see Article 43.2).

In case of **rejection of costs and reduction of the grant**, the revised final grant amount for the beneficiary concerned will be the lower of the two amounts above.

## ARTICLE 6 — ELIGIBLE AND INELIGIBLE COSTS

### 6.1 General conditions for costs to be eligible

‘**Eligible costs**’ are costs that meet the following criteria:

#### (a) for **actual costs**:

- (i) they must be actually incurred by the beneficiary;
- (ii) they must be incurred in the period set out in Article 3, with the exception of costs relating to the submission of the periodic report for the last reporting period and the final report (see Article 20);
- (iii) they must be indicated in the estimated budget set out in Annex 2;
- (iv) they must be incurred in connection with the action as described in Annex 1 and necessary for its implementation;
- (v) they must be identifiable and verifiable, in particular recorded in the beneficiary’s accounts in accordance with the accounting standards applicable in the country where the beneficiary is established and with the beneficiary’s usual cost accounting practices;
- (vi) they must comply with the applicable national law on taxes, labour and social security, and
- (vii) they must be reasonable, justified and must comply with the principle of sound financial management, in particular regarding economy and efficiency;

#### (b) for **unit costs**:

- (i) they must be calculated as follows:

{amounts per unit set out in Annex 2a or calculated by the beneficiary in accordance with its usual cost accounting practices (see Article 6.2, Point A)

multiplied by

the number of actual units};

(ii) the number of actual units must comply with the following conditions:

- the units must be actually used or produced in the period set out in Article 3;
- the units must be necessary for implementing the action or produced by it, and
- the number of units must be identifiable and verifiable, in particular supported by records and documentation (see Article 18);

(c) for **flat-rate costs**:

- (i) they must be calculated by applying the flat-rate set out in Annex 2, and
- (ii) the costs (actual costs or unit costs) to which the flat-rate is applied must comply with the conditions for eligibility set out in this Article.

## 6.2 Specific conditions for costs to be eligible

Costs are eligible if they comply with the general conditions (see above) and the specific conditions set out below for each of the following budget categories:

- A. direct personnel costs;
- B. direct costs of subcontracting;
- C. not applicable;
- D. other direct costs;
- E. indirect costs;
- F. not applicable.

‘Direct costs’ are costs that are directly linked to the action implementation and can therefore be attributed to it directly. They must not include any indirect costs (see Point E below).

‘Indirect costs’ are costs that are not directly linked to the action implementation and therefore cannot be attributed directly to it.

### A. Direct personnel costs

#### Types of eligible personnel costs

A.1 Personnel costs are eligible, if they are related to personnel working for the beneficiary under an employment contract (or equivalent appointing act) and assigned to the action (**‘costs for employees (or equivalent)’**). They must be limited to salaries (including during parental leave), social security contributions, taxes and other costs included in the **remuneration**, if they arise from national law or the employment contract (or equivalent appointing act).

Beneficiaries that are non-profit legal entities<sup>1</sup> may also declare as personnel costs **additional remuneration** for personnel assigned to the action (including payments on the basis of supplementary contracts regardless of their nature), if:

- (a) it is part of the beneficiary’s usual remuneration practices and is paid in a consistent manner whenever the same kind of work or expertise is required;



- (b) the criteria used to calculate the supplementary payments are objective and generally applied by the beneficiary, regardless of the source of funding used.

Additional remuneration for personnel assigned to the action is eligible up to the following amount:

- (a) if the person works full time and exclusively on the action during the full year: up to EUR 8 000;
- (b) if the person works exclusively on the action but not full-time or not for the full year: up to the corresponding pro-rata amount of EUR 8 000, or
- (c) if the person does not work exclusively on the action: up to a pro-rata amount calculated as follows:
- {{EUR 8 000  
divided by  
the number of annual productive hours (see below)},  
multiplied by  
the number of hours that the person has worked on the action during the year}.

A.2 The **costs for natural persons working under a direct contract** with the beneficiary other than an employment contract are eligible personnel costs, if:

- (a) the person works under the beneficiary's instructions and, unless otherwise agreed with the beneficiary, on the beneficiary's premises;
- (b) the result of the work carried out belongs to the beneficiary, and
- (c) the costs are not significantly different from those for personnel performing similar tasks under an employment contract with the beneficiary.

A.3 The **costs of personnel seconded by a third party against payment** are eligible personnel costs, if the conditions in Article 11.1 are met.

A.4 **Costs of owners** of beneficiaries that are small and medium-sized enterprises ('**SME owners**') who are working on the action and who do not receive a salary are eligible personnel costs, if they correspond to the amount per unit set out in Annex 2a multiplied by the number of actual hours worked on the action.

A.5 **Costs of 'beneficiaries that are natural persons'** not receiving a salary are eligible personnel costs, if they correspond to the amount per unit set out in Annex 2a multiplied by the number of actual hours worked on the action.

A.6 **Personnel costs for providing trans-national access to research infrastructure** are eligible only if also the conditions set out in Article 16.1.1 are met. **Personnel costs for providing virtual**

---

<sup>1</sup> For the definition, see Article 2.1(14) of the Rules for Participation Regulation No 1290/2013: '**non-profit legal entity**' means a legal entity which by its legal form is non-profit-making or which has a legal or statutory obligation not to distribute profits to its shareholders or individual members.

**access to research infrastructure** are eligible only if also the conditions set out in Article 16.2 are met.

### **Calculation**

Personnel costs must be calculated by the beneficiaries as follows:

{hourly rate  
multiplied by  
the number of actual hours worked on the action},  
plus  
for non-profit legal entities: additional remuneration to personnel assigned to the action under the conditions set out above (Point A.1)}.

The number of actual hours declared for a person must be identifiable and verifiable (see Article 18).

The total number of hours declared in EU or Euratom grants, for a person for a year, cannot be higher than the annual productive hours used for the calculations of the hourly rate. Therefore, the maximum number of hours that can be declared for the grant is:

{the number of annual productive hours for the year (see below)  
minus  
total number of hours declared by the beneficiary for that person in that year for other EU or Euratom grants}.

The ‘**hourly rate**’ is one of the following:

- (a) for personnel costs declared as **actual costs**: the hourly rate is calculated *per full financial year*, as follows:

{actual annual personnel costs (excluding additional remuneration) for the person  
divided by  
number of annual productive hours}.

using the personnel costs and the number of productive hours for each full financial year covered by the reporting period concerned. If a financial year is not closed at the end of the reporting period, the beneficiaries must use the hourly rate of the last closed financial year available.

For the ‘number of annual productive hours’, the beneficiaries may choose one of the following:

- (i) ‘fixed number of hours’: 1 720 hours for persons working full time (or corresponding pro-rata for persons not working full time);
- (ii) ‘individual annual productive hours’: the total number of hours worked by the person in the year for the beneficiary, calculated as follows:

{annual workable hours of the person (according to the employment contract, applicable collective labour agreement or national law)  
plus

overtime worked  
 minus  
 absences (such as sick leave and special leave)}.

‘Annual workable hours’ means the period during which the personnel must be working, at the employer’s disposal and carrying out his/her activity or duties under the employment contract, applicable collective labour agreement or national working time legislation.

If the contract (or applicable collective labour agreement or national working time legislation) does not allow to determine the annual workable hours, this option cannot be used;

- (iii) ‘standard annual productive hours’: the ‘standard number of annual hours’ generally applied by the beneficiary for its personnel in accordance with its usual cost accounting practices. This number must be at least 90% of the ‘standard annual workable hours’.

If there is no applicable reference for the standard annual workable hours, this option cannot be used.

For all options, the actual time spent on **parental leave** by a person assigned to the action may be deducted from the number of annual productive hours.

As an alternative, beneficiaries may calculate the hourly rate *per month*, as follows:

{actual monthly personnel cost (excluding additional remuneration) for the person  
 divided by  
 {number of annual productive hours / 12}}

using the personnel costs for each month and (one twelfth of) the annual productive hours calculated according to either option (i) or (iii) above, i.e.:

- fixed number of hours or
- standard annual productive hours.

Time spent on **parental leave** may not be deducted when calculating the hourly rate per month. However, beneficiaries may declare personnel costs incurred in periods of parental leave in proportion to the time the person worked on the action in that financial year.

If parts of a basic remuneration are generated over a period longer than a month, the beneficiaries may include only the share which is generated in the month (irrespective of the amount actually paid for that month).

Each beneficiary must use only one option (per full financial year or per month) for each full financial year;

- (b) for personnel costs declared on the basis of **unit costs**: the hourly rate is one of the following:

- (i) for SME owners or beneficiaries that are natural persons: the hourly rate set out in Annex 2a (see Points A.4 and A.5 above), or
- (ii) for personnel costs declared on the basis of the beneficiary's usual cost accounting practices: the hourly rate calculated by the beneficiary in accordance with its usual cost accounting practices, if:
  - the cost accounting practices used are applied in a consistent manner, based on objective criteria, regardless of the source of funding;
  - the hourly rate is calculated using the actual personnel costs recorded in the beneficiary's accounts, excluding any ineligible cost or costs included in other budget categories.

The actual personnel costs may be adjusted by the beneficiary on the basis of budgeted or estimated elements. Those elements must be relevant for calculating the personnel costs, reasonable and correspond to objective and verifiable information;

and

- the hourly rate is calculated using the number of annual productive hours (see above).

**B. Direct costs of subcontracting** (including related duties, taxes and charges such as non-deductible value added tax (VAT) paid by the beneficiary) are eligible if the conditions in Article 13.1.1 are met.

Subcontracting costs **for providing trans-national access to research infrastructure** are eligible only if also the conditions set out in Article 16.1.1 are met.

Subcontracting costs **for providing virtual access to research infrastructure** are eligible only if also the conditions set out in Article 16.2 are met.

### **C. Direct costs of providing financial support to third parties**

Not applicable

### **D. Other direct costs**

**D.1 Travel costs and related subsistence allowances** (including related duties, taxes and charges such as non-deductible value added tax (VAT) paid by the beneficiary) are eligible if they are in line with the beneficiary's usual practices on travel.

Travel costs **for providing trans-national access to research infrastructure** are eligible only if also the conditions set out in Article 16.1.1 are met.

**D.2 The depreciation costs of equipment, infrastructure or other assets** (new or second-hand) as recorded in the beneficiary's accounts are eligible, if they were purchased in accordance with Article 10.1.1 and written off in accordance with international accounting standards and the beneficiary's usual accounting practices.

The **costs of renting or leasing** equipment, infrastructure or other assets (including related duties,

taxes and charges such as non-deductible value added tax (VAT) paid by the beneficiary) are also eligible, if they do not exceed the depreciation costs of similar equipment, infrastructure or assets and do not include any financing fees.

The costs of equipment, infrastructure or other assets **contributed in-kind against payment** are eligible, if they do not exceed the depreciation costs of similar equipment, infrastructure or assets, do not include any financing fees and if the conditions in Article 11.1 are met.

The only portion of the costs that will be taken into account is that which corresponds to the duration of the action and rate of actual use for the purposes of the action.

As an exception, the beneficiaries must not declare such costs (i.e. costs of renting, leasing, purchasing depreciable equipment, infrastructure and other assets) for providing trans-national access to research infrastructure (see Article 16.1).

As an exception, the beneficiaries must not declare such costs (i.e. costs of renting, leasing, purchasing depreciable equipment, infrastructure and other assets) for providing virtual access to research infrastructure (see Article 16.2).

**D.3 Costs of other goods and services** (including related duties, taxes and charges such as non-deductible value added tax (VAT) paid by the beneficiary) are eligible, if they are:

- (a) purchased specifically for the action and in accordance with Article 10.1.1 or
- (b) contributed in kind against payment and in accordance with Article 11.1.

Such goods and services include, for instance, consumables and supplies, dissemination (including open access), protection of results, certificates on the financial statements (if they are required by the Agreement), certificates on the methodology, translations and publications.

Costs of other goods and services **for providing trans-national access to research infrastructure** are eligible only if also the conditions set out in Article 16.1.1 are met.

Costs of other goods and services **for providing virtual access to research infrastructure** are eligible only if also the conditions set out in Article 16.2 are met.

**D.4 Capitalised and operating costs of ‘large research infrastructure’<sup>2</sup>:** Not applicable

## **E. Indirect costs**

**Indirect costs** are eligible if they are declared on the basis of the flat-rate of 25% of the eligible direct costs (see Article 5.2 and Points A to D above), from which are excluded:

- (a) costs of subcontracting and
- (b) costs of in-kind contributions provided by third parties which are not used on the beneficiary’s premises;

---

<sup>2</sup> ‘**Large research infrastructure**’ means research infrastructure of a total value of at least EUR 20 million, for a beneficiary, calculated as the sum of historical asset values of each individual research infrastructure of that beneficiary, as they appear in its last closed balance sheet before the date of the signature of the Agreement or as determined on the basis of the rental and leasing costs of the research infrastructure.

(c) not applicable;

(d) not applicable.

Beneficiaries receiving an operating grant<sup>4</sup> financed by the EU or Euratom budget cannot declare indirect costs for the period covered by the operating grant.

## F. Specific cost category(ies)

Not applicable

### 6.3 Conditions for costs of linked third parties to be eligible

**Costs incurred by linked third parties** are eligible if they fulfil — *mutatis mutandis* — the general and specific conditions for eligibility set out in this Article (Article 6.1 and 6.2) and Article 14.1.1.

### 6.4 Conditions for in-kind contributions provided by third parties free of charge to be eligible

**In-kind contributions provided free of charge** are eligible direct costs (for the beneficiary or linked third party), if the costs incurred by the third party fulfil — *mutatis mutandis* — the general and specific conditions for eligibility set out in this Article (Article 6.1 and 6.2) and Article 12.1.

### 6.5 Ineligible costs

‘**Ineligible costs**’ are:

- (a) costs that do not comply with the conditions set out above (Article 6.1 to 6.4), in particular:
  - (i) costs related to return on capital;
  - (ii) debt and debt service charges;
  - (iii) provisions for future losses or debts;
  - (iv) interest owed;
  - (v) doubtful debts;
  - (vi) currency exchange losses;
  - (vii) bank costs charged by the beneficiary’s bank for transfers from the Commission;
  - (viii) excessive or reckless expenditure;
  - (ix) deductible VAT;
  - (x) costs incurred during suspension of the implementation of the action (see Article 49);

---

<sup>4</sup> For the definition, see Article 121(1)(b) of Regulation (EU, Euratom) No 966/2012 of the European Parliament and of the Council of 25 October 2012 on the financial rules applicable to the general budget of the Union and repealing Council Regulation (EC, Euratom) No 1605/2002 (‘**Financial Regulation No 966/2012**’)(OJ L 218, 26.10.2012, p.1): ‘**operating grant**’ means direct financial contribution, by way of donation, from the budget in order to finance the functioning of a body which pursues an aim of general EU interest or has an objective forming part of and supporting an EU policy.

- (b) costs declared under another EU or Euratom grant (including grants awarded by a Member State and financed by the EU or Euratom budget and grants awarded by bodies other than the Commission for the purpose of implementing the EU or Euratom budget); in particular, indirect costs if the beneficiary is already receiving an operating grant financed by the EU or Euratom budget in the same period.

## **6.6 Consequences of declaration of ineligible costs**

Declared costs that are ineligible will be rejected (see Article 42).

This may also lead to any of the other measures described in Chapter 6.

## **CHAPTER 4 RIGHTS AND OBLIGATIONS OF THE PARTIES**

### **SECTION 1 RIGHTS AND OBLIGATIONS RELATED TO IMPLEMENTING THE ACTION**

#### **ARTICLE 7 — GENERAL OBLIGATION TO PROPERLY IMPLEMENT THE ACTION**

##### **7.1 General obligation to properly implement the action**

The beneficiaries must implement the action as described in Annex 1 and in compliance with the provisions of the Agreement and all legal obligations under applicable EU, international and national law.

##### **7.2 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

#### **ARTICLE 8 — RESOURCES TO IMPLEMENT THE ACTION — THIRD PARTIES INVOLVED IN THE ACTION**

The beneficiaries must have the appropriate resources to implement the action.

If it is necessary to implement the action, the beneficiaries may:

- purchase goods, works and services (see Article 10);
- use in-kind contributions provided by third parties against payment (see Article 11);
- use in-kind contributions provided by third parties free of charge (see Article 12);
- call upon subcontractors to implement action tasks described in Annex 1 (see Article 13);
- call upon linked third parties to implement action tasks described in Annex 1 (see Article 14).

In these cases, the beneficiaries retain sole responsibility towards the Commission and the other beneficiaries for implementing the action.

## **ARTICLE 9 — IMPLEMENTATION OF ACTION TASKS BY BENEFICIARIES NOT RECEIVING EU FUNDING**

Not applicable

## **ARTICLE 10 — PURCHASE OF GOODS, WORKS OR SERVICES**

### **10.1 Rules for purchasing goods, works or services**

10.1.1 If necessary to implement the action, the beneficiaries may purchase goods, works or services.

The beneficiaries must make such purchases ensuring the best value for money or, if appropriate, the lowest price. In doing so, they must avoid any conflict of interests (see Article 35).

The beneficiaries must ensure that the Commission, the European Court of Auditors (ECA) and the European Anti-Fraud Office (OLAF) can exercise their rights under Articles 22 and 23 also towards their contractors.

10.1.2 Beneficiaries that are ‘contracting authorities’ within the meaning of Directive 2004/18/EC<sup>5</sup> (or 2014/24/EU<sup>6</sup>) or ‘contracting entities’ within the meaning of Directive 2004/17/EC<sup>7</sup> (or 2014/25/EU<sup>8</sup>) must comply with the applicable national law on public procurement.

### **10.2 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under Article 10.1.1, the costs related to the contract concerned will be ineligible (see Article 6) and will be rejected (see Article 42).

If a beneficiary breaches any of its obligations under Article 10.1.2, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

## **ARTICLE 11 — USE OF IN-KIND CONTRIBUTIONS PROVIDED BY THIRD PARTIES AGAINST PAYMENT**

### **11.1 Rules for the use of in-kind contributions against payment**

---

<sup>5</sup> Directive 2004/18/EC of the European Parliament and of the Council of 31 March 2004 on the coordination of procedures for the award of public work contracts, public supply contracts and public service contracts (OJ L 134, 30.04.2004, p. 114).

<sup>6</sup> Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement and repealing Directive 2004/18/EC. (OJ L 94, 28.03.2014, p. 65).

<sup>7</sup> Directive 2004/17/EC of the European Parliament and of the Council of 31 March 2004 coordinating the procurement procedures of entities operating in the water, energy, transport and postal services sectors (OJ L 134, 30.04.2004, p. 1)

<sup>8</sup> Directive 2014/25/EU of the European Parliament and of the Council of 26 February 2014 on procurement by entities operating in the water, energy, transport and postal services sectors and repealing Directive 2004/17/EC (OJ L 94, 28.03.2014, p. 243).



If necessary to implement the action, the beneficiaries may use in-kind contributions provided by third parties against payment.

The beneficiaries may declare costs related to the payment of in-kind contributions as eligible (see Article 6.1 and 6.2), up to the third parties' costs for the seconded persons, contributed equipment, infrastructure or other assets or other contributed goods and services.

The third parties and their contributions must be set out in Annex 1. The Commission may however approve in-kind contributions not set out in Annex 1 without amendment (see Article 55), if:

- they are specifically justified in the periodic technical report and
- their use does not entail changes to the Agreement which would call into question the decision awarding the grant or breach the principle of equal treatment of applicants.

The beneficiaries must ensure that the Commission, the European Court of Auditors (ECA) and the European Anti-Fraud Office (OLAF) can exercise their rights under Articles 22 and 23 also towards the third parties.

## **11.2 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under this Article, the costs related to the payment of the in-kind contribution will be ineligible (see Article 6) and will be rejected (see Article 42).

Such breaches may also lead to any of the other measures described in Chapter 6.

## **ARTICLE 12 — USE OF IN-KIND CONTRIBUTIONS PROVIDED BY THIRD PARTIES FREE OF CHARGE**

### **12.1 Rules for the use of in-kind contributions free of charge**

If necessary to implement the action, the beneficiaries may use in-kind contributions provided by third parties free of charge.

The beneficiaries may declare costs incurred by the third parties for the seconded persons, contributed equipment, infrastructure or other assets or other contributed goods and services as eligible in accordance with Article 6.4.

The third parties and their contributions must be set out in Annex 1. The Commission may however approve in-kind contributions not set out in Annex 1 without amendment (see Article 55), if:

- they are specifically justified in the periodic technical report and
- their use does not entail changes to the Agreement which would call into question the decision awarding the grant or breach the principle of equal treatment of applicants.

The beneficiaries must ensure that the Commission, the European Court of Auditors (ECA) and the European Anti-Fraud Office (OLAF) can exercise their rights under Articles 22 and 23 also towards the third parties.

### **12.2 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under this Article, the costs incurred by the third parties related to the in-kind contribution will be ineligible (see Article 6) and will be rejected (see Article 42).

Such breaches may also lead to any of the other measures described in Chapter 6.

## **ARTICLE 13 — IMPLEMENTATION OF ACTION TASKS BY SUBCONTRACTORS**

### **13.1 Rules for subcontracting action tasks**

13.1.1 If necessary to implement the action, the beneficiaries may award subcontracts covering the implementation of certain action tasks described in Annex 1.

Subcontracting may cover only a limited part of the action.

The beneficiaries must award the subcontracts ensuring the best value for money or, if appropriate, the lowest price. In doing so, they must avoid any conflict of interests (see Article 35).

The tasks to be implemented and the estimated cost for each subcontract must be set out in Annex 1 and the total estimated costs of subcontracting per beneficiary must be set out in Annex 2. The Commission may however approve subcontracts not set out in Annex 1 and 2 without amendment (see Article 55), if:

- they are specifically justified in the periodic technical report and
- they do not entail changes to the Agreement which would call into question the decision awarding the grant or breach the principle of equal treatment of applicants.

The beneficiaries must ensure that the Commission, the European Court of Auditors (ECA) and the European Anti-Fraud Office (OLAF) can exercise their rights under Articles 22 and 23 also towards their subcontractors.

13.1.2 The beneficiaries must ensure that their obligations under Articles 35, 36, 38 and 46 also apply to the subcontractors.

Beneficiaries that are ‘contracting authorities’ within the meaning of Directive 2004/18/EC (or 2014/24/EU) or ‘contracting entities’ within the meaning of Directive 2004/17/EC (or 2014/25/EU) must comply with the applicable national law on public procurement.

### **13.2 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under Article 13.1.1, the costs related to the subcontract concerned will be ineligible (see Article 6) and will be rejected (see Article 42).

If a beneficiary breaches any of its obligations under Article 13.1.2, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

## **ARTICLE 14 — IMPLEMENTATION OF ACTION TASKS BY LINKED THIRD PARTIES**

### **14.1 Rules for calling upon linked third parties to implement part of the action**

14.1.1 The following **affiliated entities**<sup>10</sup> and **third parties with a legal link to a beneficiary**<sup>11</sup> (**‘linked third parties’**) may implement the action tasks attributed to them in Annex 1:

- RAE GESTION SL (RAEG), affiliated or linked to RAE

The linked third parties may declare as eligible the costs they incur for implementing the action tasks in accordance with Article 6.3.

The beneficiaries must ensure that the Commission, the European Court of Auditors (ECA) and the European Anti-Fraud Office (OLAF) can exercise their rights under Articles 22 and 23 also towards their linked third parties.

14.1.2 The beneficiaries must ensure that their obligations under Articles 18, 20, 35, 36 and 38 also apply to their linked third parties.

## 14.2 Consequences of non-compliance

If any obligation under Article 14.1.1 is breached, the costs of the linked third party will be ineligible (see Article 6) and will be rejected (see Article 42).

If any obligation under Article 14.1.2 is breached, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

## ARTICLE 15 — FINANCIAL SUPPORT TO THIRD PARTIES

### 15.1 Rules for providing financial support to third parties

Not applicable

### 15.2 Financial support in the form of prizes

Not applicable

### 15.3 Consequences of non-compliance

---

<sup>10</sup> For the definition see Article 2.1(2) Rules for Participation Regulation No 1290/2013: **‘affiliated entity’** means any legal entity that is:

- under the direct or indirect control of a participant, or
- under the same direct or indirect control as the participant, or
- directly or indirectly controlling a participant.

‘Control’ may take any of the following forms:

- (a) the direct or indirect holding of more than 50% of the nominal value of the issued share capital in the legal entity concerned, or of a majority of the voting rights of the shareholders or associates of that entity;
- (b) the direct or indirect holding, in fact or in law, of decision-making powers in the legal entity concerned.

However the following relationships between legal entities shall not in themselves be deemed to constitute controlling relationships:

- (a) the same public investment corporation, institutional investor or venture-capital company has a direct or indirect holding of more than 50% of the nominal value of the issued share capital or a majority of voting rights of the shareholders or associates;
- (b) the legal entities concerned are owned or supervised by the same public body.

<sup>11</sup> **‘Third party with a legal link to a beneficiary’** is any legal entity which has a legal link to the beneficiary implying collaboration that is not limited to the action.

Not applicable

## ARTICLE 16 — PROVISION OF TRANS-NATIONAL OR VIRTUAL ACCESS TO RESEARCH INFRASTRUCTURE

### 16.1 Rules for providing trans-national access to research infrastructure

16.1.1 ‘**Access providers**’<sup>12</sup> must provide access to research infrastructure or installations<sup>13</sup> in accordance with the following conditions:

(a) access which must be provided:

The access must be free of charge, trans-national access to research infrastructure or installations for selected user-groups.

This access must include the logistical, technological and scientific support and the specific training that is usually provided to external researchers using the infrastructure.

(b) categories of users that may have access:

Trans-national access must be provided to selected ‘**user-groups**’, i.e. teams of one or more researchers (users) led by a ‘**user group leader**’.

The user group leader and the majority of the users must work in a country other than the country(ies) where the installation is located.

This rule does not apply:

- if access is provided by an International organisation, the Joint Research Centre (JRC), an ERIC or similar legal entities;
- in case of remote access to a set of installations located in different countries offering the same type of service.

Only user groups that are allowed to disseminate the results they have generated under the action may benefit from the access, unless the users are working for SMEs.

Access for user groups with a majority of users not working in a EU or associated country is limited to 20% of the total amount of units of access provided under the grant, unless a higher percentage is foreseen in Annex 1;

(c) procedure and criteria for selecting user groups:

The user groups must request access by submitting (in writing) a description of the work that they wish to carry out and the names, nationalities and home institutions of the users.

The user groups must be selected by a **selection panel** set up by the access providers.

---

<sup>12</sup> ‘**Access provider**’ means a beneficiary or linked third party that is in charge of providing access to one or more research infrastructures or installations, or part of them, as described in Annex 1.

<sup>13</sup> ‘**Installation**’ means a part or a service of a research infrastructure that could be used independently from the rest. A research infrastructure consists of one or more installations.

The selection panel must be composed of international experts in the field, at least half of them independent from the beneficiaries, unless otherwise specified in Annex 1.

The selection panel must assess all proposals received and recommend a short-list of the user groups that should benefit from access.

The selection panel must base its selection on scientific merit, taking into account that priority should be given to user groups composed of users who:

- have not previously used the installation and
- are working in countries where no equivalent research infrastructure exist.

It will apply the principles of transparency, fairness and impartiality.

(d) other conditions:

The access provider must request written approval from the Commission (see Article 52) for the selection of user groups requiring visits to the installation(s) exceeding 3 months, unless such visits are foreseen in Annex 1.

16.1.2 In addition, the access provider must:

- advertise widely, including on a dedicated website, the access offered under the Agreement;
- promote equal opportunities in advertising the access and take into account the gender dimension when defining the support provided to users;
- ensure that users comply with the terms and conditions of this Agreement;
- ensure that its obligations under Articles 35, 36, 38 and 46 also apply to the users.

## **16.2 Rules for providing virtual access to research infrastructure**

‘**Access providers**’ must provide access to research infrastructure or installations in accordance with the following conditions:

(a) access which must be provided:

The access must be free of charge, virtual access to research infrastructure or installations.

‘Virtual access’ means open and free access through communication networks to resources needed for research, without selecting the researchers to whom access is provided;

(b) other conditions:

The access provider must have the virtual access services assessed periodically by a board composed of international experts in the field, at least half of whom must be independent from the beneficiaries, unless otherwise specified in Annex 1.

## **16.3 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under Articles 16.1.1 and 16.2, the costs of access will be ineligible (see Article 6) and will be rejected (see Article 42).

If a beneficiary breaches any of its obligations under Articles 16.1.2, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

## **SECTION 2 RIGHTS AND OBLIGATIONS RELATED TO THE GRANT ADMINISTRATION**

### **ARTICLE 17 — GENERAL OBLIGATION TO INFORM**

#### **17.1 General obligation to provide information upon request**

The beneficiaries must provide — during implementation of the action or afterwards and in accordance with Article 41.2 — any information requested in order to verify eligibility of the costs, proper implementation of the action and compliance with any other obligation under the Agreement.

#### **17.2 Obligation to keep information up to date and to inform about events and circumstances likely to affect the Agreement**

Each beneficiary must keep information stored in the Participant Portal Beneficiary Register (via the electronic exchange system; see Article 52) up to date, in particular, its name, address, legal representatives, legal form and organisation type.

Each beneficiary must immediately inform the coordinator — which must immediately inform the Commission and the other beneficiaries — of any of the following:

- (a) **events** which are likely to affect significantly or delay the implementation of the action or the EU's financial interests, in particular:
  - (i) changes in its legal, financial, technical, organisational or ownership situation or those of its linked third parties and
  - (ii) changes in the name, address, legal form, organisation type of its linked third parties;
- (b) **circumstances** affecting:
  - (i) the decision to award the grant or
  - (ii) compliance with requirements under the Agreement.

#### **17.3 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

### **ARTICLE 18 — KEEPING RECORDS — SUPPORTING DOCUMENTATION**

## 18.1 Obligation to keep records and other supporting documentation

The beneficiaries must — for a period of five years after the payment of the balance — keep records and other supporting documentation in order to prove the proper implementation of the action and the costs they declare as eligible.

They must make them available upon request (see Article 17) or in the context of checks, reviews, audits or investigations (see Article 22).

If there are on-going checks, reviews, audits, investigations, litigation or other pursuits of claims under the Agreement (including the extension of findings; see Article 22), the beneficiaries must keep the records and other supporting documentation until the end of these procedures.

The beneficiaries must keep the original documents. Digital and digitalised documents are considered originals if they are authorised by the applicable national law. The Commission may accept non-original documents if it considers that they offer a comparable level of assurance.

### 18.1.1 Records and other supporting documentation on the scientific and technical implementation

The beneficiaries must keep records and other supporting documentation on scientific and technical implementation of the action in line with the accepted standards in the respective field.

### 18.1.2 Records and other documentation to support the costs declared

The beneficiaries must keep the records and documentation supporting the costs declared, in particular the following:

- (a) for **actual costs**: adequate records and other supporting documentation to prove the costs declared, such as contracts, subcontracts, invoices and accounting records. In addition, the beneficiaries' usual cost accounting practices and internal control procedures must enable direct reconciliation between the amounts declared, the amounts recorded in their accounts and the amounts stated in the supporting documentation;
- (b) for **unit costs**: adequate records and other supporting documentation to prove the number of units declared. This documentation must include records of the names, nationalities, and home institutions of users, as well as the nature and quantity of access provided to them. Beneficiaries do not need to identify the actual eligible costs covered or to keep or provide supporting documentation (such as accounting statements) to prove the amount per unit.

In addition, **for direct personnel costs declared as unit costs calculated in accordance with the beneficiary's usual cost accounting practices**, the beneficiaries must keep adequate records and documentation to prove that the cost accounting practices used comply with the conditions set out in Article 6.2, Point A.

The beneficiaries and linked third parties may submit to the Commission, for approval, a certificate (drawn up in accordance with Annex 6) stating that their usual cost accounting practices comply with these conditions (**'certificate on the methodology'**). If the certificate is approved, costs declared in line with this methodology will not be challenged subsequently, unless the beneficiaries have concealed information for the purpose of the approval.

- (c) for **flat-rate costs**: adequate records and other supporting documentation to prove the eligibility



of the costs to which the flat-rate is applied. The beneficiaries do not need to identify the costs covered or provide supporting documentation (such as accounting statements) to prove the amount declared at a flat-rate.

In addition, for **personnel costs** (declared as actual costs or on the basis of unit costs), the beneficiaries must keep **time records** for the number of hours declared. The time records must be in writing and approved by the persons working on the action and their supervisors, at least monthly. In the absence of reliable time records of the hours worked on the action, the Commission may accept alternative evidence supporting the number of hours declared, if it considers that it offers an adequate level of assurance.

As an exception, for **persons working exclusively on the action**, there is no need to keep time records, if the beneficiary signs a **declaration** confirming that the persons concerned have worked exclusively on the action.

For costs declared by linked third parties (see Article 14), it is the beneficiary that must keep the originals of the financial statements and the certificates on the financial statements of the linked third parties.

## 18.2 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, costs insufficiently substantiated will be ineligible (see Article 6) and will be rejected (see Article 42), and the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

## ARTICLE 19 — SUBMISSION OF DELIVERABLES

### 19.1 Obligation to submit deliverables

The coordinator must submit the ‘**deliverables**’ identified in Annex 1, in accordance with the timing and conditions set out in it.

### 19.2 Consequences of non-compliance

If the coordinator breaches any of its obligations under this Article, the Commission may apply any of the measures described in Chapter 6.

## ARTICLE 20 — REPORTING — PAYMENT REQUESTS

### 20.1 Obligation to submit reports

The coordinator must submit to the Commission (see Article 52) the technical and financial reports set out in this Article. These reports include requests for payment and must be drawn up using the forms and templates provided in the electronic exchange system (see Article 52).

### 20.2 Reporting periods



The action is divided into the following ‘**reporting periods**’:

- RP1: from month 1 to month 18
- RP2: from month 19 to month 36
- RP3: from month 37 to month 48

### 20.3 Periodic reports — Requests for interim payments

The coordinator must submit a periodic report within 60 days following the end of each reporting period.

The **periodic report** must include the following:

(a) a ‘**periodic technical report**’ containing:

- (i) an **explanation of the work carried out** by the beneficiaries;
- (ii) an **overview of the progress** towards the objectives of the action, including milestones and deliverables identified in Annex 1.

This report must include explanations justifying the differences between work expected to be carried out in accordance with Annex 1 and that actually carried out.

The report must detail the exploitation and dissemination of the results and — if required in Annex 1 — an updated ‘**plan for the exploitation and dissemination of the results**’.

The report must indicate the communication activities;

The report must detail the access activity, indicating the members of the selection panel, the selection procedure, the exact amount of access provided to the user groups, the description of their work, and information on the users (including names, nationality and home institutions).

The reports must detail the access activity, with statistics on the virtual access provided in the period, including quantity, geographical distribution of users and, whenever possible, information/statistics on scientific outcomes (publications, patents, etc) acknowledging the use of the infrastructure;

- (iii) a **summary** for publication by the Commission;
- (iv) the answers to the ‘**questionnaire**’, covering issues related to the action implementation and the economic and societal impact, notably in the context of the Horizon 2020 key performance indicators and the Horizon 2020 monitoring requirements;

(b) a ‘**periodic financial report**’ containing:

- (i) an ‘**individual financial statement**’ (see Annex 4) from each beneficiary and from each linked third party, for the reporting period concerned.

The individual financial statement must detail the eligible costs (actual costs, unit costs and flat-rate costs; see Article 6) for each budget category (see Annex 2).

The beneficiaries and linked third parties must declare all eligible costs, even if — for actual costs, unit costs and flat-rate costs — they exceed the amounts indicated in the estimated

budget (see Annex 2). Amounts which are not declared in the individual financial statement will not be taken into account by the Commission.

If an individual financial statement is not submitted for a reporting period, it may be included in the periodic financial report for the next reporting period.

The individual financial statements of the last reporting period must also detail the **receipts of the action** (see Article 5.3.3).

Each beneficiary and each linked third party must **certify** that:

- the information provided is full, reliable and true;
  - the costs declared are eligible (see Article 6);
  - the costs can be substantiated by adequate records and supporting documentation (see Article 18) that will be produced upon request (see Article 17) or in the context of checks, reviews, audits and investigations (see Article 22), and
  - for the last reporting period: that all the receipts have been declared (see Article 5.3.3);
- (ii) an **explanation of the use of resources** and the information on subcontracting (see Article 13) and in-kind contributions provided by third parties (see Articles 11 and 12) from each beneficiary and from each linked third party, for the reporting period concerned;
- (iii) not applicable;
- (iv) a **‘periodic summary financial statement’**, created automatically by the electronic exchange system, consolidating the individual financial statements for the reporting period concerned and including — except for the last reporting period — the **request for interim payment**.

#### 20.4 Final report — Request for payment of the balance

In addition to the periodic report for the last reporting period, the coordinator must submit the final report within 60 days following the end of the last reporting period.

The **final report** must include the following:

- (a) a **‘final technical report’** with a **summary** for publication containing:
- (i) an overview of the results and their exploitation and dissemination;
  - (ii) the conclusions on the action, and
  - (iii) the socio-economic impact of the action;
- (b) a **‘final financial report’** containing:
- (i) a **‘final summary financial statement’**, created automatically by the electronic exchange system, consolidating the individual financial statements for all reporting periods and including the **request for payment of the balance** and

- (ii) a ‘**certificate on the financial statements**’ (drawn up in accordance with Annex 5) for each beneficiary and for each linked third party, if it requests a total contribution of EUR 325 000 or more, as reimbursement of actual costs and unit costs calculated on the basis of its usual cost accounting practices (see Article 5.2 and Article 6.2, Point A).

## 20.5 Information on cumulative expenditure incurred

Not applicable

## 20.6 Currency for financial statements and conversion into euro

Financial statements must be drafted in euro.

Beneficiaries and linked third parties with accounting established in a currency other than the euro must convert the costs recorded in their accounts into euro, at the average of the daily exchange rates published in the C series of the *Official Journal of the European Union*, calculated over the corresponding reporting period.

If no daily euro exchange rate is published in the *Official Journal of the European Union* for the currency in question, they must be converted at the average of the monthly accounting rates published on the Commission’s website, calculated over the corresponding reporting period.

Beneficiaries and linked third parties with accounting established in euro must convert costs incurred in another currency into euro according to their usual accounting practices.

## 20.7 Language of reports

All reports (technical and financial reports, including financial statements) must be submitted in the language of the Agreement.

## 20.8 Consequences of non-compliance

If the reports submitted do not comply with this Article, the Commission may suspend the payment deadline (see Article 47) and apply any of the other measures described in Chapter 6.

If the coordinator breaches its obligation to submit the reports and if it fails to comply with this obligation within 30 days following a written reminder, the Commission may terminate the Agreement (see Article 50) or apply any of the other measures described in Chapter 6.

# ARTICLE 21 — PAYMENTS AND PAYMENT ARRANGEMENTS

## 21.1 Payments to be made

The following payments will be made to the coordinator:

- one **pre-financing payment**;
- one or more **interim payments**, on the basis of the request(s) for interim payment (see Article 20), and

- one **payment of the balance**, on the basis of the request for payment of the balance (see Article 20).

## **21.2 Pre-financing payment — Amount — Amount retained for the Guarantee Fund**

The aim of the pre-financing is to provide the beneficiaries with a float.

It remains the property of the EU until the payment of the balance.

The amount of the pre-financing payment will be EUR **2,666,649.33** (two million six hundred and sixty six thousand six hundred and forty nine EURO and thirty three eurocents).

The Commission will — except if Article 48 applies — make the pre-financing payment to the coordinator within 30 days, either from the entry into force of the Agreement (see Article 58) or from 10 days before the starting date of the action (see Article 3), whichever is the latest.

An amount of EUR **249,998.38** (two hundred and forty nine thousand nine hundred and ninety eight EURO and thirty eight eurocents), corresponding to 5% of the maximum grant amount (see Article 5.1), is retained by the Commission from the pre-financing payment and transferred into the ‘**Guarantee Fund**’.

## **21.3 Interim payments — Amount — Calculation**

Interim payments reimburse the eligible costs incurred for the implementation of the action during the corresponding reporting periods.

The Commission will pay to the coordinator the amount due as interim payment within 90 days from receiving the periodic report (see Article 20.3), except if Articles 47 or 48 apply.

Payment is subject to the approval of the periodic report. Its approval does not imply recognition of the compliance, authenticity, completeness or correctness of its content.

The **amount due as interim payment** is calculated by the Commission in the following steps:

Step 1 – Application of the reimbursement rates

Step 2 – Limit to 90% of the maximum grant amount

### **21.3.1 Step 1 — Application of the reimbursement rates**

The reimbursement rate(s) (see Article 5.2) are applied to the eligible costs (actual costs, unit costs and flat-rate costs ; see Article 6) declared by the beneficiaries and the linked third parties (see Article 20) and approved by the Commission (see above) for the concerned reporting period.

### **21.3.2 Step 2 — Limit to 90% of the maximum grant amount**

The total amount of pre-financing and interim payments must not exceed 90% of the maximum grant amount set out in Article 5.1. The maximum amount for the interim payment will be calculated as follows:

{90% of the maximum grant amount (see Article 5.1)  
 minus  
 {pre-financing and previous interim payments}}.

#### **21.4 Payment of the balance — Amount — Calculation — Release of the amount retained for the Guarantee Fund**

The payment of the balance reimburses the remaining part of the eligible costs incurred by the beneficiaries for the implementation of the action.

If the total amount of earlier payments is greater than the final grant amount (see Article 5.3), the payment of the balance takes the form of a recovery (see Article 44).

If the total amount of earlier payments is lower than the final grant amount, the Commission will pay the balance within 90 days from receiving the final report (see Article 20.4), except if Articles 47 or 48 apply.

Payment is subject to the approval of the final report. Its approval does not imply recognition of the compliance, authenticity, completeness or correctness of its content.

The **amount due as the balance** is calculated by the Commission by deducting the total amount of pre-financing and interim payments (if any) already made, from the final grant amount determined in accordance with Article 5.3:

{final grant amount (see Article 5.3)  
 minus  
 {pre-financing and interim payments (if any) made}}.

At the payment of the balance, the amount retained for the Guarantee Fund (see above) will be released and:

- if the balance is positive: the amount released will be paid in full to the coordinator together with the amount due as the balance;
- if the balance is negative (payment of the balance taking the form of recovery): it will be deducted from the amount released (see Article 44.1.2). If the resulting amount:
  - is positive, it will be paid to the coordinator
  - is negative, it will be recovered.

The amount to be paid may however be offset — without the beneficiaries' consent — against any other amount owed by a beneficiary to the Commission or an executive agency (under the EU or Euratom budget), up to the maximum EU contribution indicated, for that beneficiary, in the estimated budget (see Annex 2).

#### **21.5 Notification of amounts due**

When making payments, the Commission will formally notify to the coordinator the amount due, specifying whether it concerns an interim payment or the payment of the balance.

For the payment of the balance, the notification will also specify the final grant amount.

In the case of reduction of the grant or recovery of undue amounts, the notification will be preceded by the contradictory procedure set out in Articles 43 and 44.

### **21.6 Currency for payments**

The Commission will make all payments in euro.

### **21.7 Payments to the coordinator — Distribution to the beneficiaries**

Payments will be made to the coordinator.

Payments to the coordinator will discharge the Commission from its payment obligation.

The coordinator must distribute the payments between the beneficiaries without unjustified delay.

Pre-financing may however be distributed only:

- (a) if the minimum number of beneficiaries set out in the call for proposals has acceded to the Agreement (see Article 56) and
- (b) to beneficiaries that have acceded to the Agreement (see Article 56).

### **21.8 Bank account for payments**

All payments will be made to the following bank account:

Name of bank: BANK OF SLOVENIA  
Full name of the account holder: JOZEF STEFAN INSTITUTE  
Full account number (including bank codes): ()  
IBAN code: SI56011006030344242

### **21.9 Costs of payment transfers**

The cost of the payment transfers is borne as follows:

- the Commission bears the cost of transfers charged by its bank;
- the beneficiary bears the cost of transfers charged by its bank;
- the party causing a repetition of a transfer bears all costs of the repeated transfer.

### **21.10 Date of payment**

Payments by the Commission are considered to have been carried out on the date when they are debited to its account.

### **21.11 Consequences of non-compliance**

21.11.1 If the Commission does not pay within the payment deadlines (see above), the beneficiaries are entitled to **late-payment interest** at the rate applied by the European Central Bank (ECB) for its main refinancing operations in euros ('reference rate'), plus three and a half points. The reference rate

is the rate in force on the first day of the month in which the payment deadline expires, as published in the C series of the *Official Journal of the European Union*.

If the late-payment interest is lower than or equal to EUR 200, it will be paid to the coordinator only upon request submitted within two months of receiving the late payment.

Late-payment interest is not due if all beneficiaries are EU Member States (including regional and local government authorities or other public bodies acting on behalf of a Member State for the purpose of this Agreement).

Suspension of the payment deadline or payments (see Articles 47 and 48) will not be considered as late payment.

Late-payment interest covers the period running from the day following the due date for payment (see above), up to and including the date of payment.

Late-payment interest is not considered for the purposes of calculating the final grant amount.

21.11.2 If the coordinator breaches any of its obligations under this Article, the grant may be reduced (see Article 43) and the Agreement or the participation of the coordinator may be terminated (see Article 50).

Such breaches may also lead to any of the other measures described in Chapter 6.

## **ARTICLE 22 — CHECKS, REVIEWS, AUDITS AND INVESTIGATIONS — EXTENSION OF FINDINGS**

### **22.1 Checks, reviews and audits by the Commission**

#### **22.1.1 Right to carry out checks**

The Commission will — during the implementation of the action or afterwards — check the proper implementation of the action and compliance with the obligations under the Agreement, including assessing deliverables and reports.

For this purpose the Commission may be assisted by external persons or bodies.

The Commission may also request additional information in accordance with Article 17. The Commission may request beneficiaries to provide such information to it directly.

Information provided must be accurate, precise and complete and in the format requested, including electronic format.

#### **22.1.2 Right to carry out reviews**

The Commission may — during the implementation of the action or afterwards — carry out reviews on the proper implementation of the action (including assessment of deliverables and reports), compliance with the obligations under the Agreement and continued scientific or technological relevance of the action.

Reviews may be started up to two years after the payment of the balance. They will be formally



notified to the coordinator or beneficiary concerned and will be considered to have started on the date of the formal notification.

If the review is carried out on a third party (see Articles 10 to 16), the beneficiary concerned must inform the third party.

The Commission may carry out reviews directly (using its own staff) or indirectly (using external persons or bodies appointed to do so). It will inform the coordinator or beneficiary concerned of the identity of the external persons or bodies. They have the right to object to the appointment on grounds of commercial confidentiality.

The coordinator or beneficiary concerned must provide — within the deadline requested — any information and data in addition to deliverables and reports already submitted (including information on the use of resources). The Commission may request beneficiaries to provide such information to it directly.

The coordinator or beneficiary concerned may be requested to participate in meetings, including with external experts.

For **on-the-spot** reviews, the beneficiaries must allow access to their sites and premises, including to external persons or bodies, and must ensure that information requested is readily available.

Information provided must be accurate, precise and complete and in the format requested, including electronic format.

On the basis of the review findings, a '**review report**' will be drawn up.

The Commission will formally notify the review report to the coordinator or beneficiary concerned, which has 30 days to formally notify observations ('**contradictory review procedure**').

Reviews (including review reports) are in the language of the Agreement.

### **22.1.3 Right to carry out audits**

The Commission may — during the implementation of the action or afterwards — carry out audits on the proper implementation of the action and compliance with the obligations under the Agreement.

Audits may be started up to two years after the payment of the balance. They will be formally notified to the coordinator or beneficiary concerned and will be considered to have started on the date of the formal notification.

If the audit is carried out on a third party (see Articles 10 to 16), the beneficiary concerned must inform the third party.

The Commission may carry out audits directly (using its own staff) or indirectly (using external persons or bodies appointed to do so). It will inform the coordinator or beneficiary concerned of the identity of the external persons or bodies. They have the right to object to the appointment on grounds of commercial confidentiality.

The coordinator or beneficiary concerned must provide — within the deadline requested — any information (including complete accounts, individual salary statements or other personal data) to



verify compliance with the Agreement. The Commission may request beneficiaries to provide such information to it directly.

For **on-the-spot** audits, the beneficiaries must allow access to their sites and premises, including to external persons or bodies, and must ensure that information requested is readily available.

Information provided must be accurate, precise and complete and in the format requested, including electronic format.

On the basis of the audit findings, a '**draft audit report**' will be drawn up.

The Commission will formally notify the draft audit report to the coordinator or beneficiary concerned, which has 30 days to formally notify observations ('**contradictory audit procedure**'). This period may be extended by the Commission in justified cases.

The '**final audit report**' will take into account observations by the coordinator or beneficiary concerned. The report will be formally notified to it.

Audits (including audit reports) are in the language of the Agreement.

The Commission may also access the beneficiaries' statutory records for the periodical assessment of unit costs or flat-rate amounts.

## **22.2 Investigations by the European Anti-Fraud Office (OLAF)**

Under Regulations No 883/2013<sup>14</sup> and No 2185/96<sup>15</sup> (and in accordance with their provisions and procedures) the European Anti-Fraud Office (OLAF) may — at any moment during implementation of the action or afterwards — carry out investigations, including on-the-spot checks and inspections, to establish whether there has been fraud, corruption or any other illegal activity affecting the financial interests of the EU.

## **22.3 Checks and audits by the European Court of Auditors (ECA)**

Under Article 287 of the Treaty on the Functioning of the European Union (TFEU) and Article 161 of the Financial Regulation No 966/2012<sup>17</sup>, the European Court of Auditors (ECA) may — at any moment during implementation of the action or afterwards — carry out audits.

The ECA has the right of access for the purpose of checks and audits.

## **22.4 Checks, reviews, audits and investigations for international organisations**

Not applicable

---

<sup>14</sup> Regulation (EU, Euratom) No 883/2013 of the European Parliament and of the Council of 11 September 2013 concerning investigations conducted by the European Anti-Fraud Office (OLAF) and repealing Regulation (EC) No 1073/1999 of the European Parliament and of the Council and Council Regulation (Euratom) No 1074/1999 (OJ L 248, 18.09.2013, p. 1).

<sup>15</sup> Council Regulation (Euratom, EC) No 2185/1996 of 11 November 1996 concerning on-the-spot checks and inspections carried out by the Commission in order to protect the European Communities' financial interests against fraud and other irregularities (OJ L 292, 15.11.1996, p. 2).

<sup>17</sup> Regulation (EU, Euratom) No 966/2012 of the European Parliament and of the Council of 25 October 2012 on the financial rules applicable to the general budget of the Union and repealing Council Regulation (EC, Euratom) No 1605/2002 (OJ L 298, 26.10.2012, p. 1).

## 22.5 Consequences of findings in checks, reviews, audits and investigations — Extension of findings

### 22.5.1 Findings in this grant

Findings in checks, reviews, audits or investigations carried out in the context of this grant may lead to the rejection of ineligible costs (see Article 42), reduction of the grant (see Article 43), recovery of undue amounts (see Article 44) or to any of the other measures described in Chapter 6.

Rejection of costs or reduction of the grant after the payment of the balance will lead to a revised final grant amount (see Article 5.4).

Findings in checks, reviews, audits or investigations may lead to a request for amendment for the modification of Annex 1 (see Article 55).

Checks, reviews, audits or investigations that find systemic or recurrent errors, irregularities, fraud or breach of obligations may also lead to consequences in other EU or Euratom grants awarded under similar conditions (**‘extension of findings from this grant to other grants’**).

Moreover, findings arising from an OLAF investigation may lead to criminal prosecution under national law.

### 22.5.2 Findings in other grants

The Commission may extend findings from other grants to this grant (**‘extension of findings from other grants to this grant’**), if:

- (a) the beneficiary concerned is found, in other EU or Euratom grants awarded under similar conditions, to have committed systemic or recurrent errors, irregularities, fraud or breach of obligations that have a material impact on this grant and
- (b) those findings are formally notified to the beneficiary concerned — together with the list of grants affected by the findings — no later than two years after the payment of the balance of this grant.

The extension of findings may lead to the rejection of costs (see Article 42), reduction of the grant (see Article 43), recovery of undue amounts (see Article 44), suspension of payments (see Article 48), suspension of the action implementation (see Article 49) or termination (see Article 50).

### 22.5.3 Procedure

The Commission will formally notify the beneficiary concerned the systemic or recurrent errors and its intention to extend these audit findings, together with the list of grants affected.

22.5.3.1 If the findings concern **eligibility of costs**: the formal notification will include:

- (a) an invitation to submit observations on the list of grants affected by the findings;
- (b) the request to submit **revised financial statements** for all grants affected;
- (c) the **correction rate for extrapolation** established by the Commission on the basis of the systemic or recurrent errors, to calculate the amounts to be rejected if the beneficiary concerned:

- (i) considers that the submission of revised financial statements is not possible or practicable or
- (ii) does not submit revised financial statements.

The beneficiary concerned has 90 days from receiving notification to submit observations, revised financial statements or to propose a duly substantiated **alternative correction method**. This period may be extended by the Commission in justified cases.

The Commission may then start a rejection procedure in accordance with Article 42, on the basis of:

- the revised financial statements, if approved;
- the proposed alternative correction method, if accepted

or

- the initially notified correction rate for extrapolation, if it does not receive any observations or revised financial statements, does not accept the observations or the proposed alternative correction method or does not approve the revised financial statements.

22.5.3.2 If the findings concern **substantial errors, irregularities or fraud or serious breach of obligations**: the formal notification will include:

- (a) an invitation to submit observations on the list of grants affected by the findings and
- (b) the flat-rate the Commission intends to apply according to the principle of proportionality.

The beneficiary concerned has 90 days from receiving notification to submit observations or to propose a duly substantiated alternative flat-rate.

The Commission may then start a reduction procedure in accordance with Article 43, on the basis of:

- the proposed alternative flat-rate, if accepted

or

- the initially notified flat-rate, if it does not receive any observations or does not accept the observations or the proposed alternative flat-rate.

## 22.6 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, any insufficiently substantiated costs will be ineligible (see Article 6) and will be rejected (see Article 42).

Such breaches may also lead to any of the other measures described in Chapter 6.

## ARTICLE 23 — EVALUATION OF THE IMPACT OF THE ACTION

### 23.1 Right to evaluate the impact of the action

The Commission may carry out interim and final evaluations of the impact of the action measured against the objective of the EU programme.

Evaluations may be started during implementation of the action and up to five years after the payment of the balance. The evaluation is considered to start on the date of the formal notification to the coordinator or beneficiaries.

The Commission may make these evaluations directly (using its own staff) or indirectly (using external bodies or persons it has authorised to do so).

The coordinator or beneficiaries must provide any information relevant to evaluate the impact of the action, including information in electronic format.

### **23.2 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under this Article, the Commission may apply the measures described in Chapter 6.

## **SECTION 3 RIGHTS AND OBLIGATIONS RELATED TO BACKGROUND AND RESULTS**

### **SUBSECTION 1 GENERAL**

#### **ARTICLE 23a — MANAGEMENT OF INTELLECTUAL PROPERTY**

##### **23a.1 Obligation to take measures to implement the Commission Recommendation on the management of intellectual property in knowledge transfer activities**

Beneficiaries that are universities or other public research organisations must take measures to implement the principles set out in Points 1 and 2 of the Code of Practice annexed to the Commission Recommendation on the management of intellectual property in knowledge transfer activities<sup>17</sup>.

This does not change the obligations set out in Subsections 2 and 3 of this Section.

The beneficiaries must ensure that researchers and third parties involved in the action are aware of them.

##### **23a.2 Consequences of non-compliance**

If a beneficiary breaches its obligations under this Article, the Commission may apply any of the measures described in Chapter 6.

### **SUBSECTION 2 RIGHTS AND OBLIGATIONS RELATED TO BACKGROUND**

#### **ARTICLE 24 — AGREEMENT ON BACKGROUND**

---

<sup>17</sup> Commission Recommendation C(2008) 1329 of 10.4.2008 on the management of intellectual property in knowledge transfer activities and the Code of Practice for universities and other public research institutions attached to this recommendation.

## 24.1 Agreement on background

The beneficiaries must identify and agree (in writing) on the background for the action (**‘agreement on background’**).

**‘Background’** means any data, know-how or information — whatever its form or nature (tangible or intangible), including any rights such as intellectual property rights — that:

- (a) is held by the beneficiaries before they acceded to the Agreement, and
- (b) is needed to implement the action or exploit the results.

## 24.2 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

## ARTICLE 25 — ACCESS RIGHTS TO BACKGROUND

### 25.1 Exercise of access rights — Waiving of access rights — No sub-licensing

To exercise access rights, this must first be requested in writing (**‘request for access’**).

**‘Access rights’** means rights to use results or background under the terms and conditions laid down in this Agreement.

Waivers of access rights are not valid unless in writing.

Unless agreed otherwise, access rights do not include the right to sub-license.

### 25.2 Access rights for other beneficiaries, for implementing their own tasks under the action

The beneficiaries must give each other access — on a royalty-free basis — to background needed to implement their own tasks under the action, unless the beneficiary that holds the background has — before acceding to the Agreement —:

- (a) informed the other beneficiaries that access to its background is subject to legal restrictions or limits, including those imposed by the rights of third parties (including personnel), or
- (b) agreed with the other beneficiaries that access would not be on a royalty-free basis.

### 25.3 Access rights for other beneficiaries, for exploiting their own results

The beneficiaries must give each other access — under fair and reasonable conditions — to background needed for exploiting their own results, unless the beneficiary that holds the background has — before acceding to the Agreement — informed the other beneficiaries that access to its background is subject to legal restrictions or limits, including those imposed by the rights of third parties (including personnel).

**‘Fair and reasonable conditions’** means appropriate conditions, including possible financial terms

or royalty-free conditions, taking into account the specific circumstances of the request for access, for example the actual or potential value of the results or background to which access is requested and/or the scope, duration or other characteristics of the exploitation envisaged.

Requests for access may be made — unless agreed otherwise — up to one year after the period set out in Article 3.

#### **25.4 Access rights for affiliated entities**

Unless otherwise agreed in the consortium agreement, access to background must also be given — under fair and reasonable conditions (see above; Article 25.3) and unless it is subject to legal restrictions or limits, including those imposed by the rights of third parties (including personnel) — to affiliated entities<sup>18</sup> established in an EU Member State or ‘**associated country**’<sup>19</sup>, if this is needed to exploit the results generated by the beneficiaries to which they are affiliated.

Unless agreed otherwise (see above; Article 25.1), the affiliated entity concerned must make the request directly to the beneficiary that holds the background.

Requests for access may be made — unless agreed otherwise — up to one year after the period set out in Article 3.

#### **25.5 Access rights for third parties**

The access provider must — unless it is subject to legal restrictions or limits, including those imposed by the rights of third parties (including personnel) — give users royalty-free access to background needed to implement the action.

The access provider must inform the users as soon as possible of any restriction which might substantially affect the granting of access rights.

#### **25.6 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

### **SUBSECTION 3 RIGHTS AND OBLIGATIONS RELATED TO RESULTS**

#### **ARTICLE 26 — OWNERSHIP OF RESULTS**

##### **26.1 Ownership by the beneficiary that generates the results**

Results are owned by the beneficiary that generates them.

‘**Results**’ means any (tangible or intangible) output of the action such as data, knowledge or

<sup>18</sup> For the definition, see ‘affiliated entity’ footnote (Article 14.1).

<sup>19</sup> For the definition, see Article 2.1(3) of the Rules for Participation Regulation No 1290/2013: ‘**associated country**’ means a third country which is party to an international agreement with the Union, as identified in Article 7 of Horizon 2020 Framework Programme Regulation No 1291/2013. Article 7 sets out the conditions for association of non-EU countries to Horizon 2020.

information — whatever its form or nature, whether it can be protected or not — that is generated in the action, as well as any rights attached to it, including intellectual property rights.

## 26.2 Joint ownership by several beneficiaries

Two or more beneficiaries own results jointly if:

- (a) they have jointly generated them and
- (b) it is not possible to:
  - (i) establish the respective contribution of each beneficiary, or
  - (ii) separate them for the purpose of applying for, obtaining or maintaining their protection (see Article 27).

The joint owners must agree (in writing) on the allocation and terms of exercise of their joint ownership (**‘joint ownership agreement’**), to ensure compliance with their obligations under this Agreement.

Unless otherwise agreed in the joint ownership agreement, each joint owner may grant non-exclusive licences to third parties to exploit jointly-owned results (without any right to sub-license), if the other joint owners are given:

- (a) at least 45 days advance notice and
- (b) fair and reasonable compensation.

Once the results have been generated, joint owners may agree (in writing) to apply another regime than joint ownership (such as, for instance, transfer to a single owner (see Article 30) with access rights for the others).

## 26.3 Rights of third parties (including personnel)

If third parties (including personnel) may claim rights to the results, the beneficiary concerned must ensure that it complies with its obligations under the Agreement.

If a third party generates results, the beneficiary concerned must obtain all necessary rights (transfer, licences or other) from the third party, in order to be able to respect its obligations as if those results were generated by the beneficiary itself.

If obtaining the rights is impossible, the beneficiary must refrain from using the third party to generate the results.

## 26.4 EU ownership, to protect results

26.4.1 The EU may — with the consent of the beneficiary concerned — assume ownership of results to protect them, if a beneficiary intends — up to four years after the period set out in Article 3 — to disseminate its results without protecting them, except in any of the following cases:

- (a) the lack of protection is because protecting the results is not possible, reasonable or justified (given the circumstances);



- (b) the lack of protection is because there is a lack of potential for commercial or industrial exploitation, or
- (c) the beneficiary intends to transfer the results to another beneficiary or third party established in an EU Member State or associated country, which will protect them.

Before the results are disseminated and unless any of the cases above under Points (a), (b) or (c) applies, the beneficiary must formally notify the Commission and at the same time inform it of any reasons for refusing consent. The beneficiary may refuse consent only if it can show that its legitimate interests would suffer significant harm.

If the Commission decides to assume ownership, it will formally notify the beneficiary concerned within 45 days of receiving notification.

No dissemination relating to these results may take place before the end of this period or, if the Commission takes a positive decision, until it has taken the necessary steps to protect the results.

26.4.2 The EU may — with the consent of the beneficiary concerned — assume ownership of results to protect them, if a beneficiary intends — up to four years after the period set out in Article 3 — to stop protecting them or not to seek an extension of protection, except in any of the following cases:

- (a) the protection is stopped because of a lack of potential for commercial or industrial exploitation;
- (b) an extension would not be justified given the circumstances.

A beneficiary that intends to stop protecting results or not seek an extension must — unless any of the cases above under Points (a) or (b) applies — formally notify the Commission at least 60 days before the protection lapses or its extension is no longer possible and at the same time inform it of any reasons for refusing consent. The beneficiary may refuse consent only if it can show that its legitimate interests would suffer significant harm.

If the Commission decides to assume ownership, it will formally notify the beneficiary concerned within 45 days of receiving notification.

## **26.5 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such breaches may also lead to the any of the other measures described in Chapter 6.

## **ARTICLE 27 — PROTECTION OF RESULTS — VISIBILITY OF EU FUNDING**

### **27.1 Obligation to protect the results**

Each beneficiary must examine the possibility of protecting its results and must adequately protect them — for an appropriate period and with appropriate territorial coverage — if:

- (a) the results can reasonably be expected to be commercially or industrially exploited and
- (b) protecting them is possible, reasonable and justified (given the circumstances).



When deciding on protection, the beneficiary must consider its own legitimate interests and the legitimate interests (especially commercial) of the other beneficiaries.

## 27.2 EU ownership, to protect the results

If a beneficiary intends not to protect its results, to stop protecting them or not seek an extension of protection, the EU may — under certain conditions (see Article 26.4) — assume ownership to ensure their (continued) protection.

## 27.3 Information on EU funding

Applications for protection of results (including patent applications) filed by or on behalf of a beneficiary must — unless the Commission requests or agrees otherwise or unless it is impossible — include the following:

“The project leading to this application has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 731015”.

## 27.4 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such a breach may also lead to any of the other measures described in Chapter 6.

# ARTICLE 28 — EXPLOITATION OF RESULTS

## 28.1 Obligation to exploit the results

Each beneficiary must — up to four years after the period set out in Article 3 — take measures aiming to ensure ‘**exploitation**’ of its results (either directly or indirectly, in particular through transfer or licensing; see Article 30) by:

- (a) using them in further research activities (outside the action);
- (b) developing, creating or marketing a product or process;
- (c) creating and providing a service, or
- (d) using them in standardisation activities.

This does not change the security obligations in Article 37, which still apply.

## 28.2 Results that could contribute to European or international standards — Information on EU funding

If results are incorporated in a standard, the beneficiary concerned must — unless the Commission requests or agrees otherwise or unless it is impossible — ask the standardisation body to include the following statement in (information related to) the standard:

“Results incorporated in this standard received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 731015”.

### **28.3 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced in accordance with Article 43.

Such a breach may also lead to any of the other measures described in Chapter 6.

## **ARTICLE 29 — DISSEMINATION OF RESULTS — OPEN ACCESS — VISIBILITY OF EU FUNDING**

### **29.1 Obligation to disseminate results**

Unless it goes against their legitimate interests, each beneficiary must — as soon as possible — ‘**disseminate**’ its results by disclosing them to the public by appropriate means (other than those resulting from protecting or exploiting the results), including in scientific publications (in any medium).

This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply.

A beneficiary that intends to disseminate its results must give advance notice to the other beneficiaries of — unless agreed otherwise — at least 45 days, together with sufficient information on the results it will disseminate.

Any other beneficiary may object within — unless agreed otherwise — 30 days of receiving notification, if it can show that its legitimate interests in relation to the results or background would be significantly harmed. In such cases, the dissemination may not take place unless appropriate steps are taken to safeguard these legitimate interests.

If a beneficiary intends not to protect its results, it may — under certain conditions (see Article 26.4.1) — need to formally notify the Commission before dissemination takes place.

### **29.2 Open access to scientific publications**

Each beneficiary must ensure open access (free of charge online access for any user) to all peer-reviewed scientific publications relating to its results.

In particular, it must:

- (a) as soon as possible and at the latest on publication, deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications;

Moreover, the beneficiary must aim to deposit at the same time the research data needed to validate the results presented in the deposited scientific publications.

- (b) ensure open access to the deposited publication — via the repository — at the latest:
  - (i) on publication, if an electronic version is available for free via the publisher, or

- (ii) within six months of publication (twelve months for publications in the social sciences and humanities) in any other case.
- (c) ensure open access — via the repository — to the bibliographic metadata that identify the deposited publication.

The bibliographic metadata must be in a standard format and must include all of the following:

- the terms “European Union (EU)” and “Horizon 2020”;
- the name of the action, acronym and grant number;
- the publication date, and length of embargo period if applicable, and
- a persistent identifier.

### 29.3 Open access to research data

Regarding the digital research data generated in the action (**‘data’**), the beneficiaries must:

- (a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:
  - (i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;
  - (ii) other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan' (see Annex 1);
- (b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).

This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply.

As an exception, the beneficiaries do not have to ensure open access to specific parts of their research data if the achievement of the action's main objective, as described in Annex 1, would be jeopardised by making those specific parts of the research data openly accessible. In this case, the data management plan must contain the reasons for not giving access.

### 29.4 Information on EU funding — Obligation and right to use the EU emblem

Unless the Commission requests or agrees otherwise or unless it is impossible, any dissemination of results (in any form, including electronic) must:

- (a) display the EU emblem and
- (b) include the following text:

“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 731015”.

When displayed together with another logo, the EU emblem must have appropriate prominence.

For the purposes of their obligations under this Article, the beneficiaries may use the EU emblem without first obtaining approval from the Commission.

This does not however give them the right to exclusive use.

Moreover, they may not appropriate the EU emblem or any similar trademark or logo, either by registration or by any other means.

### **29.5 Disclaimer excluding Commission responsibility**

Any dissemination of results must indicate that it reflects only the author's view and that the Commission is not responsible for any use that may be made of the information it contains.

### **29.6 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such a breach may also lead to any of the other measures described in Chapter 6.

## **ARTICLE 30 — TRANSFER AND LICENSING OF RESULTS**

### **30.1 Transfer of ownership**

Each beneficiary may transfer ownership of its results.

It must however ensure that its obligations under Articles 26.2, 26.4, 27, 28, 29, 30 and 31 also apply to the new owner and that this owner has the obligation to pass them on in any subsequent transfer.

This does not change the security obligations in Article 37, which still apply.

Unless agreed otherwise (in writing) for specifically-identified third parties or unless impossible under applicable EU and national laws on mergers and acquisitions, a beneficiary that intends to transfer ownership of results must give at least 45 days advance notice (or less if agreed in writing) to the other beneficiaries that still have (or still may request) access rights to the results. This notification must include sufficient information on the new owner to enable any beneficiary concerned to assess the effects on its access rights.

Unless agreed otherwise (in writing) for specifically-identified third parties, any other beneficiary may object within 30 days of receiving notification (or less if agreed in writing), if it can show that the transfer would adversely affect its access rights. In this case, the transfer may not take place until agreement has been reached between the beneficiaries concerned.

### **30.2 Granting licenses**

Each beneficiary may grant licences to its results (or otherwise give the right to exploit them), if:

- (a) this does not impede the access rights under Article 31 and
- (b) not applicable.

In addition to Points (a) and (b), exclusive licences for results may be granted only if all the other beneficiaries concerned have waived their access rights (see Article 31.1).

This does not change the dissemination obligations in Article 29 or security obligations in Article 37, which still apply.

### **30.3 Commission right to object to transfers or licensing**

Not applicable

### **30.4 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such a breach may also lead to any of the other measures described in Chapter 6.

## **ARTICLE 31 — ACCESS RIGHTS TO RESULTS**

### **31.1 Exercise of access rights — Waiving of access rights — No sub-licensing**

The conditions set out in Article 25.1 apply.

The obligations set out in this Article do not change the security obligations in Article 37, which still apply.

### **31.2 Access rights for other beneficiaries, for implementing their own tasks under the action**

The beneficiaries must give each other access — on a royalty-free basis — to results needed for implementing their own tasks under the action.

### **31.3 Access rights for other beneficiaries, for exploiting their own results**

The beneficiaries must give each other — under fair and reasonable conditions (see Article 25.3) — access to results needed for exploiting their own results.

Requests for access may be made — unless agreed otherwise — up to one year after the period set out in Article 3.

### **31.4 Access rights of affiliated entities**

Unless agreed otherwise in the consortium agreement, access to results must also be given — under fair and reasonable conditions (Article 25.3) — to affiliated entities established in an EU Member State or associated country, if this is needed for those entities to exploit the results generated by the beneficiaries to which they are affiliated.

Unless agreed otherwise (see above; Article 31.1), the affiliated entity concerned must make any such request directly to the beneficiary that owns the results.

Requests for access may be made — unless agreed otherwise — up to one year after the period set out in Article 3.

### **31.5 Access rights for the EU institutions, bodies, offices or agencies and EU Member States**

The beneficiaries must give access to their results — on a royalty-free basis — to EU institutions, bodies, offices or agencies, for developing, implementing or monitoring EU policies or programmes.

Such access rights are limited to non-commercial and non-competitive use.

This does not change the right to use any material, document or information received from the beneficiaries for communication and publicising activities (see Article 38.2).

### **31.6 Access rights for third parties**

The access provider must give the users royalty-free access to the results needed to implement the action.

### **31.7 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

## **SECTION 4 OTHER RIGHTS AND OBLIGATIONS**

### **ARTICLE 32 — RECRUITMENT AND WORKING CONDITIONS FOR RESEARCHERS**

#### **32.1 Obligation to take measures to implement the European Charter for Researchers and Code of Conduct for the Recruitment of Researchers**

The beneficiaries must take all measures to implement the principles set out in the Commission Recommendation on the European Charter for Researchers and the Code of Conduct for the Recruitment of Researchers<sup>21</sup>, in particular regarding:

- working conditions;
- transparent recruitment processes based on merit, and
- career development.

The beneficiaries must ensure that researchers and third parties involved in the action are aware of them.

#### **32.2 Consequences of non-compliance**

---

<sup>21</sup> Commission Recommendation 2005/251/EC of 11 March 2005 on the European Charter for Researchers and on a Code of Conduct for the Recruitment of Researchers (OJ L 75, 22.3.2005, p. 67).

If a beneficiary breaches its obligations under this Article, the Commission may apply any of the measures described in Chapter 6.

## **ARTICLE 33 — GENDER EQUALITY**

### **33.1 Obligation to aim for gender equality**

The beneficiaries must take all measures to promote equal opportunities between men and women in the implementation of the action. They must aim, to the extent possible, for a gender balance at all levels of personnel assigned to the action, including at supervisory and managerial level.

### **33.2 Consequences of non-compliance**

If a beneficiary breaches its obligations under this Article, the Commission may apply any of the measures described in Chapter 6.

## **ARTICLE 34 — ETHICS AND RESEARCH INTEGRITY**

### **34.1 Obligation to comply with ethical and research integrity principles**

The beneficiaries must carry out the action in compliance with:

(a) ethical principles (including the highest standards of research integrity)

and

(b) applicable international, EU and national law.

Funding will not be granted for activities carried out outside the EU if they are prohibited in all Member States or for activities which destroy human embryos (for example, for obtaining stem cells).

The beneficiaries must ensure that the activities under the action have an exclusive focus on civil applications.

The beneficiaries must ensure that the activities under the action do not:

(a) aim at human cloning for reproductive purposes;

(b) intend to modify the genetic heritage of human beings which could make such changes heritable (with the exception of research relating to cancer treatment of the gonads, which may be financed), or

(c) intend to create human embryos solely for the purpose of research or for the purpose of stem cell procurement, including by means of somatic cell nuclear transfer.

The beneficiaries must respect the highest standards of research integrity — as set out, for instance, in the European Code of Conduct for Research Integrity<sup>22</sup>.

---

<sup>22</sup> The European Code of Conduct for Research Integrity of ALLEA (All European Academies) and ESF (European Science Foundation) of March 2011.

[http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics\\_code-of-conduct\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf)

This implies notably compliance with the following essential principles:

- honesty;
- reliability;
- objectivity;
- impartiality;
- open communication;
- duty of care;
- fairness and
- responsibility for future science generations.

This means that beneficiaries must ensure that persons carrying out research tasks:

- present their research goals and intentions in an honest and transparent manner;
- design their research carefully and conduct it in a reliable fashion, taking its impact on society into account;
- use techniques and methodologies (including for data collection and management) that are appropriate for the field(s) concerned;
- exercise due care for the subjects of research — be they human beings, animals, the environment or cultural objects;
- ensure objectivity, accuracy and impartiality when disseminating the results;
- allow — in addition to the open access obligations under Article 29.3 as much as possible and taking into account the legitimate interest of the beneficiaries — access to research data, in order to enable research to be reproduced;
- make the necessary references to their work and that of other researchers;
- refrain from practicing any form of plagiarism, data falsification or fabrication;
- avoid double funding, conflicts of interest and misrepresentation of credentials or other research misconduct.

### **34.2 Activities raising ethical issues**

Activities raising ethical issues must comply with the ‘**ethics requirements**’ set out as deliverables in Annex 1.

Before the beginning of an activity raising an ethical issue, each beneficiary must have obtained:

- (a) any ethics committee opinion required under national law and



(b) any notification or authorisation for activities raising ethical issues required under national and/or European law

needed for implementing the action tasks in question.

The documents must be kept on file and be submitted upon request by the coordinator to the Commission (see Article 52). If they are not in English, they must be submitted together with an English summary, which shows that the action tasks in question are covered and includes the conclusions of the committee or authority concerned (if available).

### **34.3 Activities involving human embryos or human embryonic stem cells**

Activities involving research on human embryos or human embryonic stem cells may be carried out, in addition to Article 34.1, only if:

- they are set out in Annex 1 or
- the coordinator has obtained explicit approval (in writing) from the Commission (see Article 52).

### **34.4 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43) and the Agreement or participation of the beneficiary may be terminated (see Article 50).

Such breaches may also lead to any of the other measures described in Chapter 6.

## **ARTICLE 35 — CONFLICT OF INTERESTS**

### **35.1 Obligation to avoid a conflict of interests**

The beneficiaries must take all measures to prevent any situation where the impartial and objective implementation of the action is compromised for reasons involving economic interest, political or national affinity, family or emotional ties or any other shared interest (**'conflict of interests'**).

They must formally notify to the Commission without delay any situation constituting or likely to lead to a conflict of interests and immediately take all the necessary steps to rectify this situation.

The Commission may verify that the measures taken are appropriate and may require additional measures to be taken by a specified deadline.

### **35.2 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43) and the Agreement or participation of the beneficiary may be terminated (see Article 50).

Such breaches may also lead to any of the other measures described in Chapter 6.

## **ARTICLE 36 — CONFIDENTIALITY**

### **36.1 General obligation to maintain confidentiality**

During implementation of the action and for four years after the period set out in Article 3, the parties must keep confidential any data, documents or other material (in any form) that is identified as confidential at the time it is disclosed (**‘confidential information’**).

If a beneficiary requests, the Commission may agree to keep such information confidential for an additional period beyond the initial four years.

If information has been identified as confidential only orally, it will be considered to be confidential only if this is confirmed in writing within 15 days of the oral disclosure.

Unless otherwise agreed between the parties, they may use confidential information only to implement the Agreement.

The beneficiaries may disclose confidential information to their personnel or third parties involved in the action only if they:

- (a) need to know to implement the Agreement and
- (b) are bound by an obligation of confidentiality.

This does not change the security obligations in Article 37, which still apply.

The Commission may disclose confidential information to its staff, other EU institutions and bodies. It may disclose confidential information to third parties, if:

- (a) this is necessary to implement the Agreement or safeguard the EU's financial interests and
- (b) the recipients of the information are bound by an obligation of confidentiality.

Under the conditions set out in Article 4 of the Rules for Participation Regulation No 1290/2013<sup>23</sup>, the Commission must moreover make available information on the results to other EU institutions, bodies, offices or agencies as well as Member States or associated countries.

The confidentiality obligations no longer apply if:

- (a) the disclosing party agrees to release the other party;
- (b) the information was already known by the recipient or is given to him without obligation of confidentiality by a third party that was not bound by any obligation of confidentiality;
- (c) the recipient proves that the information was developed without the use of confidential information;
- (d) the information becomes generally and publicly available, without breaching any confidentiality obligation, or
- (e) the disclosure of the information is required by EU or national law.

## 36.2 Consequences of non-compliance

---

<sup>23</sup> Regulation (EU) No 1290/2013 of the European Parliament and of the Council of 11 December 2013 laying down the rules for participation and dissemination in "Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020)" (OJ L 347, 20.12.2013 p.81).

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

## **ARTICLE 37 — SECURITY-RELATED OBLIGATIONS**

### **37.1 Results with a security recommendation**

Not applicable

### **37.2 Classified information**

Not applicable

### **37.3 Activities involving dual-use goods or dangerous materials and substances**

Not applicable

### **37.4 Consequences of non-compliance**

Not applicable

## **ARTICLE 38 — PROMOTING THE ACTION — VISIBILITY OF EU FUNDING**

### **38.1 Communication activities by beneficiaries**

#### **38.1.1 Obligation to promote the action and its results**

The beneficiaries must promote the action and its results, by providing targeted information to multiple audiences (including the media and the public) in a strategic and effective manner.

This does not change the dissemination obligations in Article 29, the confidentiality obligations in Article 36 or the security obligations in Article 37, all of which still apply.

Before engaging in a communication activity expected to have a major media impact, the beneficiaries must inform the Commission (see Article 52).

#### **38.1.2 Information on EU funding — Obligation and right to use the EU emblem**

Unless the Commission requests or agrees otherwise or unless it is impossible, any communication activity related to the action (including in electronic form, via social media, etc.) and any infrastructure, equipment and major results funded by the grant must:

- (a) display the EU emblem and
- (b) include the following text:

For communication activities:

“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 731015”.

For infrastructure, equipment and major results:

“This [infrastructure][equipment][insert type of result] is part of a project that has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 731015”.

When displayed together with another logo, the EU emblem must have appropriate prominence.

For the purposes of their obligations under this Article, the beneficiaries may use the EU emblem without first obtaining approval from the Commission.

This does not, however, give them the right to exclusive use.

Moreover, they may not appropriate the EU emblem or any similar trademark or logo, either by registration or by any other means.

### **38.1.3 Disclaimer excluding Commission responsibility**

Any communication activity related to the action must indicate that it reflects only the author's view and that the Commission is not responsible for any use that may be made of the information it contains.

## **38.2 Communication activities by the Commission**

### **38.2.1 Right to use beneficiaries’ materials, documents or information**

The Commission may use, for its communication and publicising activities, information relating to the action, documents notably summaries for publication and public deliverables as well as any other material, such as pictures or audio-visual material received from any beneficiary (including in electronic form).

This does not change the confidentiality obligations in Article 36 and the security obligations in Article 37, all of which still apply.

If the Commission’s use of these materials, documents or information would risk compromising legitimate interests, the beneficiary concerned may request the Commission not to use it (see Article 52).

The right to use a beneficiary’s materials, documents and information includes:

- (a) **use for its own purposes** (in particular, making them available to persons working for the Commission or any other EU institution, body, office or agency or body or institutions in EU Member States; and copying or reproducing them in whole or in part, in unlimited numbers);
- (b) **distribution to the public** (in particular, publication as hard copies and in electronic or digital format, publication on the internet, as a downloadable or non-downloadable file, broadcasting by any channel, public display or presentation, communicating through press information services, or inclusion in widely accessible databases or indexes);
- (c) **editing or redrafting** for communication and publicising activities (including shortening, summarising, inserting other elements (such as meta-data, legends, other graphic, visual, audio or text elements), extracting parts (e.g. audio or video files), dividing into parts, use in a compilation);
- (d) **translation**;

- (e) giving **access in response to individual requests** under Regulation No 1049/2001<sup>25</sup>, without the right to reproduce or exploit;
- (f) **storage** in paper, electronic or other form;
- (g) **archiving**, in line with applicable document-management rules, and
- (h) the right to authorise **third parties** to act on its behalf or sub-license the modes of use set out in Points (b), (c), (d) and (f) to third parties if needed for the communication and publicising activities of the Commission.

If the right of use is subject to rights of a third party (including personnel of the beneficiary), the beneficiary must ensure that it complies with its obligations under this Agreement (in particular, by obtaining the necessary approval from the third parties concerned).

Where applicable (and if provided by the beneficiaries), the Commission will insert the following information:

“© – [year] – [name of the copyright owner]. All rights reserved. Licensed to the European Union (EU) under conditions.”

### 38.3 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

## ARTICLE 39 — PROCESSING OF PERSONAL DATA

### 39.1 Processing of personal data by the Commission

Any personal data under the Agreement will be processed by the Commission under Regulation No 45/2001<sup>26</sup> and according to the ‘notifications of the processing operations’ to the Data Protection Officer (DPO) of the Commission (publicly accessible in the DPO register).

Such data will be processed by the ‘**data controller**’ of the Commission for the purposes of implementing, managing and monitoring the Agreement or protecting the financial interests of the EU or Euratom (including checks, reviews, audits and investigations; see Article 22).

The persons whose personal data are processed have the right to access and correct their own personal data. For this purpose, they must send any queries about the processing of their personal data to the data controller, via the contact point indicated in the privacy statement(s) that are published on the Commission websites.

<sup>25</sup> Regulation (EC) No 1049/2001 of the European Parliament and of the Council of 30 May 2001 regarding public access to European Parliament, Council and Commission documents, OJ L 145, 31.5.2001, p. 43.

<sup>26</sup> Regulation (EC) No 45/2001 of the European Parliament and of the Council of 18 December 2000 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data (OJ L 8, 12.01.2001, p. 1).

They also have the right to have recourse at any time to the European Data Protection Supervisor (EDPS).

### **39.2 Processing of personal data by the beneficiaries**

The beneficiaries must process personal data under the Agreement in compliance with applicable EU and national law on data protection (including authorisations or notification requirements).

The beneficiaries may grant their personnel access only to data that is strictly necessary for implementing, managing and monitoring the Agreement.

The beneficiaries must inform the personnel whose personal data are collected and processed by the Commission. For this purpose, they must provide them with the privacy statement(s) (see above), before transmitting their data to the Commission.

### **39.3 Consequences of non-compliance**

If a beneficiary breaches any of its obligations under Article 39.2, the Commission may apply any of the measures described in Chapter 6.

## **ARTICLE 40 — ASSIGNMENTS OF CLAIMS FOR PAYMENT AGAINST THE COMMISSION**

The beneficiaries may not assign any of their claims for payment against the Commission to any third party, except if approved by the Commission on the basis of a reasoned, written request by the coordinator (on behalf of the beneficiary concerned).

If the Commission has not accepted the assignment or the terms of it are not observed, the assignment will have no effect on it.

In no circumstances will an assignment release the beneficiaries from their obligations towards the Commission.

## **CHAPTER 5 DIVISION OF BENEFICIARIES' ROLES AND RESPONSIBILITIES** **— RELATIONSHIP WITH COMPLEMENTARY BENEFICIARIES —** **RELATIONSHIP WITH PARTNERS OF A JOINT ACTION**

### **ARTICLE 41 — DIVISION OF BENEFICIARIES' ROLES AND RESPONSIBILITIES** **— RELATIONSHIP WITH COMPLEMENTARY BENEFICIARIES —** **RELATIONSHIP WITH PARTNERS OF A JOINT ACTION**

#### **41.1 Roles and responsibility towards the Commission**

The beneficiaries have full responsibility for implementing the action and complying with the Agreement.

The beneficiaries are jointly and severally liable for the **technical implementation** of the action as described in Annex 1. If a beneficiary fails to implement its part of the action, the other beneficiaries become responsible for implementing this part (without being entitled to any additional EU funding for doing so), unless the Commission expressly relieves them of this obligation.

The **financial responsibility** of each beneficiary is governed by Articles 44, 45 and 46.

#### 41.2 Internal division of roles and responsibilities

The internal roles and responsibilities of the beneficiaries are divided as follows:

(a) Each **beneficiary** must:

- (i) keep information stored in the Participant Portal Beneficiary Register (via the electronic exchange system) up to date (see Article 17);
- (ii) inform the coordinator immediately of any events or circumstances likely to affect significantly or delay the implementation of the action (see Article 17);
- (iii) submit to the coordinator in good time:
  - individual financial statements for itself and its linked third parties and, if required, certificates on the financial statements (see Article 20);
  - the data needed to draw up the technical reports (see Article 20);
  - ethics committee opinions and notifications or authorisations for activities raising ethical issues (see Article 34);
  - any other documents or information required by the Commission under the Agreement, unless the Agreement requires the beneficiary to submit this information directly to the Commission.

(b) The **coordinator** must:

- (i) monitor that the action is implemented properly (see Article 7);
- (ii) act as the intermediary for all communications between the beneficiaries and the Commission (in particular, providing the Commission with the information described in Article 17), unless the Agreement specifies otherwise;
- (iii) request and review any documents or information required by the Commission and verify their completeness and correctness before passing them on to the Commission;
- (iv) submit the deliverables and reports to the Commission (see Articles 19 and 20);
- (v) ensure that all payments are made to the other beneficiaries without unjustified delay (see Article 21);
- (vi) inform the Commission of the amounts paid to each beneficiary, when required under the Agreement (see Articles 44 and 50) or requested by the Commission.

The coordinator may not delegate or subcontract the above-mentioned tasks to any other beneficiary or third party (including linked third parties).

#### 41.3 Internal arrangements between beneficiaries — Consortium agreement

The beneficiaries must have internal arrangements regarding their operation and co-ordination to

ensure that the action is implemented properly. These internal arrangements must be set out in a written ‘**consortium agreement**’ between the beneficiaries, which may cover:

- internal organisation of the consortium;
- management of access to the electronic exchange system;
- distribution of EU funding;
- additional rules on rights and obligations related to background and results (including whether access rights remain or not, if a beneficiary is in breach of its obligations) (see Section 3 of Chapter 4);
- settlement of internal disputes;
- liability, indemnification and confidentiality arrangements between the beneficiaries.

The consortium agreement must not contain any provision contrary to the Agreement.

#### **41.4 Relationship with complementary beneficiaries — Collaboration agreement**

Not applicable

#### **41.5 Relationship with partners of a joint action — Coordination agreement**

Not applicable

## **CHAPTER 6 REJECTION OF COSTS — REDUCTION OF THE GRANT — RECOVERY — SANCTIONS — DAMAGES — SUSPENSION — TERMINATION — FORCE MAJEURE**

### **SECTION 1 REJECTION OF COSTS — REDUCTION OF THE GRANT — RECOVERY — SANCTIONS**

#### **ARTICLE 42 — REJECTION OF INELIGIBLE COSTS**

##### **42.1 Conditions**

The Commission will — after **termination of the participation of a beneficiary**, at the time of an **interim payment, at the payment of the balance or afterwards** — reject any costs which are ineligible (see Article 6), in particular following checks, reviews, audits or investigations (see Article 22).

The rejection may also be based on the **extension of findings from other grants to this grant** (see Article 22.5.2).

##### **42.2 Ineligible costs to be rejected — Calculation — Procedure**

Ineligible costs will be rejected in full.



If the rejection of costs does not lead to a recovery (see Article 44), the Commission will formally notify the coordinator or beneficiary concerned of the rejection of costs, the amounts and the reasons why (if applicable, together with the notification of amounts due; see Article 21.5). The coordinator or beneficiary concerned may — within 30 days of receiving notification — formally notify the Commission of its disagreement and the reasons why.

If the rejection of costs leads to a recovery, the Commission will follow the contradictory procedure with pre-information letter set out in Article 44.

### 42.3 Effects

If the Commission rejects costs at the time of an **interim payment or the payment of the balance**, it will deduct them from the total eligible costs declared, for the action, in the periodic or final summary financial statement (see Articles 20.3 and 20.4). It will then calculate the interim payment or payment of the balance as set out in Articles 21.3 or 21.4.

If the Commission rejects costs **after termination of the participation of a beneficiary**, it will deduct them from the costs declared by the beneficiary in the termination report and include the rejection in the calculation after termination (see Article 50.2 and 50.3).

If the Commission — **after an interim payment but before the payment of the balance** — rejects costs declared in a periodic summary financial statement, it will deduct them from the total eligible costs declared, for the action, in the next periodic summary financial statement or in the final summary financial statement. It will then calculate the interim payment or payment of the balance as set out in Articles 21.3 or 21.4.

If the Commission rejects costs **after the payment of the balance**, it will deduct the amount rejected from the total eligible costs declared, by the beneficiary, in the final summary financial statement. It will then calculate the revised final grant amount as set out in Article 5.4.

## ARTICLE 43 — REDUCTION OF THE GRANT

### 43.1 Conditions

The Commission may — **after termination of the participation of a beneficiary, at the payment of the balance or afterwards** — reduce the grant amount (see Article 5.1), if :

- (a) a beneficiary (or a natural person who has the power to represent or take decisions on its behalf) has committed:
  - (i) substantial errors, irregularities or fraud or
  - (ii) serious breach of obligations under the Agreement or during the award procedure (including improper implementation of the action, submission of false information, failure to provide required information, breach of ethical principles) or
- (b) a beneficiary (or a natural person who has the power to represent or take decision on its behalf) has committed — in other EU or Euratom grants awarded to it under similar conditions — systemic or recurrent errors, irregularities, fraud or serious breach of obligations that have a material impact on this grant (**extension of findings from other grants to this grant**; see Article 22.5.2).

### 43.2 Amount to be reduced — Calculation — Procedure

The amount of the reduction will be proportionate to the seriousness of the errors, irregularities or fraud or breach of obligations.

Before reduction of the grant, the Commission will formally notify a '**pre-information letter**' to the coordinator or beneficiary concerned:

- informing it of its intention to reduce the grant, the amount it intends to reduce and the reasons why and
- inviting it to submit observations within 30 days of receiving notification

If the Commission does not receive any observations or decides to pursue reduction despite the observations it has received, it will formally notify **confirmation** of the reduction (if applicable, together with the notification of amounts due; see Article 21).

### 43.3 Effects

If the Commission reduces the grant **after termination of the participation of a beneficiary**, it will calculate the reduced grant amount for that beneficiary and then determine the amount due to that beneficiary (see Article 50.2 and 50.3).

If the Commission reduces the grant **at the payment of the balance**, it will calculate the reduced grant amount for the action and then determine the amount due as payment of the balance (see Articles 5.3.4 and 21.4).

If the Commission reduces the grant **after the payment of the balance**, it will calculate the revised final grant amount for the beneficiary concerned (see Article 5.4). If the revised final grant amount for the beneficiary concerned is lower than its share of the final grant amount, the Commission will recover the difference (see Article 44).

## ARTICLE 44 — RECOVERY OF UNDUE AMOUNTS

### 44.1 Amount to be recovered — Calculation — Procedure

The Commission will — after **termination of the participation of a beneficiary, at the payment of the balance** or **afterwards** — claim back any amount that was paid, but is not due under the Agreement.

Each beneficiary's financial responsibility in case of recovery is limited to its own debt (including undue amounts paid by the Commission for costs declared by its linked third parties), except for the amount retained for the Guarantee Fund (see Article 21.4).

#### 44.1.1 Recovery after termination of a beneficiary's participation

If recovery takes place after termination of a beneficiary's participation (including the coordinator), the Commission will claim back the undue amount from the beneficiary concerned, by formally notifying it a debit note (see Article 50.2 and 50.3). This note will specify the amount to be recovered, the terms and the date for payment.

If payment is not made by the date specified in the debit note, the Commission will **recover** the amount:

- (a) by ‘**offsetting**’ it — without the beneficiary’s consent — against any amounts owed to the beneficiary concerned by the Commission or an executive agency (from the EU or Euratom budget).

In exceptional circumstances, to safeguard the EU’s financial interests, the Commission may offset before the payment date specified in the debit note;

- (b) not applicable;

- (c) by **taking legal action** (see Article 57) or by **adopting an enforceable decision** under Article 299 of the Treaty on the Functioning of the EU (TFEU) and Article 79(2) of the Financial regulation No 966/2012.

If payment is not made by the date specified in the debit note, the amount to be recovered (see above) will be increased by **late-payment interest** at the rate set out in Article 21.11, from the day following the payment date in the debit note, up to and including the date the Commission receives full payment of the amount.

Partial payments will be first credited against expenses, charges and late-payment interest and then against the principal.

Bank charges incurred in the recovery process will be borne by the beneficiary, unless Directive 2007/64/EC<sup>27</sup> applies.

#### 44.1.2 Recovery at payment of the balance

If the payment of the balance takes the form of a recovery (see Article 21.4), the Commission will formally notify a ‘**pre-information letter**’ to the coordinator:

- informing it of its intention to recover, the amount due as the balance and the reasons why;
- specifying that it intends to deduct the amount to be recovered from the amount retained for the Guarantee Fund;
- requesting the coordinator to submit a report on the distribution of payments to the beneficiaries within 30 days of receiving notification, and
- inviting the coordinator to submit observations within 30 days of receiving notification.

If no observations are submitted or the Commission decides to pursue recovery despite the observations it has received, it will **confirm recovery** (together with the notification of amounts due; see Article 21.5) and:

- pay the difference between the amount to be recovered and the amount retained for the Guarantee Fund, **if the difference is positive** or

---

<sup>27</sup> Directive 2007/64/EC of the European Parliament and of the Council of 13 November 2007 on payment services in the internal market amending Directives 97/7/EC, 2002/65/EC, 2005/60/EC and 2006/48/EC and repealing Directive 97/5/EC (OJ L 319, 05.12.2007, p. 1).

- formally notify to the coordinator a **debit note** for the difference between the amount to be recovered and the amount retained for the Guarantee Fund, **if the difference is negative**. This note will also specify the terms and the date for payment.

If the coordinator does not repay the Commission by the date in the debit note and has not submitted the report on the distribution of payments: the Commission will **recover** the amount set out in the debit note from the coordinator (see below).

If the coordinator does not repay the Commission by the date in the debit note, but has submitted the report on the distribution of payments: the Commission will:

- (a) identify the beneficiaries for which the amount calculated as follows is negative:

$\{ \{ \{ \text{beneficiary's costs declared in the final summary financial statement and approved by the Commission multiplied by the reimbursement rate set out in Article 5.2 for the beneficiary concerned}$

plus

$\{ \text{its linked third parties' costs declared in the final summary financial statement and approved by the Commission multiplied by the reimbursement rate set out in Article 5.2 for each linked third party concerned} \}$

divided by

$\{ \text{the EU contribution for the action calculated according to Article 5.3.1} \}$

multiplied by

$\{ \text{the final grant amount (see Article 5.3)} \}$ ,

minus

$\{ \text{pre-financing and interim payments received by the beneficiary} \}$ .

- (b) formally notify to each beneficiary identified according to point (a) a **debit note** specifying the terms and date for payment. The amount of the debit note is calculated as follows:

$\{ \{ \text{amount calculated according to point (a) for the beneficiary concerned}$

divided by

$\{ \text{the sum of the amounts calculated according to point (a) for all the beneficiaries identified according to point (a)} \}$

multiplied by

$\{ \text{the amount set out in the debit note formally notified to the coordinator} \}$ .

If payment is not made by the date specified in the debit note, the Commission will **recover** the amount:

- (a) by ‘**offsetting**’ it — without the beneficiary’s consent — against any amounts owed to the beneficiary concerned by the Commission or an executive agency (from the EU or Euratom budget).

In exceptional circumstances, to safeguard the EU’s financial interests, the Commission may offset before the payment date specified in the debit note;

(b) by **drawing on the Guarantee Fund**. The Commission will formally notify the beneficiary concerned the debit note on behalf of the Guarantee Fund and recover the amount:

- (i) not applicable;
- (ii) by **taking legal action** (see Article 57) or by **adopting an enforceable decision** under Article 299 of the Treaty on the Functioning of the EU (TFEU) and Article 79(2) of the Financial Regulation No 966/2012.

If payment is not made by the date in the debit note, the amount to be recovered (see above) will be increased by **late-payment interest** at the rate set out in Article 21.11, from the day following the payment date in the debit note, up to and including the date the Commission receives full payment of the amount.

Partial payments will be first credited against expenses, charges and late-payment interest and then against the principal.

Bank charges incurred in the recovery process will be borne by the beneficiary, unless Directive 2007/64/EC applies.

#### 44.1.3 Recovery of amounts after payment of the balance

If, for a beneficiary, the revised final grant amount (see Article 5.4) is lower than its share of the final grant amount, it must repay the difference to the Commission.

The beneficiary's share of the final grant amount is calculated as follows:

$$\left\{ \left\{ \begin{array}{l} \text{beneficiary's costs declared in the final summary financial statement and approved by the Commission} \\ \text{multiplied by the reimbursement rate set out in Article 5.2 for the beneficiary concerned} \end{array} \right\} \right.$$

plus

$$\left\{ \begin{array}{l} \text{its linked third parties' costs declared in the final summary financial statement and approved by the Commission} \\ \text{multiplied by the reimbursement rate set out in Article 5.2 for each linked third party concerned} \end{array} \right\}$$

divided by

$$\left\{ \begin{array}{l} \text{the EU contribution for the action calculated according to Article 5.3.1} \end{array} \right\}$$

multiplied by

$$\left\{ \begin{array}{l} \text{the final grant amount (see Article 5.3)} \end{array} \right\}.$$

If the coordinator has not distributed amounts received (see Article 21.7), the Commission will also recover these amounts.

The Commission will formally notify a **pre-information letter** to the beneficiary concerned:

- informing it of its intention to recover, the due amount and the reasons why and
- inviting it to submit observations within 30 days of receiving notification.

If no observations are submitted or the Commission decides to pursue recovery despite the observations it has received, it will **confirm** the amount to be recovered and formally notify to the beneficiary concerned a **debit note**. This note will also specify the terms and the date for payment.

If payment is not made by the date specified in the debit note, the Commission will **recover** the amount:

- (a) by **‘offsetting’** it — without the beneficiary’s consent — against any amounts owed to the beneficiary concerned by the Commission or an executive agency (from the EU or Euratom budget).

In exceptional circumstances, to safeguard the EU’s financial interests, the Commission may offset before the payment date specified in the debit note;

- (b) by **drawing on the Guarantee Fund**. The Commission will formally notify the beneficiary concerned the debit note on behalf of the Guarantee Fund and recover the amount:
  - (i) not applicable;
  - (ii) by **taking legal action** (see Article 57) or by **adopting an enforceable decision** under Article 299 of the Treaty on the Functioning of the EU (TFEU) and Article 79(2) of the Financial Regulation No 966/2012.

If payment is not made by the date in the debit note, the amount to be recovered (see above) will be increased by **late-payment interest** at the rate set out in Article 21.11, from the day following the date for payment in the debit note, up to and including the date the Commission receives full payment of the amount.

Partial payments will be first credited against expenses, charges and late-payment interest and then against the principal.

Bank charges incurred in the recovery process will be borne by the beneficiary, unless Directive 2007/64/EC applies.

## **ARTICLE 45 — ADMINISTRATIVE SANCTIONS**

In addition to contractual measures, the Commission may also adopt administrative sanctions under Articles 106 and 131(4) of the Financial Regulation No 966/2012 (i.e. exclusion from future procurement contracts, grants and expert contracts and/or financial penalties).

## **SECTION 2 LIABILITY FOR DAMAGES**

### **ARTICLE 46 — LIABILITY FOR DAMAGES**

#### **46.1 Liability of the Commission**

The Commission cannot be held liable for any damage caused to the beneficiaries or to third parties as a consequence of implementing the Agreement, including for gross negligence.

The Commission cannot be held liable for any damage caused by any of the beneficiaries or third parties involved in the action, as a consequence of implementing the Agreement.

#### **46.2 Liability of the beneficiaries**

Except in case of force majeure (see Article 51), the beneficiaries must compensate the Commission

for any damage it sustains as a result of the implementation of the action or because the action was not implemented in full compliance with the Agreement.

### **SECTION 3 SUSPENSION AND TERMINATION**

#### **ARTICLE 47 — SUSPENSION OF PAYMENT DEADLINE**

##### **47.1 Conditions**

The Commission may — at any moment — suspend the payment deadline (see Article 21.2 to 21.4) if a request for payment (see Article 20) cannot be approved because:

- (a) it does not comply with the provisions of the Agreement (see Article 20);
- (b) the technical or financial reports have not been submitted or are not complete or additional information is needed, or
- (c) there is doubt about the eligibility of the costs declared in the financial statements and additional checks, reviews, audits or investigations are necessary.

##### **47.2 Procedure**

The Commission will formally notify the coordinator of the suspension and the reasons why.

The suspension will **take effect** the day notification is sent by the Commission (see Article 52).

If the conditions for suspending the payment deadline are no longer met, the suspension will be **lifted** — and the remaining period will resume.

If the suspension exceeds two months, the coordinator may request the Commission if the suspension will continue.

If the payment deadline has been suspended due to the non-compliance of the technical or financial reports (see Article 20) and the revised report or statement is not submitted or was submitted but is also rejected, the Commission may also terminate the Agreement or the participation of the beneficiary (see Article 50.3.1(l)).

#### **ARTICLE 48 — SUSPENSION OF PAYMENTS**

##### **48.1 Conditions**

The Commission may — at any moment — suspend payments, in whole or in part and for one or more beneficiaries, if:

- (a) a beneficiary (or a natural person who has the power to represent or take decision on its behalf) has committed or is suspected of having committed:
  - (i) substantial errors, irregularities or fraud or
  - (ii) serious breach of obligations under the Agreement or during the award procedure



(including improper implementation of the action, submission of false information, failure to provide required information, breach of ethical principles) or

- (b) a beneficiary (or a natural person who has the power to represent or take decision on its behalf) has committed — in other EU or Euratom grants awarded to it under similar conditions — systemic or recurrent errors, irregularities, fraud or serious breach of obligations that have a material impact on this grant (**extension of findings from other grants to this grant**; see Article 22.5.2).

If payments are suspended for one or more beneficiaries, the Commission will make partial payment(s) for the part(s) not suspended. If suspension concerns the payment of the balance, — once suspension is lifted — the payment or the recovery of the amount(s) concerned will be considered the payment of the balance that closes the action.

## 48.2 Procedure

Before suspending payments, the Commission will formally notify the coordinator or beneficiary concerned:

- informing it of its intention to suspend payments and the reasons why and
- inviting it to submit observations within 30 days of receiving notification.

If the Commission does not receive observations or decides to pursue the procedure despite the observations it has received, it will formally notify **confirmation** of the suspension. Otherwise, it will formally notify that the suspension procedure is not continued.

The suspension will **take effect** the day the confirmation notification is sent by the Commission.

If the conditions for resuming payments are met, the suspension will be **lifted**. The Commission will formally notify the coordinator or beneficiary concerned.

During the suspension, the periodic report(s) for all reporting periods except the last one (see Article 20.3), must not contain any individual financial statements from the beneficiary concerned and its linked third parties. The coordinator must include them in the next periodic report after the suspension is lifted or — if suspension is not lifted before the end of the action — in the last periodic report.

The beneficiaries may suspend implementation of the action (see Article 49.1) or terminate the Agreement or the participation of the beneficiary concerned (see Article 50.1 and 50.2).

## ARTICLE 49 — SUSPENSION OF THE ACTION IMPLEMENTATION

### 49.1 Suspension of the action implementation, by the beneficiaries

#### 49.1.1 Conditions

The beneficiaries may suspend implementation of the action or any part of it, if exceptional circumstances — in particular *force majeure* (see Article 51) — make implementation impossible or excessively difficult.



### 49.1.2 Procedure

The coordinator must immediately formally notify to the Commission the suspension (see Article 52), stating:

- the reasons why and
- the expected date of resumption.

The suspension will **take effect** the day this notification is received by the Commission.

Once circumstances allow for implementation to resume, the coordinator must immediately formally notify the Commission and request an **amendment** of the Agreement to set the date on which the action will be resumed, extend the duration of the action and make other changes necessary to adapt the action to the new situation (see Article 55) — unless the Agreement or the participation of a beneficiary has been terminated (see Article 50).

The suspension will be **lifted** with effect from the resumption date set out in the amendment. This date may be before the date on which the amendment enters into force.

Costs incurred during suspension of the action implementation are not eligible (see Article 6).

## 49.2 Suspension of the action implementation, by the Commission

### 49.2.1 Conditions

The Commission may suspend implementation of the action or any part of it, if:

- (a) a beneficiary (or a natural person who has the power to represent or take decisions on its behalf) has committed or is suspected of having committed:
  - (i) substantial errors, irregularities or fraud or
  - (ii) serious breach of obligations under the Agreement or during the award procedure (including improper implementation of the action, submission of false declaration, failure to provide required information, breach of ethical principles);
- (b) a beneficiary (or a natural person who has the power to represent or take decisions on its behalf) has committed — in other EU or Euratom grants awarded to it under similar conditions — systemic or recurrent errors, irregularities, fraud or serious breach of obligations that have a material impact on this grant (**extension of findings from other grants to this grant**; see Article 22.5.2), or
- (c) the action is suspected of having lost its scientific or technological relevance.

### 49.2.2 Procedure

Before suspending implementation of the action, the Commission will formally notify the coordinator or beneficiary concerned:

- informing it of its intention to suspend the implementation and the reasons why and
- inviting it to submit observations within 30 days of receiving notification.

If the Commission does not receive observations or decides to pursue the procedure despite the observations it has received, it will formally notify **confirmation** of the suspension. Otherwise, it will formally notify that the procedure is not continued.

The suspension will **take effect** five days after confirmation notification is received (or on a later date specified in the notification).

It will be **lifted** if the conditions for resuming implementation of the action are met.

The coordinator or beneficiary concerned will be formally notified of the lifting and the Agreement will be **amended** to set the date on which the action will be resumed, extend the duration of the action and make other changes necessary to adapt the action to the new situation (see Article 55) — unless the Agreement has already been terminated (see Article 50).

The suspension will be lifted with effect from the resumption date set out in the amendment. This date may be before the date on which the amendment enters into force.

Costs incurred during suspension are not eligible (see Article 6).

The beneficiaries may not claim damages due to suspension by the Commission (see Article 46).

Suspension of the action implementation does not affect the Commission's right to terminate the Agreement or participation of a beneficiary (see Article 50), reduce the grant or recover amounts unduly paid (see Articles 43 and 44).

## **ARTICLE 50 — TERMINATION OF THE AGREEMENT OR OF THE PARTICIPATION OF ONE OR MORE BENEFICIARIES**

### **50.1 Termination of the Agreement, by the beneficiaries**

#### **50.1.1 Conditions and procedure**

The beneficiaries may terminate the Agreement.

The coordinator must formally notify termination to the Commission (see Article 52), stating:

- the reasons why and
- the date the termination will take effect. This date must be after the notification.

If no reasons are given or if the Commission considers the reasons do not justify termination, the Agreement will be considered to have been '**terminated improperly**'.

The termination will **take effect** on the day specified in the notification.

#### **50.1.2 Effects**

The coordinator must — within 60 days from when termination takes effect — submit:

- (i) a periodic report (for the open reporting period until termination; see Article 20.3) and
- (ii) the final report (see Article 20.4).

If the Commission does not receive the reports within the deadline (see above), only costs which are included in an approved periodic report will be taken into account.

The Commission will **calculate** the final grant amount (see Article 5.3) and the balance (see Article 21.4) on the basis of the reports submitted. Only costs incurred until termination are eligible (see Article 6). Costs relating to contracts due for execution only after termination are not eligible.

Improper termination may lead to a reduction of the grant (see Article 43).

After termination, the beneficiaries' obligations (in particular Articles 20, 22, 23, Section 3 of Chapter 4, 36, 37, 38, 40, 42, 43 and 44) continue to apply.

## **50.2 Termination of the participation of one or more beneficiaries, by the beneficiaries**

### **50.2.1 Conditions and procedure**

The participation of one or more beneficiaries may be terminated by the coordinator, on request of the beneficiary concerned or on behalf of the other beneficiaries.

The coordinator must formally notify termination to the Commission (see Article 52) and inform the beneficiary concerned.

If the coordinator's participation is terminated without its agreement, the formal notification must be done by another beneficiary (acting on behalf of the other beneficiaries).

The notification must include:

- the reasons why;
- the opinion of the beneficiary concerned (or proof that this opinion has been requested in writing);
- the date the termination takes effect. This date must be after the notification, and
- a request for amendment (see Article 55), with a proposal for reallocation of the tasks and the estimated budget of the beneficiary concerned (see Annexes 1 and 2) and, if necessary, the addition of one or more new beneficiaries (see Article 56). If termination takes effect after the period set out in Article 3, no request for amendment must be included unless the beneficiary concerned is the coordinator. In this case, the request for amendment must propose a new coordinator.

If this information is not given or if the Commission considers that the reasons do not justify termination, the participation will be considered to have been **terminated improperly**.

The termination will **take effect** on the day specified in the notification.

### **50.2.2 Effects**

The coordinator must — within 30 days from when termination takes effect — submit:

- (i) a report on the distribution of payments to the beneficiary concerned and
- (ii) if termination takes effect during the period set out in Article 3, a '**termination report**'

from the beneficiary concerned, for the open reporting period until termination, containing an overview of the progress of the work, an overview of the use of resources, the individual financial statement and, if applicable, the certificate on the financial statement (see Articles 20.3 and 20.4).

The information in the termination report must also be included in the periodic report for the next reporting period (see Article 20.3).

If the request for amendment is rejected by the Commission, (because it calls into question the decision awarding the grant or breaches the principle of equal treatment of applicants), the Agreement may be terminated according to Article 50.3.1(c).

If the request for amendment is accepted by the Commission, the Agreement is **amended** to introduce the necessary changes (see Article 55).

The Commission will calculate — on the basis of the periodic reports, the termination report and the report on the distribution of payments — **calculate** the amount which is due to the beneficiary and if the (pre-financing and interim) payments received by the beneficiary exceed this amount.

The **amount which is due** is calculated in the following steps:

Step 1 — Application of the reimbursement rate to the eligible costs

The grant amount for the beneficiary is calculated by applying the reimbursement rate(s) to the total eligible costs declared by the beneficiary and its linked third parties in the termination report and approved by the Commission.

Only costs incurred by the beneficiary concerned until termination takes effect are eligible (see Article 6). Costs relating to contracts due for execution only after termination are not eligible.

Step 2 — Reduction due to substantial errors, irregularities or fraud or serious breach of obligations

In case of a reduction (see Article 43), the Commission will calculate the reduced grant amount for the beneficiary by deducting the amount of the reduction (calculated in proportion to the seriousness of the errors, irregularities or fraud or breach of obligations, in accordance with Article 43.2) from the grant amount for the beneficiary.

If the payments received **exceed the amounts due**:

- if termination takes effect during the period set out in Article 3 and the request for amendment is accepted, the beneficiary concerned must repay to the coordinator the amount unduly received. The Commission will formally notify the amount unduly received and request the beneficiary concerned to repay it to the coordinator within 30 days of receiving notification. If it does not repay the coordinator, the Commission will draw upon the Guarantee Fund to pay the coordinator and then notify a **debit note** on behalf of the Guarantee Fund to the beneficiary concerned (see Article 44);
- in all other cases, in particular if termination takes effect after the period set out in Article 3, the Commission will formally notify a **debit note** to the beneficiary concerned. If payment

is not made by the date in the debit note, the Guarantee Fund will pay to the Commission the amount due and the Commission will notify a debit note on behalf of the Guarantee Fund to the beneficiary concerned (see Article 44);

- if the beneficiary concerned is the former coordinator, it must repay the new coordinator according to the procedure above, unless:
  - termination takes effect after an interim payment and
  - the former coordinator has not distributed amounts received as pre-financing or interim payments (see Article 21.7).

In this case, the Commission will formally notify a **debit note** to the former coordinator. If payment is not made by the date in the debit note, the Guarantee Fund will pay to the Commission the amount due. The Commission will then pay the new coordinator and notify a debit note on behalf of the Guarantee Fund to the former coordinator (see Article 44).

If the payments received **do not exceed the amounts due**: amounts owed to the beneficiary concerned will be included in the next interim or final payment.

If the Commission does not receive the termination report within the deadline (see above), only costs included in an approved periodic report will be taken into account.

If the Commission does not receive the report on the distribution of payments within the deadline (see above), it will consider that:

- the coordinator did not distribute any payment to the beneficiary concerned and that
- the beneficiary concerned must not repay any amount to the coordinator.

Improper termination may lead to a reduction of the grant (see Article 43) or termination of the Agreement (see Article 50).

After termination, the concerned beneficiary's obligations (in particular Articles 20, 22, 23, Section 3 of Chapter 4, 36, 37, 38, 40, 42, 43 and 44) continue to apply.

### **50.3 Termination of the Agreement or the participation of one or more beneficiaries, by the Commission**

#### **50.3.1 Conditions**

The Commission may terminate the Agreement or the participation of one or more beneficiaries, if:

- (a) one or more beneficiaries do not accede to the Agreement (see Article 56);
- (b) a change to their legal, financial, technical, organisational or ownership situation (or those of its linked third parties) is likely to substantially affect or delay the implementation of the action or calls into question the decision to award the grant;
- (c) following termination of participation for one or more beneficiaries (see above), the necessary changes to the Agreement would call into question the decision awarding the grant or breach the principle of equal treatment of applicants (see Article 55);

- (d) implementation of the action is prevented by force majeure (see Article 51) or suspended by the coordinator (see Article 49.1) and either:
  - (i) resumption is impossible, or
  - (ii) the necessary changes to the Agreement would call into question the decision awarding the grant or breach the principle of equal treatment of applicants;
- (e) a beneficiary is declared bankrupt, being wound up, having its affairs administered by the courts, has entered into an arrangement with creditors, has suspended business activities, or is subject to any other similar proceedings or procedures under national law;
- (f) a beneficiary (or a natural person who has the power to represent or take decisions on its behalf) has been found guilty of professional misconduct, proven by any means;
- (g) a beneficiary does not comply with the applicable national law on taxes and social security;
- (h) the action has lost scientific or technological relevance;
- (i) not applicable;
- (j) not applicable;
- (k) a beneficiary (or a natural person who has the power to represent or take decisions on its behalf) has committed fraud, corruption, or is involved in a criminal organisation, money laundering or any other illegal activity;
- (l) a beneficiary (or a natural person who has the power to represent or take decisions on its behalf) has committed:
  - (i) substantial errors, irregularities or fraud or
  - (ii) serious breach of obligations under the Agreement or during the award procedure (including improper implementation of the action, submission of false information, failure to provide required information, breach of ethical principles);
- (m) a beneficiary (or a natural person who has the power to represent or take decisions on its behalf) has committed — in other EU or Euratom grants awarded to it under similar conditions — systemic or recurrent errors, irregularities, fraud or serious breach of obligations that have a material impact on this grant (**extension of findings from other grants to this grant**; see Article 22.5.2).
- (n) despite a specific request by the Commission, a beneficiary does not request — through the coordinator — an amendment to the Agreement to end the participation of one of its linked third parties that is in one of the situations under points (e), (f), (g), (k), (l) or (m) and to reallocate its tasks.

### 50.3.2 Procedure

Before terminating the Agreement or participation of one or more beneficiaries, the Commission will formally notify the coordinator or beneficiary concerned:

- informing it of its intention to terminate and the reasons why and

- inviting it, within 30 days of receiving notification, to submit observations and — in case of Point (l.ii) above — to inform the Commission of the measures to ensure compliance with the obligations under the Agreement.

If the Commission does not receive observations or decides to pursue the procedure despite the observations it has received, it will formally notify to the coordinator or beneficiary concerned **confirmation** of the termination and the date it will take effect. Otherwise, it will formally notify that the procedure is not continued.

The termination will **take effect**:

- for terminations under Points (b), (c), (e), (g), (h), (j), (l.ii) and (n) above: on the day specified in the notification of the confirmation (see above);
- for terminations under Points (a), (d), (f), (i), (k), (l.i) and (m) above: on the day after the notification of the confirmation is received.

### 50.3.3 Effects

#### (a) for **termination of the Agreement**:

The coordinator must — within 60 days from when termination takes effect — submit:

- (i) a periodic report (for the last open reporting period until termination; see Article 20.3) and
- (ii) a final report (see Article 20.4).

If the Agreement is terminated for breach of the obligation to submit reports (see Articles 20.8 and 50.3.1(l)), the coordinator may not submit any reports after termination.

If the Commission does not receive the reports within the deadline (see above), only costs which are included in an approved periodic report will be taken into account.

The Commission will **calculate** the final grant amount (see Article 5.3) and the balance (see Article 21.4) on the basis of the reports submitted. Only costs incurred until termination takes effect are eligible (see Article 6). Costs relating to contracts due for execution only after termination are not eligible.

This does not affect the Commission's right to reduce the grant (see Article 43) or to impose administrative sanctions (Article 45).

The beneficiaries may not claim damages due to termination by the Commission (see Article 46).

After termination, the beneficiaries' obligations (in particular Articles 20, 22, 23, Section 3 of Chapter 4, 36, 37, 38, 40, 42, 43 and 44) continue to apply.

#### (b) for **termination of the participation of one or more beneficiaries**:

The coordinator must — within 60 days from when termination takes effect — submit:

- (i) a report on the distribution of payments to the beneficiary concerned;



- (ii) a request for amendment (see Article 55), with a proposal for reallocation of the tasks and estimated budget of the beneficiary concerned (see Annexes 1 and 2) and, if necessary, the addition of one or more new beneficiaries (see Article 56). If termination is notified after the period set out in Article 3, no request for amendment must be submitted unless the beneficiary concerned is the coordinator. In this case the request for amendment must propose a new coordinator, and
- (iii) if termination takes effect during the period set out in Article 3, a **termination report** from the beneficiary concerned, for the open reporting period until termination, containing an overview of the progress of the work, an overview of the use of resources, the individual financial statement and, if applicable, the certificate on the financial statement (see Article 20).

The information in the termination report must also be included in the periodic report for the next reporting period (see Article 20.3).

If the request for amendment is rejected by the Commission, (because it calls into question the decision awarding the grant or breaches the principle of equal treatment of applicants), the Agreement may be terminated according to Article 50.3.1(c).

If the request for amendment is accepted by the Commission, the Agreement is **amended** to introduce the necessary changes (see Article 55).

The Commission will calculate — on the basis of the periodic reports, the termination report and the report on the distribution of payments — **calculate** the amount which is due to the beneficiary and if the (pre-financing and interim) payments received by the beneficiary exceed this amount.

The **amount which is due** is calculated in the following steps:

Step 1 — Application of the reimbursement rate to the eligible costs

The grant amount for the beneficiary is calculated by applying the reimbursement rate(s) to the total eligible costs declared by the beneficiary and its linked third parties in the termination report and approved by the Commission.

Only costs incurred by the beneficiary concerned until termination takes effect are eligible (see Article 6). Costs relating to contracts due for execution only after termination are not eligible.

Step 2 — Reduction due to substantial errors, irregularities or fraud or serious breach of obligations

In case of a reduction (see Article 43), the Commission will calculate the reduced grant amount for the beneficiary by deducting the amount of the reduction (calculated in proportion to the seriousness of the errors, irregularities or fraud or breach of obligations, in accordance with Article 43.2) from the grant amount for the beneficiary.

If the payments received **exceed the amounts due**:



- if termination takes effect during the period set out in Article 3 and the request for amendment is accepted, the beneficiary concerned must repay to the coordinator the amount unduly received. The Commission will formally notify the amount unduly received and request the beneficiary concerned to repay it to the coordinator within 30 days of receiving notification. If it does not repay the coordinator, the Commission will draw upon the Guarantee Fund to pay the coordinator and then notify a **debit note** on behalf of the Guarantee Fund to the beneficiary concerned (see Article 44);
- in all other cases, in particular if termination takes effect after the period set out in Article 3, the Commission will formally notify a **debit note** to the beneficiary concerned. If payment is not made by the date in the debit note, the Guarantee Fund will pay to the Commission the amount due and the Commission will notify a debit note on behalf of the Guarantee Fund to the beneficiary concerned (see Article 44);
- if the beneficiary concerned is the former coordinator, it must repay the new coordinator according to the procedure above, unless:
  - termination takes effect after an interim payment and
  - the former coordinator has not distributed amounts received as pre-financing or interim payments (see Article 21.7).

In this case, the Commission will formally notify a **debit note** to the former coordinator. If payment is not made by the date in the debit note, the Guarantee Fund will pay to the Commission the amount due. The Commission will then pay the new coordinator and notify a debit note on behalf of the Guarantee Fund to the former coordinator (see Article 44).

If the payments received **do not exceed the amounts due**: amounts owed to the beneficiary concerned will be included in the next interim or final payment.

If the Commission does not receive the termination report within the deadline (see above), only costs included in an approved periodic report will be taken into account.

If the Commission does not receive the report on the distribution of payments within the deadline (see above), it will consider that:

- the coordinator did not distribute any payment to the beneficiary concerned and that
- the beneficiary concerned must not repay any amount to the coordinator.

After termination, the concerned beneficiary's obligations (in particular Articles 20, 22, 23, Section 3 of Chapter 4, 36, 37, 38, 40, 42, 43 and 44) continue to apply.

## **SECTION 4 FORCE MAJEURE**

### **ARTICLE 51 — FORCE MAJEURE**

'Force majeure' means any situation or event that:

- prevents either party from fulfilling their obligations under the Agreement,
- was unforeseeable, exceptional situation and beyond the parties' control,
- was not due to error or negligence on their part (or on the part of third parties involved in the action), and
- proves to be inevitable in spite of exercising all due diligence.

The following cannot be invoked as force majeure:

- any default of a service, defect in equipment or material or delays in making them available, unless they stem directly from a relevant case of force majeure,
- labour disputes or strikes, or
- financial difficulties.

Any situation constituting force majeure must be formally notified to the other party without delay, stating the nature, likely duration and foreseeable effects.

The parties must immediately take all the necessary steps to limit any damage due to force majeure and do their best to resume implementation of the action as soon as possible.

The party prevented by force majeure from fulfilling its obligations under the Agreement cannot be considered in breach of them.

## **CHAPTER 7 FINAL PROVISIONS**

### **ARTICLE 52 — COMMUNICATION BETWEEN THE PARTIES**

#### **52.1 Form and means of communication**

Communication under the Agreement (information, requests, submissions, 'formal notifications', etc.) must:

- be made in writing and
- bear the number of the Agreement.

**Until the payment of the balance:** all communication must be made through the electronic exchange system and using the forms and templates provided there.

**After the payment of the balance:** formal notifications must be made by registered post with proof of delivery ('formal notification on paper').

Communications in the electronic exchange system must be made by persons authorised according to the Participant Portal Terms & Conditions. For naming the authorised persons, each beneficiary must have designated — before the signature of this Agreement — a 'legal entity appointed representative (LEAR)'. The role and tasks of the LEAR are stipulated in his/her appointment letter (see Participant Portal Terms & Conditions).

If the electronic exchange system is temporarily unavailable, instructions will be given on the Commission websites.

## 52.2 Date of communication

**Communications** are considered to have been made when they are sent by the sending party (i.e. on the date and time they are sent through the electronic exchange system).

**Formal notifications** through the **electronic** exchange system are considered to have been made when they are received by the receiving party (i.e. on the date and time of acceptance by the receiving party, as indicated by the time stamp). A formal notification that has not been accepted within 10 days after sending is considered to have been accepted.

Formal notifications **on paper** sent by **registered post** with proof of delivery (only after the payment of the balance) are considered to have been made on either:

- the delivery date registered by the postal service or
- the deadline for collection at the post office.

If the electronic exchange system is temporarily unavailable, the sending party cannot be considered in breach of its obligation to send a communication within a specified deadline.

## 52.3 Addresses for communication

The **electronic** exchange system must be accessed via the following URL:

<https://ec.europa.eu/research/participants/portal/desktop/en/projects/>

The Commission will formally notify the coordinator and beneficiaries in advance any changes to this URL.

**Formal notifications on paper** (only after the payment of the balance) addressed **to the Commission** must be sent to the following address:

European Commission  
DIRECTORATE-GENERAL RESEARCH & INNOVATION  
Research infrastructure  
B-1049 Brussels Belgium

Formal notifications on paper (only after the payment of the balance) addressed **to the beneficiaries** must be sent to their legal address as specified in the Participant Portal Beneficiary Register.

## ARTICLE 53 — INTERPRETATION OF THE AGREEMENT

### 53.1 Precedence of the Terms and Conditions over the Annexes

The provisions in the Terms and Conditions of the Agreement take precedence over its Annexes.

Annex 2 takes precedence over Annex 1.

### 53.2 Privileges and immunities

Not applicable

## **ARTICLE 54 — CALCULATION OF PERIODS, DATES AND DEADLINES**

In accordance with Regulation No 1182/71<sup>28</sup>, periods expressed in days, months or years are calculated from the moment the triggering event occurs.

The day during which that event occurs is not considered as falling within the period.

## **ARTICLE 55 — AMENDMENTS TO THE AGREEMENT**

### **55.1 Conditions**

The Agreement may be amended, unless the amendment entails changes to the Agreement which would call into question the decision awarding the grant or breach the principle of equal treatment of applicants.

Amendments may be requested by any of the parties.

### **55.2 Procedure**

The party requesting an amendment must submit a request for amendment signed in the electronic exchange system (see Article 52).

The coordinator submits and receives requests for amendment on behalf of the beneficiaries (see Annex 3).

If a change of coordinator is requested without its agreement, the submission must be done by another beneficiary (acting on behalf of the other beneficiaries).

The request for amendment must include:

- the reasons why;
- the appropriate supporting documents;
- for a change of coordinator without its agreement: the opinion of the coordinator (or proof that this opinion has been requested in writing).

The Commission may request additional information.

If the party receiving the request agrees, it must sign the amendment in the electronic exchange system within 45 days of receiving notification (or any additional information the Commission has requested). If it does not agree, it must formally notify its disagreement within the same deadline. The deadline may be extended, if necessary for the assessment of the request. If no notification is received within the deadline, the request is considered to have been rejected

An amendment **enters into force** on the day of the signature of the receiving party.

---

<sup>28</sup> Regulation (EEC, Euratom) No 1182/71 of the Council of 3 June 1971 determining the rules applicable to periods, dates and time-limits (OJ L 124, 8.6.1971, p. 1).

An amendment **takes effect** on the date agreed by the parties or, in the absence of such an agreement, on the date on which the amendment enters into force.

## **ARTICLE 56 — ACCESSION TO THE AGREEMENT**

### **56.1 Accession of the beneficiaries mentioned in the Preamble**

The other beneficiaries must accede to the Agreement by signing the Accession Form (see Annex 3) in the electronic exchange system (see Article 52) within 30 days after its entry into force (see Article 58).

They will assume the rights and obligations under the Agreement with effect from the date of its entry into force (see Article 58).

If a beneficiary does not accede to the Agreement within the above deadline, the coordinator must — within 30 days — request an amendment to make any changes necessary to ensure proper implementation of the action. This does not affect the Commission's right to terminate the Agreement (see Article 50).

### **56.2 Addition of new beneficiaries**

In justified cases, the beneficiaries may request the addition of a new beneficiary.

For this purpose, the coordinator must submit a request for amendment in accordance with Article 55. It must include an Accession Form (see Annex 3) signed by the new beneficiary in the electronic exchange system (see Article 52).

New beneficiaries must assume the rights and obligations under the Agreement with effect from the date of their accession specified in the Accession Form (see Annex 3).

## **ARTICLE 57 — APPLICABLE LAW AND SETTLEMENT OF DISPUTES**

### **57.1 Applicable law**

The Agreement is governed by the applicable EU law, supplemented if necessary by the law of Belgium.

### **57.2 Dispute settlement**

If a dispute concerning the interpretation, application or validity of the Agreement cannot be settled amicably, the General Court — or, on appeal, the Court of Justice of the European Union — has sole jurisdiction. Such actions must be brought under Article 272 of the Treaty on the Functioning of the EU (TFEU).

As an exception, if such a dispute is between the Commission and CENTAR ZA DIGITALNE HUMANISTICKE NAUKE, K DICTIONARIES LTD, the competent Belgian courts have sole jurisdiction.

If a dispute concerns administrative sanctions, offsetting or an enforceable decision under Article 299 TFEU (see Articles 44, 45 and 46), the beneficiaries must bring action before the General Court — or, on appeal, the Court of Justice of the European Union — under Article 263 TFEU.

## **ARTICLE 58 — ENTRY INTO FORCE OF THE AGREEMENT**

The Agreement will enter into force on the day of signature by the Commission or the coordinator, depending on which is later.

### **SIGNATURES**

For the coordinator

For the Commission



**EUROPEAN COMMISSION**  
**DIRECTORATE-GENERAL RESEARCH & INNOVATION**  
Research infrastructure



## **ANNEX 1 (part A)**

**Research and Innovation action**

**NUMBER — 731015 — ELEXIS**

## Table of Contents

1.1. The project summary.....	3
1.2. The list of beneficiaries.....	4
1.3. Workplan Tables - Detailed implementation.....	5
1.3.1. WT1 List of work packages.....	5
1.3.2. WT2 List of deliverables.....	6
1.3.3. WT3 Work package descriptions.....	10
Work package 1.....	10
Work package 2.....	15
Work package 3.....	19
Work package 4.....	24
Work package 5.....	29
Work package 6.....	33
Work package 7.....	37
Work package 8.....	44
Work package 9.....	50
Work package 10.....	56
Work package 11.....	60
1.3.4. WT4 List of milestones.....	62
1.3.5. WT5 Critical Implementation risks and mitigation actions.....	64
1.3.6. WT6 Summary of project effort in person-months.....	65
1.3.7. WT7 Tentative schedule of project reviews.....	66
1.3.8. WT8 Summary of transnational / virtual access provision per installation.....	67



# 1.1. The project summary

Project Number <sup>1</sup>	731015	Project Acronym <sup>2</sup>	ELEXIS
-----------------------------	--------	------------------------------	--------

## One form per project

### General information

Project title <sup>3</sup>	European Lexicographic Infrastructure
Starting date <sup>4</sup>	01/02/2018
Duration in months <sup>5</sup>	48
Call (part) identifier <sup>6</sup>	H2020-INFRAIA-2017-1-two-stage
Topic	INFRAIA-02-2017 Integrating Activities for Starting Communities
Fixed EC Keywords	Knowledge infrastructure
Free keywords	lexicography, computational linguistics, artificial intelligence, linked (open) data, semantic web, lesser-resourced languages

### Abstract <sup>7</sup>

The project proposes to integrate, extend and harmonise national and regional efforts in the field of lexicography, both modern and historical, with the goal of creating a sustainable infrastructure which will (1) enable efficient access to high quality lexical data in the digital age, and (2) bridge the gap between more advanced and lesser-resourced scholarly communities working on lexicographic resources. The need for such an infrastructure has clearly emerged out of the lexicographic community within the European Network of e-Lexicography COST Action which will end in 2017.

Current lexicographic resources, both modern and historical, have different levels of structuring and are not equally suitable for application in other fields, e.g. Natural Language Processing. The project will develop strategies, tools and standards for extracting, structuring and linking lexicographic resources to unlock their full potential for Linked Open Data and the Semantic Web, as well as in the context of digital humanities.

The project will help researchers create, access, share, link, analyse, and interpret heterogeneous lexicographic data across national borders, paving the way for ambitious, transnational, data-driven advancements in the field, while significantly reducing a duplication of effort across disciplinary boundaries.

ELEXIS will be carried out by a balanced consortium with distributed geographical origins. It is composed of content-holding institutions and researchers with complementary backgrounds - lexicography, digital humanities, language technology and standardisation - a crucial feature required to address the multi-disciplinary objectives of the project. In cooperation with CLARIN and DARIAH, it will focus on defining and providing common interoperability standards, workflows, conceptual models and data services as well as training and education activities focusing on user needs and cross-disciplinary fertilisations.

## 1.2. List of Beneficiaries

Project Number <sup>1</sup>	731015	Project Acronym <sup>2</sup>	ELEXIS
-----------------------------	--------	------------------------------	--------

### List of Beneficiaries

No	Name	Short name	Country	Project entry month <sup>8</sup>	Project exit month
1	INSTITUT JOZEF STEFAN	JSI	Slovenia	1	48
2	LEXICAL COMPUTING CZ SRO	LC	Czech Republic	1	48
3	STICHTING INSTITUUT VOOR DE NEDERLANDSE TAAL	IVDNT	Netherlands	1	48
4	UNIVERSITA DEGLI STUDI DI ROMA LA SAPIENZA	UNIROMA1	Italy	1	48
5	NATIONAL UNIVERSITY OF IRELAND GALWAY	NUI GALWAY	Ireland	1	48
6	OESTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN	OEAW	Austria	1	48
7	CENTAR ZA DIGITALNE HUMANISTICKE NAUKE	BCDH	Serbia	1	48
8	MAGYAR TUDOMANYOS AKADEMIA, NYELVTUDOMANYI INTEZET	MTANYTI	Hungary	1	48
9	INSTITUTE FOR BULGARIAN LANGUAGE PROF LYUBOMIR ANDREYCHIN	IBL	Bulgaria	1	48
10	UNIVERSIDADE NOVA DE LISBOA	FCSH-UNL	Portugal	1	48
11	K DICTIONARIES LTD	K Dictionaries	Israel	1	48
12	CONSIGLIO NAZIONALE DELLE RICERCHE	CNR	Italy	1	48
13	DET DANSKE SPROG- OG LITTERATURSELSKAB	DSL	Denmark	1	48
14	KOBENHAVNS UNIVERSITET	UCPH	Denmark	1	48
15	UNIVERSITAT TRIER	UT	Germany	1	48
16	EESTI KEELE INSTITUUT	EKI	Estonia	1	48
17	REAL ACADEMIA ESPANOLA	RAE	Spain	1	48

# 1.3. Workplan Tables - Detailed implementation

## 1.3.1. WT1 List of work packages

WP Number <sup>9</sup>	WP Title	Lead beneficiary <sup>10</sup>	Person-months <sup>11</sup>	Start month <sup>12</sup>	End month <sup>13</sup>
WP1	JRA Lexicographic data and workflow	3 - IVDNT	103.00	1	48
WP2	JRA Interoperability and Linked (Open) Data	5 - NUI GALWAY	97.00	1	48
WP3	JRA Lexicographic Data for NLP	4 - UNIROMA1	70.00	9	48
WP4	JRA NLP for Lexicography	2 - LC	84.00	1	48
WP5	NA Training and Education	7 - BCDH	70.00	3	42
WP6	NA Integration, evaluation and sustainability	1 - JSI	46.00	1	48
WP7	NA Dissemination and community building	6 - OEAW	77.00	1	48
WP8	VA Virtual Access	2 - LC	150.00	6	48
WP9	TA Trans-National Access	14 - UCPH	15.00	6	48
WP10	Coordination and Management	1 - JSI	26.00	1	48
WP11	Ethics requirements	1 - JSI	N/A	1	48
<b>Total</b>			738.00		

### 1.3.2. WT2 list of deliverables

Deliverable Number <sup>14</sup>	Deliverable Title	WP number <sup>9</sup>	Lead beneficiary	Type <sup>15</sup>	Dissemination level <sup>16</sup>	Due Date (in months) <sup>17</sup>
D1.1	Lexicographic practices in Europe: A Survey of User Needs	WP1	3 - IVDNT	Report	Public	12
D1.2	Best practices for lexicography – intermediate report	WP1	3 - IVDNT	Report	Public	24
D1.3	Tools for the automatic segmentation and identification of lexicographic content	WP1	3 - IVDNT	Other	Public	24
D1.4	ELEXIS conversion tools	WP1	3 - IVDNT	Other	Public	36
D1.5	Best practices for lexicography – final report	WP1	3 - IVDNT	ORDP: Open Research Data Pilot	Public	48
D1.6	Lexicographic data in ELEXIS	WP1	3 - IVDNT	Report	Public	48
D2.1	Interface for interoperable lexical resources	WP2	5 - NUI GALWAY	Report	Public	12
D2.2	Interoperable interface for Lemon and TEI resources	WP2	5 - NUI GALWAY	Other	Public	24
D2.3	Lexical Resource Linking Service	WP2	5 - NUI GALWAY	Report	Public	36
D2.4	Cross-lingual Lexical Resource Linking Web Service	WP2	5 - NUI GALWAY	Other	Public	42
D2.5	Validation and benchmarking services	WP2	5 - NUI GALWAY	Other	Public	48
D3.1	Lexical-semantic analytics for NLP: sense clustering	WP3	4 - UNIROMA1	Other	Public	18
D3.2	Multilingual Word Sense Disambiguation and Entity Linking algorithms – initial report	WP3	4 - UNIROMA1	Report	Public	24
D3.3	Lexical-semantic analytics for NLP: domain labeling	WP3	4 - UNIROMA1	Other	Public	30
D3.4	Multilingual semantic parsing – initial report	WP3	4 - UNIROMA1	Report	Public	30

<b>Deliverable Number<sup>14</sup></b>	<b>Deliverable Title</b>	<b>WP number<sup>9</sup></b>	<b>Lead beneficiary</b>	<b>Type<sup>15</sup></b>	<b>Dissemination level<sup>16</sup></b>	<b>Due Date (in months)<sup>17</sup></b>
D3.5	Multilingual Word Sense Disambiguation and Entity Linking algorithms – final report (	WP3	4 - UNIROMA1	Report	Public	42
D3.6	Multilingual semantic parsing – final report	WP3	4 - UNIROMA1	Report	Public	48
D3.7	Lexical-semantic analytics for NLP: diachronic distribution of senses	WP3	4 - UNIROMA1	Other	Public	48
D3.8	Report on Lexical-semantic analytics for NLP	WP3	4 - UNIROMA1	Report	Public	48
D4.1	Online Dictionary Post-Editing and Presentation Module	WP4	2 - LC	Other	Public	18
D4.2	Dictionary Drafting Module	WP4	2 - LC	Other	Public	24
D4.3	Crowdsourcing Module	WP4	2 - LC	Other	Public	24
D4.4	Dictionary Enhancement Module	WP4	2 - LC	Other	Public	36
D4.5	Sample Dictionary Drafts	WP4	2 - LC	Other	Public	42
D4.6	Semantically Annotated Corpora	WP4	2 - LC	Other	Public	48
D4.7	Evaluation and assessment of methods for automatic drafting of lexicographic resources	WP4	2 - LC	Other	Public	48
D4.8	Evaluation and assessment of methods for automatic enriching of lexicographic resources	WP4	2 - LC	Report	Public	48
D4.9	Evaluation and assessment of methods for crowdsourcing in lexicography	WP4	2 - LC	Report	Public	48
D5.1	ELEXIS Skillset Report	WP5	7 - BCDH	Report	Public	12
D5.2	Guidelines for Producing ELEXIS Tutorials and Instruction Manuals	WP5	7 - BCDH	Report	Public	24

<b>Deliverable Number<sup>14</sup></b>	<b>Deliverable Title</b>	<b>WP number<sup>9</sup></b>	<b>Lead beneficiary</b>	<b>Type<sup>15</sup></b>	<b>Dissemination level<sup>16</sup></b>	<b>Due Date (in months)<sup>17</sup></b>
D5.3	Overview of Online Tutorials and Instruction Manuals	WP5	7 - BCDH	Report	Public	42
D6.1	Early ELEXIS interoperability report	WP6	1 - JSI	Report	Public	18
D6.2	Recommendations on legal and IPR issues for lexicography	WP6	1 - JSI	Report	Public	24
D6.3	Intermediate ELEXIS interoperability report	WP6	1 - JSI	Report	Public	30
D6.4	Report on data seal of compliance	WP6	1 - JSI	Report	Public	36
D6.5	Final ELEXIS interoperability report including interaction with CLARIN/ DARIAH services	WP6	1 - JSI	Report	Public	48
D7.1	Visual identity and design manual	WP7	6 - OEAW	Report	Public	3
D7.2	Project Website	WP7	6 - OEAW	Report	Public	3
D7.3	Updated communication plan	WP7	6 - OEAW	Report	Public	4
D7.4	Print publicity	WP7	6 - OEAW	Report	Public	6
D7.5	First year dissemination and communication report and updated communication plan	WP7	6 - OEAW	Report	Public	14
D7.6	Second year dissemination and communication report	WP7	6 - OEAW	Report	Public	26
D7.7	Survey	WP7	6 - OEAW	Report	Public	28
D7.8	Third year dissemination and communication report	WP7	6 - OEAW	Report	Public	38
D7.9	Final communication report	WP7	6 - OEAW	Report	Public	48
D8.1	Periodic assessment of LEX1, LEX2 and LEX3 – first report	WP8	2 - LC	Report	Public	12
D8.2	Periodic assessment of LEX1, LEX2 and LEX3 – second report	WP8	2 - LC	Report	Public	24

<b>Deliverable Number<sup>14</sup></b>	<b>Deliverable Title</b>	<b>WP number<sup>9</sup></b>	<b>Lead beneficiary</b>	<b>Type<sup>15</sup></b>	<b>Dissemination level<sup>16</sup></b>	<b>Due Date (in months)<sup>17</sup></b>
D8.3	Periodic assessment of LEX1, LEX2 and LEX3 – third report	WP8	2 - LC	Report	Public	36
D8.4	Periodic assessment of LEX1, LEX2 and LEX3 – final report	WP8	2 - LC	Report	Public	48
D9.1	Report on trans-national access – year 1	WP9	14 - UCPH	Report	Public	18
D9.2	Report on trans-national access – year 2	WP9	14 - UCPH	Report	Public	30
D9.3	Report on trans-national access – year 3	WP9	14 - UCPH	Report	Public	42
D9.4	Final report on trans-national access	WP9	14 - UCPH	Report	Public	48
D10.1	Project manual and quality assurance plan	WP10	1 - JSI	Report	Confidential, only for members of the consortium (including the Commission Services)	12
D10.2	Data management plan	WP10	1 - JSI	Report	Public	24
D11.1	H - Requirement No. 1	WP11	1 - JSI	Ethics	Confidential, only for members of the consortium (including the Commission Services)	48
D11.2	POPD - Requirement No. 2	WP11	1 - JSI	Ethics	Confidential, only for members of the consortium (including the Commission Services)	48
D11.3	POPD - Requirement No. 3	WP11	1 - JSI	Ethics	Confidential, only for members of the consortium (including the Commission Services)	48
D11.4	POPD - Requirement No. 4	WP11	1 - JSI	Ethics	Confidential, only for members of the consortium (including the Commission Services)	48

### 1.3.3. WT3 Work package descriptions

<b>Work package number</b> <sup>9</sup>	WP1	<b>Lead beneficiary</b> <sup>10</sup>	3 - IVDNT
<b>Work package title</b>	JRA Lexicographic data and workflow		
<b>Start month</b>	1	<b>End month</b>	48

#### Objectives

The objectives of this work package are to:

- establish an inventory of the needs of the lexicographic community to support them in creating dictionaries.
- create robust guidelines and best practices in order to promote clearly defined workflows for lexicography.
- develop and test new methods and tools for the conversion, automatic segmentation and identification of lexicographic content.

#### Description of work and role of partners

**WP1 - JRA Lexicographic data and workflow** [Months: 1-48]

IVDNT, JSI, LC, NUI GALWAY, OEAW, BCDH, MTANYTI, IBL, FCSH-UNL, K Dictionaries, CNR, DSL, UT, EKI, RAE

The objectives of WP1 are predominantly directed towards lexicographic partners in the consortium. With the exception of two (technology) partners, all members of the consortium are included since the goal is to collect lexicographic data, to prepare common standards and solutions for the development of lexicographic resources, and to harmonise lexicographic content. The lexicographic landscape in Europe is currently rather heterogeneous, it is characterised by stand-alone lexicographic resources, which are typically encoded in incompatible data structures due to the isolation of efforts, prohibiting reuse of this valuable data in other fields, such as natural language processing, linked open data and the Semantic Web, as well as in the context of digital humanities. The majority of work in WP1 is dedicated to overcoming this problem.

Task: T1.1 User Needs (M1-M12), leader: EKI

This task will generate an overview of lexicographic practices across Europe both for born-digital and retro-digitised resources. We will build on the results of the COST action European Network in e-Lexicography, in particular the deliverables produced by Working Group 2 on retro-digitised dictionaries and Working Group 3 on innovative e-dictionaries. In addition, a European-wide survey will be carried out focussing on lexicographic workflow, metadata and data formats used within the European institutions dealing with lexicography.

Task: T1.2 Data Collection and Preparation (M1-M48), leader: INT

Within this task, the lexicographic data that will form the core content of the ELEXIS infrastructure will be collected and prepared for integration in the infrastructure.

Lexicographic data will be provided by

A. consortium partners (cf. the list of lexicographic resources in the consortium)

B. observing institutions with resources included in the European Dictionary Portal (<http://www.dictionaryportal.eu/en/>) resulting from the COST ENeL action

C. other open access resources containing lexicographic data (CLARIN, DARIAH, etc.).

These lexicographic resources have generally been compiled within national and regional projects and as such they are typically encoded in their own custom data format. As part of the data collection and preparation task these resources will be converted to a more uniform data format to allow seamless integration in linked open data.

In order to automatise the process of preparing lexicographic data for inclusion as linked open data, tools will be developed and a workflow will be established. The tools and the data will be made available in the ELEXIS infrastructure as part of LEX1 infrastructure.

Task: T1.3 Best Practices for Lexicography (M3-M48), leader: INT

The aim of this task is to facilitate the creation of lexicographic resources in European institutions, by creating robust documentation, guidelines and collections of best practices in order to promote clearly defined workflows for producing, describing and annotating lexicographic resources (both synchronic and diachronic) in accordance with international standards and interoperability formats.

Task: T1.4 Data Structuring for existing resources (M1-M24), leader: JSI

The goal of this task is to develop and test approaches for the automatic segmentation and structuring of content in digitised lexical resources, either because the resource has been retro-digitised or because it has been created in a non-structured data format (e.g. Word). Given the high level of inconsistency in this kind of material, we will explore the



possibility of using machine learning to learn dictionary structure, by training on existing high-quality lexicographic resources.

### Participation per Partner

Partner number and short name	WP1 effort
1 - JSI	9.00
2 - LC	5.00
3 - IVDNT	20.00
5 - NUI GALWAY	2.00
6 - OEAW	3.00
7 - BCDH	11.00
8 - MTANYTI	5.00
9 - IBL	11.00
10 - FCSH-UNL	3.00
11 - K Dictionaries	7.00
12 - CNR	2.00
13 - DSL	7.00
15 - UT	6.00
16 - EKI	9.00
17 - RAE	3.00
<b>Total</b>	<b>103.00</b>

### List of deliverables

Deliverable Number <sup>14</sup>	Deliverable Title	Lead beneficiary	Type <sup>15</sup>	Dissemination level <sup>16</sup>	Due Date (in months) <sup>17</sup>
D1.1	Lexicographic practices in Europe: A Survey of User Needs	3 - IVDNT	Report	Public	12
D1.2	Best practices for lexicography – intermediate report	3 - IVDNT	Report	Public	24
D1.3	Tools for the automatic segmentation and identification of lexicographic content	3 - IVDNT	Other	Public	24
D1.4	ELEXIS conversion tools	3 - IVDNT	Other	Public	36
D1.5	Best practices for lexicography – final report	3 - IVDNT	ORDP: Open Research Data Pilot	Public	48

**List of deliverables**

<b>Deliverable Number<sup>14</sup></b>	<b>Deliverable Title</b>	<b>Lead beneficiary</b>	<b>Type<sup>15</sup></b>	<b>Dissemination level<sup>16</sup></b>	<b>Due Date (in months)<sup>17</sup></b>
D1.6	Lexicographic data in ELEXIS	3 - IVDNT	Report	Public	48

**Description of deliverables**

This work package will result in the following deliverables:

D1.1 Lexicographic practices in Europe: A Survey of User Needs (M12) (document)  
Report containing an overview of the survey results including information on lexicographic workflow, metadata and data formats used within European institutions.

D1.2 Best practices for lexicography – intermediate report (M24) (document)  
An intermediate report detailing best practices for lexicography (both multilingual and monolingual, as well as historical and contemporary lexicography).

D1.3 Tools for the automatic segmentation and identification of lexicographic content (M24) (software)  
Software tools for automatic segmentation and identification of lexicographic content as part of LEX1 infrastructure.

D1.4 ELEXIS conversion tools (M36) (software)  
Conversion software to harmonise the different data formats as part of LEX1 infrastructure.

D1.5 Best practices for lexicography – final report (M48) (document)  
The final report detailing best practices for lexicography (both multilingual and monolingual, as well as historical and contemporary lexicography).

D1.6 Lexicographic data in ELEXIS (M48) (document)  
Documentation on the lexicographic data collected for the ELEXIS infrastructure.

D1.1 : Lexicographic practices in Europe: A Survey of User Needs [12]  
Report containing an overview of the survey results including information on lexicographic workflow, metadata and data formats used within European institutions.

D1.2 : Best practices for lexicography – intermediate report [24]  
An intermediate report detailing best practices for lexicography (both multilingual and monolingual, as well as historical and contemporary lexicography).

D1.3 : Tools for the automatic segmentation and identification of lexicographic content [24]  
Software tools for automatic segmentation and identification of lexicographic content as part of LEX1 infrastructure.

D1.4 : ELEXIS conversion tools [36]  
Conversion software to harmonise the different data formats as part of LEX1 infrastructure.

D1.5 : Best practices for lexicography – final report [48]  
The final report detailing best practices for lexicography (both multilingual and monolingual, as well as historical and contemporary lexicography).

D1.6 : Lexicographic data in ELEXIS [48]  
Documentation on the lexicographic data collected for the ELEXIS infrastructure.

**Schedule of relevant Milestones**

<b>Milestone number<sup>18</sup></b>	<b>Milestone title</b>	<b>Lead beneficiary</b>	<b>Due Date (in months)</b>	<b>Means of verification</b>
MS1	Project inception	1 - JSI	3	WP2, WP5: partners fully resourcing project WP8: Website established WP9: VA LEX2 platform started WP10: TNA selection procedure

Schedule of relevant Milestones

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
				established WP11: Project satisfactorily commenced
MS2	Project establishment	1 - JSI	12	WP2: user needs established (deliverable), data collection, data structuring and best practices in progress WP3: linking of resources started, common models and protocols in progress WP4: Word Sense Disambiguation and Entity Linking started WP5: all tasks in progress WP6: skills analysis finished (deliverable), training materials and measures started WP7: all tasks in progress WP8: communication tools set up, dissemination plan finished, dissemination and monitoring in progress WP9: VA LEX1 and LEX3 platforms started WP10: TNA visiting grants in progress WP11: 1st periodic progress report delivered
MS3	Project mid-term	1 - JSI	24	WP2: data structuring finished (deliverable), data collection and best practices in progress WP3: common models and protocols finished (deliverable), linking of resources in progress WP4: all tasks in progress WP5: analytics for lexicography and crowdsourcing and gamification tasks finished (deliverables), remaining tasks in progress WP7: copyright and legal issues, data seal of compliance finished (deliverables), remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 2nd periodic progress report delivered
MS4	Project pre-final	1 - JSI	42	WP2: data collection and best practices in progress

**Schedule of relevant Milestones**

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
				WP3: validation and quality assurance started, remaining tasks in progress WP4: Word Sense Disambiguation and Entity Linking finished (deliverable), remaining tasks in progress WP5: acquisition of lexical data from corpora finished (deliverable), enrichment task in progress WP6: training and education concluded (deliverables) WP7: remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 3rd periodic progress report delivered
MS5	Project impact	1 - JSI	48	All results of the project as defined in the work plan are available (deliverables)

<b>Work package number</b> <sup>9</sup>	WP2	<b>Lead beneficiary</b> <sup>10</sup>	5 - NUI GALWAY
<b>Work package title</b>	JRA Interoperability and Linked (Open) Data		
<b>Start month</b>	1	<b>End month</b>	48

### Objectives

The objectives of this work package are to:

- ensure structural and concept interoperability of existing lexical resources.
- provide cross-lingual linking for resources through BabelNet.
- validate the resources (quality assurance).

### Description of work and role of partners

#### **WP2 - JRA Interoperability and Linked (Open) Data** [Months: 1-48]

NUI GALWAY, JSI, LC, IVDNT, UNIROMA1, OEAW, MTANYTI, IBL, FCSH-UNL , K Dictionaries, CNR, DSL, UCPH

The goal of this work package is to improve the interoperability of existing lexical resources and those developed in the project by means of shared models, quality checks and most importantly linking both between resources and by means of a single shared conceptualisation, as provided by BabelNet. As a first step we will ensure structural interoperability, by developing common models and protocols based on both RDF and TEI lexical resources. We will then provide conceptual interoperability, by linking these resources so that lexical entries, senses and even fundamental concepts in different lexical resources can be easily linked, using a semi-automatic approach. Furthermore, we will exploit BabelNet as an existing multilingual resource to provide cross-lingual linking for resources. These will be implemented by an infrastructure of Web services, which will validate the resources and assist in the linking of resources.

Task: T2.1: Common models and protocols for lexicon access (1M-24M), leader: CNR-ILC

This task will define a set of common protocols, in the form of REST API calls that can allow dictionaries involved in the project to be accessed through a single interface. This model will be based on existing web standards and models including RDF, SPARQL and Lemon. For example, a standard SPARQL query will be implemented to get all lexical entries with a particular form or part-of-speech, which clients can implement either by converting their data to RDF or by using a custom interface. Furthermore, the task will define common metadata and concept properties for use within the project. The outputs of this task will be technical documentation describing the formats and tools to allow resources in Lemon RDF or TEI to be compliant with this protocol.

Task: T2.2: Semi-automatic linking of lexical resources (12M-48M), leader: NUIG

Linking lexical resources is a challenge that requires impractical amounts of human efforts, but is still not easy to solve automatically. We will develop a semi-automatic system that will make the linking problem viable for large resources, by using state-of-the-art semantic and natural language processing techniques, especially deep learning methods such as LSTMs, with a human in the loop. Furthermore, we will apply constraint-based optimisation of the linking, which can quickly find the correct mapping in an active learning setting with only a small amount of human input. As such, we will develop a single tool where a user can upload two lexical resources and interactively link them. We will then evaluate this tool by developing gold standard mappings in the context of a shared task.

Task 2.3: Cross-lingual mapping through shared conceptualisation (18M-48M), leader: UNIROMA1

In order to link lexical resources across languages, we will use one highly multilingual lexicon, BabelNet, as the basis for a cross-lingual mapping system. As such, we will extend our linking tools to cross language boundaries by pivoting through BabelNet. We will further allow for resources linked through BabelNet to be used to be submitted to BabelNet, so they can extend the resource in future releases. This will extend the tool developed in Task 2.2 with cross-lingual features and allow results to be submitted for inclusion in BabelNet.

Task 2.4: Validation and quality assurance for lexical resources (36M-48M), leader: OEAW

We will develop tools to automatically verify the quality of lexical resources at three levels: Firstly, the technical quality, which means ensuring that the resource maintains the validity of its output and does not make errors in encoding, this will be achieved by Web services that validate TEI and RDF data as provided by producers. Secondly, operational quality ensures that the lexical resources remains available and responsive as they are deployed on the Web, in particular, a service will measure uptime of each resource. Finally, scientific quality ensures that the results of the service are correct in the task they try to perform and will work by creating benchmarks for tasks in WP2, with Web services to automatically check resource performance against existing gold standards.

**Participation per Partner**

Partner number and short name	WP2 effort
1 - JSI	3.00
2 - LC	4.00
3 - IVDNT	4.00
4 - UNIROMA1	10.00
5 - NUI GALWAY	24.00
6 - OEAW	16.00
8 - MTANYTI	4.00
9 - IBL	6.00
10 - FCSH-UNL	2.00
11 - K Dictionaries	6.00
12 - CNR	7.00
13 - DSL	7.00
14 - UCPH	4.00
<b>Total</b>	<b>97.00</b>

**List of deliverables**

Deliverable Number <sup>14</sup>	Deliverable Title	Lead beneficiary	Type <sup>15</sup>	Dissemination level <sup>16</sup>	Due Date (in months) <sup>17</sup>
D2.1	Interface for interoperable lexical resources	5 - NUI GALWAY	Report	Public	12
D2.2	Interoperable interface for Lemon and TEI resources	5 - NUI GALWAY	Other	Public	24
D2.3	Lexical Resource Linking Service	5 - NUI GALWAY	Report	Public	36
D2.4	Cross-lingual Lexical Resource Linking Web Service	5 - NUI GALWAY	Other	Public	42
D2.5	Validation and benchmarking services	5 - NUI GALWAY	Other	Public	48

**Description of deliverables**

This work package will result in the following deliverables:  
 D2.1 Interface for interoperable lexical resources (M12) (document)  
 Documentation of the interface and the necessary steps for implementation  
 D2.2 Interoperable interface for Lemon and TEI resources (M24) (software)  
 Software to enable a Lemon and TEI-based lexical resources to be easily hosted and queried, based on existing SPARQL endpoints and RDF display technologies.  
 D2.3 Lexical Resource Linking Service (M36) (document)

Report and evaluation of service for linking lexical resources for linking lexical resources based on existing language alignment software

D2.4 Cross-lingual Lexical Resource Linking Web Service (M42) (software)

Extension of the linking service to support BabelNet and cross-lingual linking

D2.5 Validation and benchmarking services (M48) (software)

Services that check the validity of returned data, the uptime of services in the architecture and benchmark resources in existing NLP tasks, as appropriate.

D2.1 : Interface for interoperable lexical resources [12]

Documentation of the interface and the necessary steps for implementation.

D2.2 : Interoperable interface for Lemon and TEI resources [24]

Software to enable a Lemon and TEI-based lexical resources to be easily hosted and queried, based on existing SPARQL endpoints and RDF display technologies.

D2.3 : Lexical Resource Linking Service [36]

Report and evaluation of service for linking lexical resources for linking lexical resources based on existing language alignment software.

D2.4 : Cross-lingual Lexical Resource Linking Web Service [42]

Extension of the linking service to support BabelNet and cross-lingual linking.

D2.5 : Validation and benchmarking services [48]

Services that check the validity of returned data, the uptime of services in the architecture and benchmark resources in existing NLP tasks, as appropriate.

### Schedule of relevant Milestones

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS2	Project establishment	1 - JSI	12	WP2: user needs established (deliverable), data collection, data structuring and best practices in progress WP3: linking of resources started, common models and protocols in progress WP4: Word Sense Disambiguation and Entity Linking started WP5: all tasks in progress WP6: skills analysis finished (deliverable), training materials and measures started WP7: all tasks in progress WP8: communication tools set up, dissemination plan finished, dissemination and monitoring in progress WP9: VA LEX1 and LEX3 platforms started WP10: TNA visiting grants in progress WP11: 1st periodic progress report delivered
MS3	Project mid-term	1 - JSI	24	WP2: data structuring finished (deliverable), data collection and best practices

Schedule of relevant Milestones

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
				in progress WP3: common models and protocols finished (deliverable), linking of resources in progress WP4: all tasks in progress WP5: analytics for lexicography and crowdsourcing and gamification tasks finished (deliverables), remaining tasks in progress WP7: copyright and legal issues, data seal of compliance finished (deliverables), remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 2nd periodic progress report delivered
MS4	Project pre-final	1 - JSI	42	WP2: data collection and best practices in progress WP3: validation and quality assurance started, remaining tasks in progress WP4: Word Sense Disambiguation and Entity Linking finished (deliverable), remaining tasks in progress WP5: acquisition of lexical data from corpora finished (deliverable), enrichment task in progress WP6: training and education concluded (deliverables) WP7: remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 3rd periodic progress report delivered
MS5	Project impact	1 - JSI	48	All results of the project as defined in the work plan are available (deliverables)



<b>Work package number</b> <sup>9</sup>	WP3	<b>Lead beneficiary</b> <sup>10</sup>	4 - UNIROMA1
<b>Work package title</b>	JRA Lexicographic Data for NLP		
<b>Start month</b>	9	<b>End month</b>	48

**Objectives**

The objectives of this work package are to:

- show the impact of quality lexicographic data from the ELEXIS infrastructure in key natural language processing tasks such as word sense disambiguation, entity linking and semantic parsing
- show the impact of different kinds of analytics obtained from lexicographic data on NLP
- show that the ELEXIS infrastructure provides a key contribution to addressing the “knowledge acquisition bottleneck” issue that is pervasive in NLP, and particularly in computational lexical semantics.

**Description of work and role of partners**

**WP3 - JRA Lexicographic Data for NLP** [Months: 9-48]  
**UNIROMA1**, JSI, LC, IVDNT, OEAW, BCDH, IBL, K Dictionaries, DSL, UCPH

This WP will study and validate the innovative use of the ELEXIS lexicographic infrastructure in multiple NLP tasks:

- Word sense disambiguation and Entity Linking
- Multilingual semantic parsing
- Lexical-semantic analytics for NLP

The first two tasks are core NLP tasks whose performance can be greatly improved thanks to an interconnected network of resources. The third task, intertwined with task 4.1, analyses the multilingual lexical-semantic resources to produce benefits for NLP, from sense clustering to the diachronic predominance of senses.

Task: T3.1 Word Sense Disambiguation and Entity Linking (M9-M42), leader: JSI

Word Sense Disambiguation (WSD) is the task of automatically determining the meaning of words occurring in context. A long-standing issue of WSD is that, in order to supervise systems, huge amounts of sense-annotated sentences need to be manually created. This endeavour, which as of today, is incomplete even for English, needs to be repeated for each new domain and language, something that makes the task arduous to replicate in most European languages. Analogously, entity linking, the task of identifying key terms within text to be linked to encyclopedic repositories like Wikipedia, suffers from lower coverage in resource-poor languages. In this task we will develop and implement novel algorithms that will use the ELEXIS lexicographic resources to bootstrap large training datasets for WSD and entity linking in dozens of languages. Quantitative evaluation will be determined using standard multilingual datasets in the field, in particular from the Senseval and SemEval competitions.

Task: 3.2 Multilingual Semantic Parsing (M12-M48), leader: UNIROMA1

Recently, a task referred to as semantic parsing has been proposed that aims to map sentences to formal representations of their meaning. Semantic parsing has deeper relationships to syntactic parsing than WSD, and is therefore the task that, more than others, would seem to hold the potential to achieve the ambitious objective of machine reading, one of the long-standing goals of Artificial Intelligence (Schubert, 2006). However, most semantic parsing approaches in the literature either work in a supervised fashion with even higher annotation costs than those of WSD (Dong and Lapata, 2016; Jia and Liang, 2016) or require knowledge resources such as Freebase which seem to work only in domain-restricted specific tasks such as question answering (Reddy et al., 2016). With this task, we will develop innovative algorithms that exploit the huge multilingual network of interlinked lexical knowledge to perform multilingual semantic parsing. Key to our approach will be the exploitation of bilingual and multilingual information to achieve wide coverage and state-of-the-art performance without relying on manually curated training data.

Task: T3.3: Lexical-semantic Analytics for NLP (M12-M48), leader: UNIROMA1

In this task, first, we will compute analytics for words, senses, domains and other lexical-semantic information (including phrases, collocations, etc.) across the analyzed lexical resources. Second, we will show how this information helps boost NLP applications in three different directions:

A. sense clustering, where development of semi-automatic procedures to bring together subtle sense distinctions in clusters of meanings will be shown to improve the performance of tasks such as Word Sense Disambiguation (Navigli, 2006);

B. domain labeling of text, where the aggregated information obtained from the lexicographic network of resources will be shown to improve automatic tagging of text with domain labels in arbitrary languages thanks to developing innovative neural techniques.

C. diachronic distribution of senses, i.e., ranking senses by frequency of use over time: the use of the most frequent sense in NLP is a solid baseline used in WSD and other tasks. However, it is not systematic and is useful only for the English language. We will develop novel techniques for aggregating the predominance information of senses a) from the multitude of resources b) considering evolution over time, so as to have important impact on disambiguation and corpus analysis.

#### Participation per Partner

Partner number and short name	WP3 effort
1 - JSI	8.00
2 - LC	6.00
3 - IVDNT	2.00
4 - UNIROMA1	30.00
6 - OEAW	3.00
7 - BCDH	4.00
9 - IBL	5.00
11 - K Dictionaries	5.00
13 - DSL	3.00
14 - UCPH	4.00
<b>Total</b>	<b>70.00</b>

#### List of deliverables

Deliverable Number <sup>14</sup>	Deliverable Title	Lead beneficiary	Type <sup>15</sup>	Dissemination level <sup>16</sup>	Due Date (in months) <sup>17</sup>
D3.1	Lexical-semantic analytics for NLP: sense clustering	4 - UNIROMA1	Other	Public	18
D3.2	Multilingual Word Sense Disambiguation and Entity Linking algorithms – initial report	4 - UNIROMA1	Report	Public	24
D3.3	Lexical-semantic analytics for NLP: domain labeling	4 - UNIROMA1	Other	Public	30
D3.4	Multilingual semantic parsing – initial report	4 - UNIROMA1	Report	Public	30
D3.5	Multilingual Word Sense Disambiguation and Entity Linking algorithms – final report (	4 - UNIROMA1	Report	Public	42
D3.6	Multilingual semantic parsing – final report	4 - UNIROMA1	Report	Public	48

**List of deliverables**

<b>Deliverable Number<sup>14</sup></b>	<b>Deliverable Title</b>	<b>Lead beneficiary</b>	<b>Type<sup>15</sup></b>	<b>Dissemination level<sup>16</sup></b>	<b>Due Date (in months)<sup>17</sup></b>
D3.7	Lexical-semantic analytics for NLP: diachronic distribution of senses	4 - UNIROMA1	Other	Public	48
D3.8	Report on Lexical-semantic analytics for NLP	4 - UNIROMA1	Report	Public	48

**Description of deliverables**

This work package will result in the following deliverables:

D3.1 Lexical-semantic analytics for NLP: sense clustering (M18) (software)

Release of software for sense clustering.

D3.2 Multilingual Word Sense Disambiguation and Entity Linking algorithms – initial report (M24)

First report on software for multilingual word sense disambiguation and entity linking (document and software).

D3.3 Lexical-semantic analytics for NLP: domain labeling (M30) (software)

Release of software for domain labeling.

D3.4 Multilingual semantic parsing – initial report (M30)

First report on the development of multilingual semantic parsing software (document and software).

D3.5 Multilingual Word Sense Disambiguation and Entity Linking algorithms – final report (M42)

Final report on software for multilingual word sense disambiguation and entity linking (document and software).

D3.6 Multilingual semantic parsing – final report (M48)

Final report on the development of multilingual semantic parsing software (document and software).

D3.7 Lexical-semantic analytics for NLP: diachronic distribution of senses (M48) (software)

Release of software for diachronic distribution of senses.

D3.8 Report on Lexical-semantic analytics for NLP (M48)

Overall report on the development of lexical-semantic analytics techniques for NLP (sense clustering, domain labeling and diachronic distribution of senses).

D3.1 : Lexical-semantic analytics for NLP: sense clustering [18]

Release of software for sense clustering.

D3.2 : Multilingual Word Sense Disambiguation and Entity Linking algorithms – initial report [24]

First report on software for multilingual word sense disambiguation and entity linking (document and software).

D3.3 : Lexical-semantic analytics for NLP: domain labeling [30]

Release of software for domain labeling.

D3.4 : Multilingual semantic parsing – initial report [30]

First report on the development of multilingual semantic parsing software (document and software).

D3.5 : Multilingual Word Sense Disambiguation and Entity Linking algorithms – final report ( [42]

Final report on software for multilingual word sense disambiguation and entity linking (document and software).

D3.6 : Multilingual semantic parsing – final report [48]

Final report on the development of multilingual semantic parsing software (document and software).

D3.7 : Lexical-semantic analytics for NLP: diachronic distribution of senses [48]

Release of software for diachronic distribution of senses.

D3.8 : Report on Lexical-semantic analytics for NLP [48]

Overall report on the development of lexical-semantic analytics techniques for NLP (sense clustering, domain labeling and diachronic distribution of senses).

Schedule of relevant Milestones

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS2	Project establishment	1 - JSI	12	WP2: user needs established (deliverable), data collection, data structuring and best practices in progress WP3: linking of resources started, common models and protocols in progress WP4: Word Sense Disambiguation and Entity Linking started WP5: all tasks in progress WP6: skills analysis finished (deliverable), training materials and measures started WP7: all tasks in progress WP8: communication tools set up, dissemination plan finished, dissemination and monitoring in progress WP9: VA LEX1 and LEX3 platforms started WP10: TNA visiting grants in progress WP11: 1st periodic progress report delivered
MS3	Project mid-term	1 - JSI	24	WP2: data structuring finished (deliverable), data collection and best practices in progress WP3: common models and protocols finished (deliverable), linking of resources in progress WP4: all tasks in progress WP5: analytics for lexicography and crowdsourcing and gamification tasks finished (deliverables), remaining tasks in progress WP7: copyright and legal issues, data seal of compliance finished (deliverables), remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 2nd periodic progress report delivered
MS4	Project pre-final	1 - JSI	42	WP2: data collection and best practices in progress WP3: validation and quality assurance started, remaining

**Schedule of relevant Milestones**

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
				tasks in progress WP4: Word Sense Disambiguation and Entity Linking finished (deliverable), remaining tasks in progress WP5: acquisition of lexical data from corpora finished (deliverable), enrichment task in progress WP6: training and education concluded (deliverables) WP7: remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 3rd periodic progress report delivered
MS5	Project impact	1 - JSI	48	All results of the project as defined in the work plan are available (deliverables)

<b>Work package number</b> <sup>9</sup>	WP4	<b>Lead beneficiary</b> <sup>10</sup>	2 - LC
<b>Work package title</b>	JRA NLP for Lexicography		
<b>Start month</b>	1	<b>End month</b>	48

### Objectives

The objectives of this work package are to:

- develop methods and tools for automatic acquisition of lexicographic data using advanced NLP techniques
- develop methods and tools for introducing crowdsourcing and gamification in lexicographic process
- develop methods and tools to enrich lexicographic resources with multi-modal data

### Description of work and role of partners

#### **WP4 - JRA NLP for Lexicography** [Months: 1-48]

LC, JSI, IVDNT, UNIROMA1, NUI GALWAY, OEAW, MTANYTI, IBL, K Dictionaries, EKI

The main objective of this work package is to develop tools for automatic acquisition of lexicographic data using advanced Natural Language Processing (NLP) techniques (»dictionary-on-the-fly«). This entails both development of new tools as well as enhancing of the current ones, and their deployment within the virtual access infrastructure described in WP8 so that they become available and easily accessible for the European research community.

Since about the mid 90's modern lexicography became entirely corpus-based: lexicographers look for evidence into large collections of textual data (text corpora) to justify their decisions about how words behave in language. Concurrently, there was a huge impact due to technological innovation on the business of publishing dictionaries (as reviewed e.g. by Rundell, 2013). These two factors represent both a research opportunity and a practical need for more automated techniques for building dictionaries from corpora.

Task: T4.1 Lexico-Semantic Analytics for Lexicography (M1-M42), leader: LC

The goal of this task is to provide tools for advanced automatic lexico-semantic analytics for lexicographic purposes. These tools will enable further tasks in this work package by providing the underlying analytic framework including:

- provision of tools for segmentation/tokenisation, part-of-speech tagging and lemmatisation
- provision of tools for the analysis of contextual behaviour of words, i.e. analysis of collocations, colligations, text types and genres/registers
- provision of tools for corpus-based pattern analysis suitable for automatic word sense clustering. In the first place the focus will be on the analysis of lexico-semantic patterns suitable for automatic clustering of corpus examples according to the different word senses they belong to (see e.g. Hanks, 2013).

Task: T4.2 Acquisition of lexicographic data from corpora (M1-M48), leader: LC

In this task we will focus on all steps of a dictionary building process with the aims of automating them:

- corpus creation
- headword list development
- analysis of the corpus:
  - word senses and other lexical units
  - features of lexical units: collocations, colligations or preferred text types
  - provision of definitions (or translations)
  - provision of corpus-based examples (»knowledge-rich examples«)
- final editing of the dictionary

The ultimate goal of this task is to shift all lexicographic work and intellectual input into the post-editing phase instead of manual analysis of input data before creating a dictionary draft. Existing systems will be enhanced for this task, both for the corpus analysis part (Sketch Engine, see Kilgarrieff 2014) and for dictionary presentation and post-editing (Lexonomy, see Mechura, 2017, to appear).

Task: T4.3 Enrichment of lexicographic resources (M6-M36), leader: KD

The goal of this task is to devise automatic method for enriching existing lexicographic resources, including multi-modal data (sound, image, video, social media etc). This entails many of the items listed in the task T4.1 with the goal to automatically provide additional information such as corpus examples, morphology, collocations, linking to other dictionaries, information on current usage etc. by mapping entities into existing entries in the given lexicographic resource.

This includes a subtask focusing on diachronic lexicography, i.e. study how language changes over time, which is important for updating existing dictionaries and resources as well as for providing information on current and past trends in language.

Task: T4.4 Crowdsourcing and Gamification (M6-M24), leader: UNIROMA1

The goal of this task is to make the interlinked data available in the ELEXIS platform as homogeneously and semantically interoperable as possible. To achieve this goal, we will:

- 1) develop new techniques for automatic knowledge acquisition, by performing lexico-semantic integration of lexicographic resources into general-purpose knowledge bases like BabelNet,
- 2) explore new crowdsourcing techniques aimed at reducing the noise that typically comes from crowdworkers by involving the lexicographic community in the annotation process while at the same time focusing on user experience and novel error checking mechanisms,
- 3) starting from the work of Luis Von Ahn (von Ahn and Dabbish, 2004; von Ahn, 2006) and subsequent work in the field, we will develop innovative video games with a purpose whose objective is to bridge the gap between lexical resources monolingually and across languages.

The task will also develop algorithms that will take advantage of the three kinds of outputs (i.e., lexical knowledge from automatic techniques, lexical knowledge from crowdworkers, lexical knowledges from games with a purpose) and seamlessly adjudicate disagreements and integrate the information into the ELEXIS platform.

#### Participation per Partner

Partner number and short name	WP4 effort
1 - JSI	11.00
2 - LC	25.00
3 - IVDNT	6.00
4 - UNIROMA1	17.00
5 - NUI GALWAY	6.00
6 - OEAW	4.00
8 - MTANYTI	2.00
9 - IBL	2.00
11 - K Dictionaries	8.00
16 - EKI	3.00
<b>Total</b>	<b>84.00</b>

#### List of deliverables

Deliverable Number <sup>14</sup>	Deliverable Title	Lead beneficiary	Type <sup>15</sup>	Dissemination level <sup>16</sup>	Due Date (in months) <sup>17</sup>
D4.1	Online Dictionary Post-Editing and Presentation Module	2 - LC	Other	Public	18
D4.2	Dictionary Drafting Module	2 - LC	Other	Public	24
D4.3	Crowdsourcing Module	2 - LC	Other	Public	24
D4.4	Dictionary Enhancement Module	2 - LC	Other	Public	36
D4.5	Sample Dictionary Drafts	2 - LC	Other	Public	42

List of deliverables

Deliverable Number <sup>14</sup>	Deliverable Title	Lead beneficiary	Type <sup>15</sup>	Dissemination level <sup>16</sup>	Due Date (in months) <sup>17</sup>
D4.6	Semantically Annotated Corpora	2 - LC	Other	Public	48
D4.7	Evaluation and assessment of methods for automatic drafting of lexicographic resources	2 - LC	Other	Public	48
D4.8	Evaluation and assessment of methods for automatic enriching of lexicographic resources	2 - LC	Report	Public	48
D4.9	Evaluation and assessment of methods for crowdsourcing in lexicography	2 - LC	Report	Public	48

Description of deliverables

This work package will result in the following deliverables:

D4.1 Online Dictionary Post-Editing and Presentation Module (M18) (software)

Fully operable user-friendly online dictionary editing platform (Lexonomy).

D4.2 Dictionary Drafting Module (M24) (software)

A dictionary drafting module interconnecting a corpus management system with a dictionary editing platform.

D4.3 Crowdsourcing Module (M24) (software)

A crowdsourcing module interconnecting a corpus management system with a crowdsourcing platform.

D4.4 Dictionary Enhancement Module (M36) (software)

A dictionary enhancement module providing the capability of accessing corpora in a corpus management system from within existing dictionaries in a dictionary writing system.

D4.5 Sample Dictionary Drafts (M42) (lexicographic data)

Sample dictionaries drafted for consortium languages using the D4.2 module which will be publicly available for reviewing the methodology by the community.

D4.6 Semantically Annotated Corpora (M48) (corpus data)

Provision of text corpora (mostly based on online texts) with semantic annotation, not older than 3 years for consortium languages (and other EU languages, including lesser resourced).

D4.7 Evaluation and assessment of methods for automatic drafting of lexicographic resources (M48) (document)

A report evaluating methods for automatic drafting of lexicographic resources

D4.8 Evaluation and assessment of methods for automatic enriching of lexicographic resources (M48) (document)

A report evaluating methods for automatic enriching of lexicographic resources

D4.9 Evaluation and assessment of methods for crowdsourcing in lexicography (M48) (document)

A report evaluating methods for crowdsourcing in lexicography

D4.1 : Online Dictionary Post-Editing and Presentation Module [18]

Fully operable user-friendly online dictionary editing platform (Lexonomy).

D4.2 : Dictionary Drafting Module [24]

A dictionary drafting module interconnecting a corpus management system with a dictionary editing platform.

D4.3 : Crowdsourcing Module [24]

A crowdsourcing module interconnecting a corpus management system with a crowdsourcing platform.

D4.4 : Dictionary Enhancement Module [36]



A dictionary enhancement module providing the capability of accessing corpora in a corpus management system from within existing dictionaries in a dictionary writing system.

D4.5 : Sample Dictionary Drafts [42]

Sample dictionaries drafted for consortium languages using the D4.2 module which will be publicly available for reviewing the methodology by the community.

D4.6 : Semantically Annotated Corpora [48]

Provision of text corpora (mostly based on online texts) with semantic annotation, not older than 3 years for consortium languages (and other EU languages, including lesser resourced).

D4.7 : Evaluation and assessment of methods for automatic drafting of lexicographic resources [48]

A report evaluating methods for automatic drafting of lexicographic resources.

D4.8 : Evaluation and assessment of methods for automatic enriching of lexicographic resources [48]

A report evaluating methods for automatic enriching of lexicographic resources.

D4.9 : Evaluation and assessment of methods for crowdsourcing in lexicography [48]

A report evaluating methods for crowdsourcing in lexicography.

### Schedule of relevant Milestones

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS1	Project inception	1 - JSI	3	WP2, WP5: partners fully resourcing project WP8: Website established WP9: VA LEX2 platform started WP10: TNA selection procedure established WP11: Project satisfactorily commenced
MS2	Project establishment	1 - JSI	12	WP2: user needs established (deliverable), data collection, data structuring and best practices in progress WP3: linking of resources started, common models and protocols in progress WP4: Word Sense Disambiguation and Entity Linking started WP5: all tasks in progress WP6: skills analysis finished (deliverable), training materials and measures started WP7: all tasks in progress WP8: communication tools set up, dissemination plan finished, dissemination and monitoring in progress WP9: VA LEX1 and LEX3 platforms started WP10: TNA visiting grants in progress WP11: 1st periodic progress report delivered
MS3	Project mid-term	1 - JSI	24	WP2: data structuring finished (deliverable), data

Schedule of relevant Milestones

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
				collection and best practices in progress WP3: common models and protocols finished (deliverable), linking of resources in progress WP4: all tasks in progress WP5: analytics for lexicography and crowdsourcing and gamification tasks finished (deliverables), remaining tasks in progress WP7: copyright and legal issues, data seal of compliance finished (deliverables), remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 2nd periodic progress report delivered
MS4	Project pre-final	1 - JSI	42	WP2: data collection and best practices in progress WP3: validation and quality assurance started, remaining tasks in progress WP4: Word Sense Disambiguation and Entity Linking finished (deliverable), remaining tasks in progress WP5: acquisition of lexical data from corpora finished (deliverable), enrichment task in progress WP6: training and education concluded (deliverables) WP7: remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 3rd periodic progress report delivered

<b>Work package number</b> <sup>9</sup>	WP5	<b>Lead beneficiary</b> <sup>10</sup>	7 - BCDH
<b>Work package title</b>	NA Training and Education		
<b>Start month</b>	3	<b>End month</b>	42

**Objectives**

The objectives of this work package are to:

- evaluate the skills researchers need in order to benefit from and contribute to the lexicographic infrastructure.
- develop online training materials covering the full spectrum of producing, exploiting and disseminating lexicographic data.
- develop and enhance online tutorials for a variety of ELEXIS tools and services.
- organise and conduct ELEXIS summer schools and workshops.

**Description of work and role of partners**

**WP5 - NA Training and Education** [Months: 3-42]  
**BCDH, JSI, LC, IVDNT, NUI GALWAY, OEAW, MTANYTI, FCSH-UNL , CNR, UT**  
 While digital research infrastructures have the potential to help scholars push new frontiers of knowledge, the significant variation in the level of expertise and resources available to researchers across Europe is still a major obstacle to the wider and more productive use of integrative technologies, tools and services. The same gap — between those who actively deploy state-of-the-art tools and those who are technologically less mature — is a major challenge for the establishment and successful, sustainable functioning of European digital research infrastructures.  
 ELEXIS champions the idea of strengthening the point of contact between digital research infrastructures and various research communities at two levels: 1) through the production, dissemination, and promotion of high-quality, extensible and localisable educational materials which draw upon the best practices and models from the wealth of European-funded digital content and research projects; and 2) by implementing a series of workshops, seminars and summer schools which will train participants in topics, methods and technologies relevant to the conceptual understanding of and active contribution to the ELEXIS infrastructure.  
 With pedagogical activities and materials being crucial to ELEXIS' sustainability, the “Training and Education” work package will leverage and build upon the significant experiences of DARIAH VCC2 Education and Research Liaison, as well as #dariahTeach, an Erasmus+ funded Strategic Partnership which is developing open-source multilingual training materials in the field of Digital Humanities.  
 By channeling some of its educational outputs through #dariahTeach, ELEXIS will promote the development of fundamental skills necessary to create, implement and engage with lexicographic resources as well as tools, services, methods and technologies which are pertinent to a wide range of disciplines within the arts and humanities.  
 ELEXIS workshops and summer schools, organised both independently and in association with the existing programs such as emLEX (European Master in Lexicography), the European Summer University in Digital Humanities at Leipzig University, and the International Terminology Summer School, will guide the participants through the entire process of collecting, annotating, analyzing and disseminating lexical datasets, with a particular focus on standardisation and sustainability efforts.  
 Task: T5.1 Skills Assessment (M3-M12), leader: BCDH  
 Skills required for the active participation in the ELEXIS research infrastructure and use of its offers and services will be identified and assessed across Europe by means of qualitative interviews with lexicographers, service providers, members of the ELEXIS working groups as well as external partners, including members of DARIAH and CLARIN. The skills assessment performed in this task will be used as a basis for Tasks 5.2 and 5.3  
 Task: T5.2 Training Materials (M13-M42), leader: BCDH  
 Tutorials and instruction manuals for ELEXIS services will be created, assessed, revised and disseminated, partly in cooperation with #dariahTeach, in order to facilitate more widespread use of the ELEXIS offer across existing and new communities.  
 Task: T5.3 Training Measures (M13-M42), leader: BCDH  
 A series of workshops and summer schools will be organised to develop methodological and technological skills needed for the productive use of and contribution to ELEXIS.

**Participation per Partner**

Partner number and short name	WP5 effort
1 - JSI	2.00
2 - LC	12.00
3 - IVDNT	4.00
5 - NUI GALWAY	5.00
6 - OEAW	4.00
7 - BCDH	24.00
8 - MTANYTI	7.00
10 - FCSH-UNL	5.00
12 - CNR	4.00
15 - UT	3.00
<b>Total</b>	<b>70.00</b>

#### List of deliverables

Deliverable Number <sup>14</sup>	Deliverable Title	Lead beneficiary	Type <sup>15</sup>	Dissemination level <sup>16</sup>	Due Date (in months) <sup>17</sup>
D5.1	ELEXIS Skillset Report	7 - BCDH	Report	Public	12
D5.2	Guidelines for Producing ELEXIS Tutorials and Instruction Manuals	7 - BCDH	Report	Public	24
D5.3	Overview of Online Tutorials and Instruction Manuals	7 - BCDH	Report	Public	42

#### Description of deliverables

This work package will result in the following deliverables:

D5.1 ELEXIS Skillset Report (M12)

Primary data for this report will be gathered via qualitative interviews with lexicographers, service providers, ELEXIS working groups, DARIAH and CLARIN.

D5.2 Guidelines for Producing ELEXIS Tutorials and Instruction Manuals (M24)

A set of practical recommendations for developing tutorials and instruction manuals for ELEXIS tools and services that lower the barrier for participation in the infrastructure.

D5.3 Overview of Online Tutorials and Instruction Manuals (M42)

A clear, structured overview of tutorials and instruction manuals developed within the project will be made available on the ELEXIS website.

D5.1 : ELEXIS Skillset Report [12]

Primary data for this report will be gathered via qualitative interviews with lexicographers, service providers, ELEXIS working groups, DARIAH and CLARIN.

D5.2 : Guidelines for Producing ELEXIS Tutorials and Instruction Manuals [24]

A set of practical recommendations for developing tutorials and instruction manuals for ELEXIS tools and services that lower the barrier for participation in the infrastructure.

D5.3 : Overview of Online Tutorials and Instruction Manuals [42]

A clear, structured overview of tutorials and instruction manuals developed within the project will be made available on the ELEXIS website.

**Schedule of relevant Milestones**

<b>Milestone number<sup>18</sup></b>	<b>Milestone title</b>	<b>Lead beneficiary</b>	<b>Due Date (in months)</b>	<b>Means of verification</b>
MS2	Project establishment	1 - JSI	12	WP2: user needs established (deliverable), data collection, data structuring and best practices in progress WP3: linking of resources started, common models and protocols in progress WP4: Word Sense Disambiguation and Entity Linking started WP5: all tasks in progress WP6: skills analysis finished (deliverable), training materials and measures started WP7: all tasks in progress WP8: communication tools set up, dissemination plan finished, dissemination and monitoring in progress WP9: VA LEX1 and LEX3 platforms started WP10: TNA visiting grants in progress WP11: 1st periodic progress report delivered
MS3	Project mid-term	1 - JSI	24	WP2: data structuring finished (deliverable), data collection and best practices in progress WP3: common models and protocols finished (deliverable), linking of resources in progress WP4: all tasks in progress WP5: analytics for lexicography and crowdsourcing and gamification tasks finished (deliverables), remaining tasks in progress WP7: copyright and legal issues, data seal of compliance finished (deliverables), remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 2nd periodic progress report delivered

**Schedule of relevant Milestones**

<b>Milestone number<sup>18</sup></b>	<b>Milestone title</b>	<b>Lead beneficiary</b>	<b>Due Date (in months)</b>	<b>Means of verification</b>
MS4	Project pre-final	1 - JSI	42	WP2: data collection and best practices in progress WP3: validation and quality assurance started, remaining tasks in progress WP4: Word Sense Disambiguation and Entity Linking finished (deliverable), remaining tasks in progress WP5: acquisition of lexical data from corpora finished (deliverable), enrichment task in progress WP6: training and education concluded (deliverables) WP7: remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 3rd periodic progress report delivered
MS5	Project impact	1 - JSI	48	All results of the project as defined in the work plan are available (deliverables)

<b>Work package number</b> <sup>9</sup>	WP6	<b>Lead beneficiary</b> <sup>10</sup>	1 - JSI
<b>Work package title</b>	NA Integration, evaluation and sustainability		
<b>Start month</b>	1	<b>End month</b>	48

**Objectives**

The objectives of this work package are to:

- provide impact assessment and quality control for ELEXIS services.
- provide quality control for ELEXIS lexicographic data.
- solve IPR and legal issues in relation to lexicographic data.
- ensure sustainability of the infrastructure after the project.

**Description of work and role of partners**

**WP6 - NA Integration, evaluation and sustainability** [Months: 1-48]  
**JSI, LC, IVDNT, NUI GALWAY, OEAW, BCDH, MTANYTI, DSL, UCPH**  
 Description of work (where appropriate, broken down into tasks), lead partner and role of participants  
 Task: T6.1 Impact assessment and reporting (M1-M48), leader: JSI  
 This task is dedicated to continuous assessment and reporting on services developed and maintained by ELEXIS. Assessment reports include different methods: technical report, external evaluation, evaluation/feedback from users etc. These reports are targeted at evaluating and reporting on the impact of the whole ELEXIS infrastructure as an ecosystem of interconnected elements, they are not dedicated to particular tools and modules in the infrastructure. The goal is to give the overall assessment of the system, secure interoperability and measure the success of the infrastructure in relation to project goals.  
 Task: T6.2 Copyright and legal issues (M1-M24), leader: JSI  
 Solving IPR issues is an extremely important part of ELEXIS endeavours to bring lexicographic data to a wider research community. This is not a completely isolated effort since CLARIN already addresses this issue through the work of CLARIN legal issues committee (CLIC), which will provide the basis for our work. However, lexicographic data is traditionally even less accessible than computational linguistics data found in CLARIN. We will form a dedicated legal issues group that will specialise in legal questions and first launch a wide-spread campaign to be able to assess and make generalisation about possible solutions. On this basis recommendations on IPR issues in lexicography will be published and disseminated in the lexicographic community. The group will include also representatives from industrial partners, as this issue is particularly important also for the disappearing dictionary publishing industry.  
 As part of this work we will also investigate legal impacts of the forthcoming General Data Protection Regulation (GDPR) of the EU (No. 2016/679) in the context of lexicographic data, and we will devise best practices and guidelines how to make lexicographic resources compliant with GDPR.  
 Task: T6.3 Data seal of compliance (M1-M24), leader: BCDH  
 For users of the infrastructure it is important to be able to learn about the quality of data included at any point in time. This issue will be addressed in this task through data seal of compliance providing formal proof that a particular resource is compatible with the standards developed and recommended by ELEXIS. A group will be formed devising the criteria for assessing the quality of data which will be then implemented in a service similar to Data Seal of Approval for repositories (<https://www.datasealofapproval.org/>).  
 Task: 6.4 Compatibility with CLARIN/DARIAH services (1-48), leader:UCPH  
 ELEXIS will use services that are already available and used in CLARIN, in particular federated identity, persistent identifiers, content search or Web service chaining. This task is dedicated to establishing protocols and interoperability with CLARIN services. Decentralised services, such as repositories, will be handled by CLARIN national consortia. An ELEXIS-CLARIN subgroup will be formed within this task consisting of representatives of national CLARINs, particularly those who have CLARIN B centres, to prepare a strategy of integration of ELEXIS services into CLARIN at the end of the project. It is expected that sustainability of ELEXIS (life after four years) will be handled through national consortia. This task is led by UCPH who is the founding member of CLARIN.

**Participation per Partner**

Partner number and short name	WP6 effort
1 - JSI	16.00
2 - LC	4.00
3 - IVDNT	7.00
5 - NUI GALWAY	2.00
6 - OEAW	5.00
7 - BCDH	3.00
8 - MTANYTI	3.00
13 - DSL	2.00
14 - UCPH	4.00
<b>Total</b>	<b>46.00</b>

#### List of deliverables

Deliverable Number <sup>14</sup>	Deliverable Title	Lead beneficiary	Type <sup>15</sup>	Dissemination level <sup>16</sup>	Due Date (in months) <sup>17</sup>
D6.1	Early ELEXIS interoperability report	1 - JSI	Report	Public	18
D6.2	Recommendations on legal and IPR issues for lexicography	1 - JSI	Report	Public	24
D6.3	Intermediate ELEXIS interoperability report	1 - JSI	Report	Public	30
D6.4	Report on data seal of compliance	1 - JSI	Report	Public	36
D6.5	Final ELEXIS interoperability report including interaction with CLARIN/DARIAH services	1 - JSI	Report	Public	48

#### Description of deliverables

D6.1 Early ELEXIS interoperability report (M18)

This deliverable will provide overall assessment of ELEXIS technology developed by M15.

D6.2 Recommendations on legal and IPR issues for lexicography (M24)

This deliverable will provide recommendations on handling legal issues for lexicographic data, with the emphasis on possible measures enabling open access to the data.

D6.3 Intermediate ELEXIS interoperability report (M30)

This deliverable will provide assessment of ELEXIS technology developed by M27.

D6.4 Report on data seal of compliance (M36)

This deliverable will provide the criteria for the ELEXIS data seal of compliance and the tools for the evaluation of lexicographic data.

D6.5 Final ELEXIS interoperability report including interaction with CLARIN/DARIAH services (M48)

This deliverable will provide final assessment of ELEXIS technology and the description of technical interaction with the services provided by CLARIN and DARIAH.



D6.1 : Early ELEXIS interoperability report [18]

This deliverable will provide overall assessment of ELEXIS technology developed by M15.

D6.2 : Recommendations on legal and IPR issues for lexicography [24]

This deliverable will provide recommendations on handling legal issues for lexicographic data, with the emphasis on possible measures enabling open access to the data.

D6.3 : Intermediate ELEXIS interoperability report [30]

This deliverable will provide assessment of ELEXIS technology developed by M27.

D6.4 : Report on data seal of compliance [36]

This deliverable will provide the criteria for the ELEXIS data seal of compliance and the tools for the evaluation of lexicographic data.

D6.5 : Final ELEXIS interoperability report including interaction with CLARIN/DARIAH services [48]

This deliverable will provide final assessment of ELEXIS technology and the description of technical interaction with the services provided by CLARIN and DARIAH.

### Schedule of relevant Milestones

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS2	Project establishment	1 - JSI	12	WP2: user needs established (deliverable), data collection, data structuring and best practices in progress WP3: linking of resources started, common models and protocols in progress WP4: Word Sense Disambiguation and Entity Linking started WP5: all tasks in progress WP6: skills analysis finished (deliverable), training materials and measures started WP7: all tasks in progress WP8: communication tools set up, dissemination plan finished, dissemination and monitoring in progress WP9: VA LEX1 and LEX3 platforms started WP10: TNA visiting grants in progress WP11: 1st periodic progress report delivered
MS3	Project mid-term	1 - JSI	24	WP2: data structuring finished (deliverable), data collection and best practices in progress WP3: common models and protocols finished (deliverable), linking of resources in progress WP4: all tasks in progress WP5: analytics for lexicography and crowdsourcing and gamification tasks finished

Schedule of relevant Milestones

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
				(deliverables), remaining tasks in progress WP7: copyright and legal issues, data seal of compliance finished (deliverables), remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 2nd periodic progress report delivered
MS4	Project pre-final	1 - JSI	42	WP2: data collection and best practices in progress WP3: validation and quality assurance started, remaining tasks in progress WP4: Word Sense Disambiguation and Entity Linking finished (deliverable), remaining tasks in progress WP5: acquisition of lexical data from corpora finished (deliverable), enrichment task in progress WP6: training and education concluded (deliverables) WP7: remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 3rd periodic progress report delivered
MS5	Project impact	1 - JSI	48	All results of the project as defined in the work plan are available (deliverables)

<b>Work package number</b> <sup>9</sup>	WP7	<b>Lead beneficiary</b> <sup>10</sup>	6 - OEAW
<b>Work package title</b>	NA Dissemination and community building		
<b>Start month</b>	1	<b>End month</b>	48

### Objectives

The objectives of this work package are to:

- disseminate the project goals and outcomes effectively.
- build a loyal community of stakeholders to serve as intermediaries for reaching other communities.
- set up efficient tools for the communication with different stakeholders within academia, education, industry, the cultural heritage sector and the general public.
- organise joint events in order to support dynamic knowledge exchange and networking.
- exploit synergies in liaisons and collaborations with other infrastructure initiatives and international networks.
- recruit new users.

### Description of work and role of partners

#### **WP7 - NA Dissemination and community building** [Months: 1-48]

**OEAW**, JSI, LC, IVDNT, UNIROMA1, NUI GALWAY, BCDH, MTANYTI, IBL, FCSH-UNL , K Dictionaries, CNR, DSL, UCPH, UT, EKI, RAE

Task: T7.1 Dissemination and communication plan (M1-M14), leader: OEAW

This task concerns updating and adapting the dissemination and communication plan, outlined in the proposal. The dissemination and communication plan will also include a social media strategy.

Task: T7.2 Setting up communication tools (M1-M6), leader: LC

This task concerns setting up communication tools in order to communicate with different stakeholders and to disseminate the project goals and outcomes.

In detail, the task includes

- o Setting up a SEO optimised, responsive project website. The website design will be formalised in Deliverable D7.2.
- o Setting up a database of up-to-date contacts in order to reach the largest possible audience who will be regularly updated using a mass mailing online tool.
- o Setting up project specific social media channels (e.g. Twitter, YouTube).
- o Setting up a newsletter template according to the project design

Task: T7.3 Community building and fostering (M1-M48), leader: OEAW

The key activity for community building is the ongoing digital marketing and communication campaign described in T7.4 which will keep the community up-to-date with any new developments and provide support. A discussion forum will be set up on the website to encourage a continuous information exchange and networking opportunities.

In addition to the above, an annual community event will be organised by the consortium, hosted on a rotational basis by different consortium members, in order to build and foster the community by encouraging personal contact:

- o Kick-off meeting: one day meeting hosted by Magyar Tudományos Akadémia Nyelvtudományi Intézet (Research Institute for Linguistics of the Hungarian Academy of Sciences) in Budapest within the first month of the project
- o First ELEXIS event: two days event in the second year of the project, hosted by OEAW in Vienna
- o Second ELEXIS event: two days event in the third year of the project, hosted by CNR in Pisa
- o Final ELEXIS event: two days event in the final year of the project, hosted by LC in Brno

The task also concerns the organisation of joint events (e.g. workshops, panels, conferences, project booth) co-located with international or national events to reach other communities and to enlarge the community of interest (by the project alone or in cooperation with other initiatives), as well as the coordinated participation of consortium members in such events. Relevant international events are:

- o EURALEX and eLex for traditional lexicography
- o LREC for language resources in general
- o LSP and TKE for terminology and knowledge engineering in particular
- o ALC, Coling etc. to reach out to the NLP community in general
- o SEMANTiCS and LDK to Semantic Web community in particular
- o International Corpus Linguistics Conference and International Conference on Corpus Linguistics organised by AELINCO to reach out to the corpus linguistic community
- o DH conference or EADH Day at international level and regional Digital Humanities conferences such as DHd, DH Benelux, DHN, DHA to reach out to the Digital Humanities community

- o Language technology industry conferences like Language Technology Industry Summit, EAMT
- o ISO TC37 Meetings to reach out to the community involved in standardisation
- o CLARIN annual meeting, Inclusive Infrastructures for the Humanities: networking opportunity with other infrastructures and initiatives (e.g. CLARIN, DARIAH, DIXIT)

Task: T7.4 Dissemination and Outreach (M1-M48), leader: LC

The goal is to communicate and disseminate the project goals and outcomes effectively, to promote the organised events, and to curate the presence on social media and on research portals. This involves the development of promotion materials, both digital and print. The promotion of events not organised by WP5 e.g. the Summer Schools and training events also falls under this task. It also concerns networking with other initiatives, the engagement with the general public and the scientific dissemination. The task will be divided into the following subtask:

Subtask: 7.4.1 Digital marketing and social media

Digital promotion material and communication will include:

- o Regular news items on the project website
- o Videos, such as mini-interviews with experts from the field, mini-tutorials and hints & tips videos focused on deploying the infrastructure in various scenarios
- o Regular social media posts for example on twitter by the project on the ELEXIS account and by the consortium members on their institutional accounts or private accounts
- o Sharing the above mentioned videos on social media like twitter: This posts will strive to become viral by including video content, currently the most shared content on the web
- o Pictures and photo material as well as graphics representative for the ELEXIS project
- o Presence of the project on research portals and within interest groups
- o Twitter lists
- o Newsletter updates to the existing subscribers
- o Online questionnaire to assess the community's experience and opinion of the project's intermediate outcomes

In addition to the general digital promotion material and communication activities described above, the Trier Center for Digital Humanities (TCDH) will develop a digital crossword puzzle drawing attention to the linguistic diversity of the European dictionary landscape and a digital quiz testing people's knowledge about dictionaries. With this special promotion material we will draw attention to the project and the importance of cross-border infrastructural efforts in the research of lexicographic resources and their analysis and to engage with the website visitors.

Subtask: 7.4.2 Publicity items and direct mail

Print publicity will be used to maximise the impact of the dissemination activity to establish an initial contact with organisations where digital marketing proves unsuccessful. Print materials will also be available for download from the website. Print materials will include:

- o project general info brochure
- o topic specific one-page inserts into the brochure

A limited number of giveaways with project logo, strap line and website URL will be produced. Preference will be given to items with long-term and frequent use such as flash drives, sturdy bags from recycled material or 'hints & tips' booklets.

Subtask: 7.4.3 Networking

This task also concerns activities aimed at networking with other infrastructures, apart from organizing joint events, like CLARIN ERIC, DARIAH-EU and international initiatives like ELRA, RDA, dariahTeach.

Subtask: 7.4.4 Scientific dissemination

This task also concerns the scientific dissemination of the project results (presentations at peer-reviewed conferences and publications). Especially scientific partners will contribute to the scientific dissemination. Papers and articles will be published in open access journals and conferences, or a pre-print versions will be provided online on the ELEXIS website and, if available, archived in the institutional repository of the authors.

We will also publish information on ELEXIS in Kernerman Dictionary News. Each issue of Kernerman Dictionary News, appearing in July each year, is printed in 1,000-1,500 copies and is distributed to all the participants in all major lexicography conferences worldwide, as well as in other relevant events and venues. It is also available online.

Subtask: 7.4.5 Engaging with the public

In order to engage with the public, the ELEXIS project will participate in events like the European Researchers' Night ([http://ec.europa.eu/research/researchersnight/index\\_en.htm](http://ec.europa.eu/research/researchersnight/index_en.htm)) that takes place in several places across Europe every year in fall. There are also similar national initiatives, in Austria for example Lange Nacht der Forschung or the Kinder Uni to engage with children.

Task: T7.5 Monitoring (scientific) dissemination activities (M1-M48), leader: OEAW

The outreach and involvement will be measured with the help of the reporting system of the mass mailing tool, with website analytics and with the standard reporting features of social media portals and the figures (number of participants etc.) collected at ELEXIS events and other events. A project survey distributed in month 28 will include evaluation

of the dissemination activities (D7.6). Each year a dissemination and communication report and final communication report will be elaborated which will also include scientific publications and a collection of digital and nondigital media clippings (D7.5, D7.7, D7.8, D7.9).

#### Participation per Partner

Partner number and short name	WP7 effort
1 - JSI	4.00
2 - LC	17.00
3 - IVDNT	3.00
4 - UNIROMA1	2.00
5 - NUI GALWAY	3.00
6 - OEAW	22.00
7 - BCDH	2.00
8 - MTANYTI	5.00
9 - IBL	2.00
10 - FCSH-UNL	2.00
11 - K Dictionaries	2.00
12 - CNR	4.00
13 - DSL	2.00
14 - UCPH	2.00
15 - UT	1.00
16 - EKI	2.00
17 - RAE	2.00
<b>Total</b>	<b>77.00</b>

#### List of deliverables

Deliverable Number <sup>14</sup>	Deliverable Title	Lead beneficiary	Type <sup>15</sup>	Dissemination level <sup>16</sup>	Due Date (in months) <sup>17</sup>
D7.1	Visual identity and design manual	6 - OEAW	Report	Public	3
D7.2	Project Website	6 - OEAW	Report	Public	3
D7.3	Updated communication plan	6 - OEAW	Report	Public	4
D7.4	Print publicity	6 - OEAW	Report	Public	6
D7.5	First year dissemination and communication report and updated communication plan	6 - OEAW	Report	Public	14

List of deliverables

Deliverable Number <sup>14</sup>	Deliverable Title	Lead beneficiary	Type <sup>15</sup>	Dissemination level <sup>16</sup>	Due Date (in months) <sup>17</sup>
D7.6	Second year dissemination and communication report	6 - OEAW	Report	Public	26
D7.7	Survey	6 - OEAW	Report	Public	28
D7.8	Third year dissemination and communication report	6 - OEAW	Report	Public	38
D7.9	Final communication report	6 - OEAW	Report	Public	48

Description of deliverables

This work package will result in the following deliverables:

**D7.1 Visual identity and design manual (M3)**

A bespoke visual identity will be created including the website design, digital publicity (newsletters, presentation templates), print publicity design, the project logo and strapline so that any material, events or outcome are instantly associated with the project. A design manual will be produced to ensure consistent branding across the whole consortium.

**D7.2 Project Website (M3)**

The project website will serve as the main hub for anyone interested in the project, its progress, activities or outcomes. It will also serve as a hub for sharing and networking among the community. The website will be SEO optimised and responsive and will provide general information about the project, information about the infrastructure for virtual access and transnational access. The website will provide targeted information for the individuated target audiences. A support section will allow visitors to submit any queries related to technical issues as well as general questions. The Social Media channels of the project will be integrated into the website and there will be a section to subscribe to the newsletter and mailing lists.

**D7.3 Updated communication plan (M4)**

The deliverable D7.3 outlines the communication strategy (including the social media strategy) of the project. It will also describe measures to evaluate and assess the impact of each dissemination and communication activity. Furthermore, it includes the first year of activities and identifies additional available opportunities that could not be assessed at the proposal stage, matching them with the respective project communication goals.

**D7.4 Print publicity (M6)**

The print publicity will consist of an 8-page general project brochure with single page topic-specific inserts. This will allow a cost effective customisation of the publicity depending on the target audience.

**D7.5 First year dissemination and communication report and updated communication plan (M14)**

This report includes the activities of the first year of the project.

**D7.6 Second year dissemination and communication report (M26)**

This report includes the activities of the second year of the project.

**D7.7 Survey (M28)**

In addition to the feedback received via the website forum and support sections, the community's experience and opinion of the project's intermediate outcomes will be assessed via a survey distributed as an online questionnaire. The results will be used to reflect on the work done previously and to fine-tune the project's final outcomes to respond best to the needs of the community. An evaluation of the communication activities will also be part of the survey.

**D7.8 Third year dissemination and communication report (M38)**

This report includes the activities of the third year of the project.

**D7.9 Final communication report (M48)**

The report includes the activities since the last communication report and includes a full list of scientific publications. The report will also include the assessment of the communication activities (e.g. providing statistics).

D7.1 : Visual identity and design manual [3]

A bespoke visual identity will be created including the website design, digital publicity (newsletters, presentation templates), print publicity design, the project logo and strapline so that any material, events or outcome are instantly associated with the project. A design manual will be produced to ensure consistent branding across the whole consortium.

**D7.2 : Project Website [3]**

The project website will serve as the main hub for anyone interested in the project, its progress, activities or outcomes. It will also serve as a hub for sharing and networking among the community. The website will be SEO optimised and responsive and will provide general information about the project, information about the infrastructure for virtual access and transnational access. The website will provide targeted information for the individuated target audiences. A support section will allow visitors to submit any queries related to technical issues as well as general questions. The Social Media channels of the project will be integrated into the website and there will be a section to subscribe to the newsletter and mailing lists.

**D7.3 : Updated communication plan [4]**

The deliverable D7.3 outlines the communication strategy (including the social media strategy) of the project. It will also describe measures to evaluate and assess the impact of each dissemination and communication activity. Furthermore, it includes the first year of activities and identifies additional available opportunities that could not be assessed at the proposal stage, matching them with the respective project communication goals.

**D7.4 : Print publicity [6]**

The print publicity will consist of an 8-page general project brochure with single page topic-specific inserts. This will allow a cost effective customisation of the publicity depending on the target audience.

**D7.5 : First year dissemination and communication report and updated communication plan [14]**

This report includes the activities of the first year of the project.

**D7.6 : Second year dissemination and communication report [26]**

This report includes the activities of the second year of the project.

**D7.7 : Survey [28]**

In addition to the feedback received via the website forum and support sections, the community’s experience and opinion of the project’s intermediate outcomes will be assessed via a survey distributed as an online questionnaire. The results will be used to reflect on the work done previously and to fine-tune the project’s final outcomes to respond best to the needs of the community. An evaluation of the communication activities will also be part of the survey.

**D7.8 : Third year dissemination and communication report [38]**

This report includes the activities of the third year of the project.

**D7.9 : Final communication report [48]**

The report includes the activities since the last communication report and includes a full list of scientific publications. The report will also include the assessment of the communication activities (e.g. providing statistics).

**Schedule of relevant Milestones**

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS1	Project inception	1 - JSI	3	WP2, WP5: partners fully resourcing project WP8: Website established WP9: VA LEX2 platform started WP10: TNA selection procedure established WP11: Project satisfactorily commenced
MS2	Project establishment	1 - JSI	12	WP2: user needs established (deliverable), data collection,



Schedule of relevant Milestones

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
				data structuring and best practices in progress WP3: linking of resources started, common models and protocols in progress WP4: Word Sense Disambiguation and Entity Linking started WP5: all tasks in progress WP6: skills analysis finished (deliverable), training materials and measures started WP7: all tasks in progress WP8: communication tools set up, dissemination plan finished, dissemination and monitoring in progress WP9: VA LEX1 and LEX3 platforms started WP10: TNA visiting grants in progress WP11: 1st periodic progress report delivered
MS3	Project mid-term	1 - JSI	24	WP2: data structuring finished (deliverable), data collection and best practices in progress WP3: common models and protocols finished (deliverable), linking of resources in progress WP4: all tasks in progress WP5: analytics for lexicography and crowdsourcing and gamification tasks finished (deliverables), remaining tasks in progress WP7: copyright and legal issues, data seal of compliance finished (deliverables), remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 2nd periodic progress report delivered
MS4	Project pre-final	1 - JSI	42	WP2: data collection and best practices in progress WP3: validation and quality assurance started, remaining tasks in progress WP4: Word Sense Disambiguation and



**Schedule of relevant Milestones**

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
				Entity Linking finished (deliverable), remaining tasks in progress WP5: acquisition of lexical data from corpora finished (deliverable), enrichment task in progress WP6: training and education concluded (deliverables) WP7: remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 3rd periodic progress report delivered
MS5	Project impact	1 - JSI	48	All results of the project as defined in the work plan are available (deliverables)

<b>Work package number</b> <sup>9</sup>	WP8	<b>Lead beneficiary</b> <sup>10</sup>	2 - LC
<b>Work package title</b>	VA Virtual Access		
<b>Start month</b>	6	<b>End month</b>	48

### Objectives

Provision of access to the following infrastructure(s): LEX1

Description of the infrastructure

Name of the infrastructure (and its installations, if applicable): LEX1

Location (town, country) of the infrastructure: Ljubljana, Slovenia

Web site address: /to be determined/

Annual operating costs (excl. investment costs) of the infrastructure (€): 38,333 €

Description of the infrastructure:

The LEX1 infrastructure will be set up during the project and will consist of several sets of tools:

(1) tools for automatic segmentation and structuring of lexicographic content for dictionaries that are currently produced in digital environments and are encoded in their own custom data format.

(2) conversion and alignment tools that will provide users of the infrastructure with the possibility to harmonise and convert their lexicographic resources to a uniform data format that allows their seamless integration in Linked Open Data.

(3) linking tools that will provide conceptual interoperability, enabling linking of ELEXIS lexicographic resources. This will provide the possibility to link lexical entries, senses and fundamental concepts in different lexical resources, using a semi-automatic approach. BabelNet, as an existing multilingual resource to provide cross-lingual linking, will be exploited for this purpose.

(4) ELEXIS matrix dictionary: extensive linking of existing lexicographic resources by pivoting through BabelNet will enable the creation of what we call ELEXIS matrix dictionary – a universal repository of linked senses, meaning descriptions, etymological data, collocations, phraseology, translation equivalents, examples of usage and all other types of lexical information found in all types of existing lexicographic resources, monolingual, multilingual, modern, historical etc., available through a RESTful web service as part of LEX1.

Services currently offered by the infrastructure:

The infrastructure is currently not available and will be set up in the first year of the project with the developments of the tools within JRA work packages, and based on pre-existing software in the consortium (cf. list on page 21). Annual operating costs are calculated on the basis of estimated costs of a system administrator, support personnel and management administrator for three years of the project (year 2-4).

Start of construction: M1

Availability: M12-M48 (adding parts of the infrastructure in sequence as they are developed in JRA WPs)

Provision of access to the following infrastructure(s): LEX2

Description of the infrastructure

Name of the infrastructure (and its installations, if applicable): LEX2, Sketch Engine

Location (town, country) of the infrastructure: Brno, Czech Republic

Web site address: <http://www.sketchengine.co.uk>

Annual operating costs (excl. investment costs) of the infrastructure (€): 55,357 €

Description of the infrastructure:

(1) Sketch Engine as a leading corpus management system will be used as the core for the LEX2 infrastructure that will include ELEXIS word sense disambiguation and entity linking tools dedicated to semantic processing of corpus data, including tasks such as sense clustering, domain labeling of text, diachronic distribution of senses (ranking senses by frequency of use over time) etc. The infrastructure includes also ELEXIS »dictionary-on-the-fly« segment. The service will be able to produce a proto-dictionary with sense distribution, extracted definitions, collocations, multi-word expressions, (good dictionary) examples, translation equivalents and data in other modalities.

(2) In addition, two complementary sets of tools will be provided: lexicographic workflow tools and crowdsourcing and gamification tools. The first will include a user-friendly open source online dictionary writing system (Lexonomy), integrated with the Sketch Engine, with the aim to provide the central dictionary writing platform for new lexicography which also includes new possibilities of online collaboration. The other will provide tools for new techniques of dictionary creation, such as explicit or implicit crowdsourcing (gamification).

Services currently offered by the infrastructure:

Sketch Engine is a leading corpus management system operating since 2003. It provides online access to large text corpora for over 80 languages and allows users to build their own corpora. It exploits state-of-the-art natural language

processing pipelines for automatic annotation of text corpora, including Part-of-Speech tagging, lemmatisation and shallow parsing. It has been designed for, and from 2003 used by, lexicographers from all around the world, including publishing houses such as Oxford University Press, Cambridge University Press, Macmillan, HarperCollins, Le Robert etc, for improving and speeding up the dictionary building process. Over the 13 years of existence, it added many advanced analytical functions for studying how language works. In recent years its user base was significantly expanding into the translation and localisation market by enhancing the workflow of translators and terminologists.

Sketch Engine provides access to large text corpora for over 80 languages, allows its users to automatically build new corpora (either from their own texts, or from the web), and provides many advanced analytical functions to study the corpora. It gained its name after word sketches, one-page summaries of a word's collocations derived from a corpus. Sketch Engine streamlines all stages of the process of dictionary building, by (semi-)automating the acquisition of corpus-derived: headword list, dictionary examples, collocations, definitions, labels, word-sense clusters.

It has been used since 2003 by lexicographers and linguists all around the world to carry out lexicographic and linguistic research as part of dozens of research projects, including:

- Macmillan English Dictionary (2008)
- Oxford Children Dictionary (since 2010)
- DANTE lexicographic database (2010)
- Frequency Dictionary for Dutch (2013)
- Estonian Collocation Dictionary (2015)
- other lexicographic projects on Arabic, Chinese, Czech, Dutch, English, Estonian, German, Hindi, Irish, Japanese, Latin, Polish, Portuguese, Romanian, Slovak, Slovene, Spanish, Tibetan, Turkish, Vietnamese and others; a comprehensive bibliography is provided on the Sketch Engine website.

As of 2016 Sketch Engine has about 45,000 users, and answers 500,000 requests daily.

Availability: M3-M48 (adding new parts of the infrastructure in sequence as they are developed in JRA WPs)

Provision of access to the following infrastructure(s): LEX3

Description of the infrastructure

Name of the infrastructure: LEX3

Location (town, country) of the infrastructure: Trier, Germany

Web site address: /to be determined/

Annual operating costs (excl. investment costs) of the infrastructure (€): 11,786 €

Description of the infrastructure:

LEX3 infrastructure will be dedicated to retrodigitised dictionaries and will include (1) tools for automatic segmentation and structuring of content in retro-digitised dictionaries, and (2) Legacy Dictionary Viewer: a generic, modular dictionary publication tool for retrodigitised dictionaries which offers interfaces for the analysis and profiling of the underlying lexical data. Based on international standards and wide-spread encoding schemes, including the best-practice guidelines developed in WP1, the viewer will allow users — lexicographers and digital humanists alike — to publish user-friendly, online editions or retrodigitised dictionaries without facing extensive technical or financial obstacles. This is particularly important in the context of lesser-resourced languages and/or institutions that don't have the manpower or the technical expertise to build their own dictionary platforms. The viewer will include different visualisation, geolocation and profiling tools that make it possible for endusers to explore and navigate the dictionary content in novel ways that go beyond the dominant look-up paradigm. Another important aspect of the LDV will be the handling of dictionary-specific characters for which there currently does not exist an appropriate representation in UNICODE. In order to deal with this issue, the viewer will be connected to a central XML-based character repository, which will simplify the process of encoding legacy dictionaries. The LDV will be used in Training Measures in WP5 and potentially find an important pedagogical role beyond ELEXIS as a tool for the study of historical dictionaries.

Services currently offered by the infrastructure:

The infrastructure is currently not available and will be set up in the first year of the project along with the developments of the tools within JRA work packages. The costs are calculated on the basis of estimated cost of a system administrator, support personnel and management administrator for three years of the project (year 2-4).

Start of construction: M1

Availability: M12-M48

## Description of work and role of partners

**WP8 - VA Virtual Access** [Months: 6-48]

LC, JSI, UT

Task 8.1: Integrated access to the LEX1 part of the infrastructure (M12-M48), leader: JSI

Task 8.2: Integrated access to the LEX2 part of the infrastructure (M3-M48), leader: LC

Task 8.3: Integrated access to the LEX3 part of the infrastructure (M12-M48), leader: TCDH

Modality of access under this proposal:

In LEX1 virtual access infrastructure we will use two modalities of access: we will provide access to ELEXIS data (e.g. matrix dictionary) through a single interface (web site address to be determined) and standard SPARQL query service will be implemented for this type of access. For structuring, conversion and linking tools we will provide a web form where the user will be able to upload his/her own resource either by providing us a direct download link or by uploading the file using an HTTP POST request. After submitting the file, the service will create a worker task that will process the file. Depending on the size of the input file, the task could run for several hours. After the task is finished and the resulting data set is available, we will notify the user by sending him an email with the link where the file can be downloaded. Relevant lexical information from the converted data sets will be included in the ELEXIS matrix dictionary through BabelNet concepts.

Access to LEX2 (Sketch Engine) will be provided through the Single-Sign-On (SSO) mechanism, since Sketch Engine is an existing Service Provider within the world-wide eduGAIN network. To activate the SSO access, the institution will go through a one-time procedure of providing their Identity Provider ID to Sketch Engine whose support team will set the access up. Sketch Engine does not require any kind of scheduling and stands to the disposition of users anytime. Since Sketch Engine is a Web service offering wide range of different functions, we define the unit of access as one hour of consecutive use by one user.

In LEX3 virtual access infrastructure we will also use two modalities of access: similar to LEX1 we will provide a web form where the user will be able to upload his/her own resource in case of tools for automatic segmentation and structuring of content in retro-digitised dictionaries, and a notification with the link where the processed file can be accessed. Legacy Dictionary Viewer will be implemented as an online dictionary publication tool with an upload option, and visualisation service.

For virtual access infrastructure in ELEXIS (LEX1, LEX2 and LEX3) we use actual cost to declare access cost. With the exception of the Sketch Engine, the infrastructure is not implemented yet and actual cost is the most convenient mode of calculation of access cost with an infrastructure in development.

Support offered under this proposal:

In virtual access work package, a significant part of budget is allocated to user support (cf. Estimation of the Access Costs for Virtual Access). As ELEXIS is connecting (technologically) more advanced research communities (NLP, AI, SW) and less advanced and resourced lexicographic communities, we expect that the latter will need support in various forms (online, call-in, crowdsourced etc).

Sketch Engine as part of LEX2 already offers access to state-of-the-art natural language processing tools that users can instantly apply to their corpora. Sketch Engine provides a 9/5 technical support team which at the moment routinely handles about 25 support requests daily.

Outreach to new users:

We will use the outreach measures described in task T7.4 (Dissemination and Outreach) in WP7, led by Lexical Computing, to publicise virtual access activities and attract new users of LEX1, LEX2 and LEX3. Sketch Engine as part of LEX2 infrastructure is already a well-known product among lexicographers (as a paid service). Upon establishing this infrastructure, existing contacts will be used to notify research teams across whole Europe about the instant availability of the service at no charge to them. We expect an increase in the number of users, which will be closely monitored through the SSO mechanisms.

The vast majority (> 99 %) of existing users of Sketch Engine comes from outside the hosting country of the infrastructure (Czech Republic).

Review procedure under this proposal:

Virtual access activities in WP8 will be regularly reviewed by a team of external experts selected from the ELEXIS International Advisory Board. As top experts in both relevant fields – lexicography and NLP, they will be familiar with the goals of the project, but also independent and non-members of the consortium. Reports will be available as project deliverables. In the case of the Sketch Engine access, any eduGAIN member (Identity Provider) from the European Union will be granted access, with no difference in treatment.

Participation per Partner

Partner number and short name	WP8 effort
1 - JSI	52.00
2 - LC	90.00
15 - UT	8.00

Partner number and short name	WP8 effort
<b>Total</b>	150.00

**List of deliverables**

Deliverable Number <sup>14</sup>	Deliverable Title	Lead beneficiary	Type <sup>15</sup>	Dissemination level <sup>16</sup>	Due Date (in months) <sup>17</sup>
D8.1	Periodic assessment of LEX1, LEX2 and LEX3 – first report	2 - LC	Report	Public	12
D8.2	Periodic assessment of LEX1, LEX2 and LEX3 – second report	2 - LC	Report	Public	24
D8.3	Periodic assessment of LEX1, LEX2 and LEX3 – third report	2 - LC	Report	Public	36
D8.4	Periodic assessment of LEX1, LEX2 and LEX3 – final report	2 - LC	Report	Public	48

**Description of deliverables**

D8.1 Periodic assessment of LEX1, LEX2 and LEX3 – first report (M12)  
 First report on virtual access services in ELEXIS by an external board (select member of ELEXIS International Advisory Board).

D8.2 Periodic assessment of LEX1, LEX2 and LEX3 – second report (M24)  
 Intermediate report on virtual access services in ELEXIS by an external board (select member of ELEXIS International Advisory Board).

D8.3 Periodic assessment of LEX1, LEX2 and LEX3 – third report (M36)  
 Intermediate report on virtual access services in ELEXIS by an external board (select member of ELEXIS International Advisory Board).

D8.4 Periodic assessment of LEX1, LEX2 and LEX3 – final report (M48)  
 Final report on virtual access services in ELEXIS by an external board (select member of ELEXIS International Advisory Board).

D8.1 : Periodic assessment of LEX1, LEX2 and LEX3 – first report [12]  
 First report on virtual access services in ELEXIS by an external board (select member of ELEXIS International Advisory Board).

D8.2 : Periodic assessment of LEX1, LEX2 and LEX3 – second report [24]  
 Intermediate report on virtual access services in ELEXIS by an external board (select member of ELEXIS International Advisory Board).

D8.3 : Periodic assessment of LEX1, LEX2 and LEX3 – third report [36]  
 Intermediate report on virtual access services in ELEXIS by an external board (select member of ELEXIS International Advisory Board).

D8.4 : Periodic assessment of LEX1, LEX2 and LEX3 – final report [48]  
 Final report on virtual access services in ELEXIS by an external board (select member of ELEXIS International Advisory Board).

Schedule of relevant Milestones

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS1	Project inception	1 - JSI	3	WP2, WP5: partners fully resourcing project WP8: Website established WP9: VA LEX2 platform started WP10: TNA selection procedure established WP11: Project satisfactorily commenced
MS2	Project establishment	1 - JSI	12	WP2: user needs established (deliverable), data collection, data structuring and best practices in progress WP3: linking of resources started, common models and protocols in progress WP4: Word Sense Disambiguation and Entity Linking started WP5: all tasks in progress WP6: skills analysis finished (deliverable), training materials and measures started WP7: all tasks in progress WP8: communication tools set up, dissemination plan finished, dissemination and monitoring in progress WP9: VA LEX1 and LEX3 platforms started WP10: TNA visiting grants in progress WP11: 1st periodic progress report delivered
MS3	Project mid-term	1 - JSI	24	WP2: data structuring finished (deliverable), data collection and best practices in progress WP3: common models and protocols finished (deliverable), linking of resources in progress WP4: all tasks in progress WP5: analytics for lexicography and crowdsourcing and gamification tasks finished (deliverables), remaining tasks in progress WP7: copyright and legal issues, data seal of compliance finished (deliverables), remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting

Schedule of relevant Milestones

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
				grants in progress WP11: 2nd periodic progress report delivered
MS4	Project pre-final	1 - JSI	42	WP2: data collection and best practices in progress WP3: validation and quality assurance started, remaining tasks in progress WP4: Word Sense Disambiguation and Entity Linking finished (deliverable), remaining tasks in progress WP5: acquisition of lexical data from corpora finished (deliverable), enrichment task in progress WP6: training and education concluded (deliverables) WP7: remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 3rd periodic progress report delivered
MS5	Project impact	1 - JSI	48	All results of the project as defined in the work plan are available (deliverables)



<b>Work package number</b> <sup>9</sup>	WP9	<b>Lead beneficiary</b> <sup>10</sup>	14 - UCPH
<b>Work package title</b>	TA Trans-National Access		
<b>Start month</b>	6	<b>End month</b>	48

### Objectives

An important activity of ELEXIS is to provide trans-national access for researchers or research teams to research facilities, lexicographical resources which are not fully accessible online or where professional on the spot expertise is needed in order to ensure and optimise mutual knowledge exchange and cross-fertilisation. We expect that the planned TNA activities as a whole will have a long-term impact especially for the under-resourced languages and will all in all strengthen the infrastructure and collaborative network provided by ELEXIS.

Objectives of the trans-national activities can be summarised as follows:

to offer opportunities to researchers or research teams to access research facilities with an excellent combination of advanced technology and expertise

to support training of new specialists in the field of e-lexicography in order to conduct high-quality research and ensure sustainability of infrastructure

to ensure support for excellent scholarly research projects and innovative enterprises and also support the complex multi-disciplinary research

to encourage the integrative use of technology and methodologies

to improve the overall services available to the research community

to exchange knowledge and experience and to work towards future common projects and objectives

to create an interdisciplinary community, collaborating on activities that are fully or partially of relevance to the proposed work.

to create knowledge at the interaction between academia and societal knowledge

Provision of access to the following infrastructure: ELEXIS-NL

Description of the infrastructure

Name of the infrastructure: Institute for Dutch Language (INT)

Location (town, country) of the infrastructure: Leiden, Netherlands

Web site address: <http://ivdnt.org/>

Description of the infrastructure:

INT is the central institution dedicated to lexicographic description of Dutch

Services currently offered by the infrastructure: cf. list of lexicographic resources on page 21.

Provision of access to the following infrastructure: ELEXIS-DK

Description of the infrastructure

Name of the infrastructure: Danish Society for Language and Literature (DSL) and University of Copenhagen (UCPH)

Location (town, country) of the infrastructure: Copenhagen, Denmark

Web site address: <http://dsl.dk/>, <http://cst.ku.dk>

Description of the infrastructure:

DSL is the central institution dedicated to lexicographic description of Danish. With UCPH they have further developed a number of NLP lexicographical resources (<http://cst.ku.dk>)

Services currently offered by the infrastructure: cf. list of lexicographic resources on page 21.

Provision of access to the following infrastructure: ELEXIS-EE

Description of the infrastructure

Name of the infrastructure: Institute for Estonian Language (EKI)

Location (town, country) of the infrastructure: Tallinn, Estonia

Web site address: <https://www.eki.ee/EN/>

Description of the infrastructure:

EKI is the central institution dedicated to lexicographic description of Estonian

Services currently offered by the infrastructure: cf. list of lexicographic resources on page 21.

Provision of access to the following infrastructure: ELEXIS-SP

Description of the infrastructure

Name of the infrastructure: Spanish Royal Academy (RAE)

Location (town, country) of the infrastructure: Madrid, Spain

Web site address: <http://www.rae.es/>

Description of the infrastructure:

RAE is the central institution dedicated to lexicographic description of Spanish



Services currently offered by the infrastructure: cf. list of lexicographic resources on page 21.

Provision of access to the following infrastructure: ELEXIS-HU

Description of the infrastructure

Name of the infrastructure: Hungarian Academy of Sciences (RILMTA)

Location (town, country) of the infrastructure: Budapest, Hungary

Web site address: <http://www.nytud.hu/>

Description of the infrastructure:

RILMTA is the central institution dedicated to lexicographic description of Hungarian

Services currently offered by the infrastructure: cf. list of lexicographic resources on page 21.

Provision of access to the following infrastructure: ELEXIS-BG

Description of the infrastructure

Name of the infrastructure: Institute of Bulgarian Language Lyubomir Andreychin (IBL)

Location (town, country) of the infrastructure: Sofia, Bulgaria

Web site address: <http://ibl.bas.bg/en/>

Description of the infrastructure:

IBL is the central institution dedicated to lexicographic description of Bulgarian

Services currently offered by the infrastructure: cf. list of lexicographic resources on page 21.

Provision of access to the following infrastructure: ELEXIS-AT

Description of the infrastructure

Name of the infrastructure: Austrian Academy of Sciences (OEAW)

Location (town, country) of the infrastructure: Vienna, Austria

Web site address: <http://www.oeaw.ac.at/en/>

Description of the infrastructure:

OEAW is involved in lexicographic description of German dialects in Austria

Services currently offered by the infrastructure: cf. list of lexicographic resources on page 21.

Provision of access to the following infrastructure: ELEXIS-DE

Description of the infrastructure

Name of the infrastructure: Trier Center for Digital Humanities (TCDH)

Location (town, country) of the infrastructure: Trier, Germany

Web site address: <http://kompetenzzentrum.uni-trier.de/en/>

Description of the infrastructure:

TCDH is involved in lexicographic work on German historical dictionaries

Services currently offered by the infrastructure: cf. list of lexicographic resources on page 21.

Provision of access to the following infrastructure: ELEXIS-RS

Description of the infrastructure

Name of the infrastructure: Belgrade Center for Digital Humanities (BCDH)

Location (town, country) of the infrastructure: Belgrade, Serbia

Web site address: <http://www.humanistika.org/>

Description of the infrastructure:

BCDH is involved in lexicographic work on Serbian historical dictionaries

Services currently offered by the infrastructure: cf. list of lexicographic resources on page 21.

Provision of access to the following infrastructure: ELEXIS-IL

Description of the infrastructure

Name of the infrastructure: K-Dictionaries (KD)

Location (town, country) of the infrastructure: Tel Aviv, Israel

Web site address: <http://kdictionaries-online.com/>

Description of the infrastructure:

KD is involved in work on multilingual dictionaries

Services currently offered by the infrastructure: cf. list of lexicographic resources on page 21.

Provision of access to the following infrastructure: ELEXIS-SI

Description of the infrastructure

Name of the infrastructure: »Jožef Stefan« Institute (JSI)

Location (town, country) of the infrastructure: Ljubljana, Slovenia

Web site address: <https://www.ijs.si/ijsw>

Description of the infrastructure:

JSI is involved in work on dictionaries of modern Slovene

Services currently offered by the infrastructure: cf. list of lexicographic resources on page 21.

**Description of work and role of partners**

**WP9 - TA Trans-National Access** [Months: 6-48]  
**UCPH, JSI, IVDNT, OEAW, BCDH, MTANYTI, IBL, K Dictionaries, CNR, DSL, UT, EKI, RAE**  
 Modality of access under this proposal:  
 ELEXIS will announce visiting grants with an average duration of two weeks for researchers to experiment with and work on lexicographical data in a context of mutual knowledge exchange with the hosting institutions. More specifically, 5 calls for visiting grants will be launched twice a year during the project period amounting to an overall number of 30-40 grants. The calls will include descriptions of the particular lexicographical resources, tools, and expertise that are made available.  
 Support offered under this proposal:  
 In ELEXIS, we define each partner with lexicographic data as the regional or national lexicographic infrastructure for the particular language, as most partners are also national institutions responsible for language description. Researchers within the EU member states (or associated countries) will be invited to apply for free-of-charge access to and support in one of the above mentioned providers of lexicographical infrastructures. During the grant visits, the hosting institutions will provide support in terms of lexicographical and IT man power expertise.  
 Outreach to new users:  
 ELEXIS trans-national access calls will be advertised via the project website, the ELEXIS newsletters and via social media.  
 Review procedure under this proposal:  
 The incoming applications will be sent to the selected access providers/hosts for approval, and a TNA assessment committee lead by UCPH and consisting of two additional ELEXIS researchers will assess the scientific rationale behind the applications. Finally, it will be the obligation of UCPH to notify the applicants regarding acceptance/refusal.  
 Each grant holder will be obliged to fill out a report on the outcomes of his/her visit which will be summarised in the yearly deliverables of the WP.

**Participation per Partner**

Partner number and short name	WP9 effort
1 - JSI	1.00
3 - IVDNT	1.00
6 - OEAW	1.00
7 - BCDH	1.00
8 - MTANYTI	1.00
9 - IBL	1.00
11 - K Dictionaries	1.00
12 - CNR	1.00
13 - DSL	1.00
14 - UCPH	3.00
15 - UT	1.00
16 - EKI	1.00
17 - RAE	1.00
<b>Total</b>	<b>15.00</b>

**List of deliverables**

<b>Deliverable Number<sup>14</sup></b>	<b>Deliverable Title</b>	<b>Lead beneficiary</b>	<b>Type<sup>15</sup></b>	<b>Dissemination level<sup>16</sup></b>	<b>Due Date (in months)<sup>17</sup></b>
D9.1	Report on trans-national access – year 1	14 - UCPH	Report	Public	18
D9.2	Report on trans-national access – year 2	14 - UCPH	Report	Public	30
D9.3	Report on trans-national access – year 3	14 - UCPH	Report	Public	42
D9.4	Final report on trans-national access	14 - UCPH	Report	Public	48

**Description of deliverables**

D9.1 Report on trans-national access – year 1 (M18)  
 Cumulative report on research visits finished from M6-M18.  
 D9.2 Report on trans-national access – year 2 (M30)  
 Cumulative report on research visits finished from M18-M30.  
 D9.3 Report on trans-national access – year 3 (M42)  
 Cumulative report on research visits finished from M30-M42.  
 D9.4 Final report on trans-national access (M48)  
 Cumulative report on trans-national access in ELEXIS.

D9.1 : Report on trans-national access – year 1 [18]  
 Cumulative report on research visits finished from M6-M18.

D9.2 : Report on trans-national access – year 2 [30]  
 Cumulative report on research visits finished from M18-M30.

D9.3 : Report on trans-national access – year 3 [42]  
 Cumulative report on research visits finished from M30-M42.

D9.4 : Final report on trans-national access [48]  
 Cumulative report on trans-national access in ELEXIS.

**Schedule of relevant Milestones**

<b>Milestone number<sup>18</sup></b>	<b>Milestone title</b>	<b>Lead beneficiary</b>	<b>Due Date (in months)</b>	<b>Means of verification</b>
MS1	Project inception	1 - JSI	3	WP2, WP5: partners fully resourcing project WP8: Website established WP9: VA LEX2 platform started WP10: TNA selection procedure established WP11: Project satisfactorily commenced
MS2	Project establishment	1 - JSI	12	WP2: user needs established (deliverable), data collection, data structuring and best practices in progress WP3: linking of resources started, common models and

Schedule of relevant Milestones

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
				protocols in progress WP4: Word Sense Disambiguation and Entity Linking started WP5: all tasks in progress WP6: skills analysis finished (deliverable), training materials and measures started WP7: all tasks in progress WP8: communication tools set up, dissemination plan finished, dissemination and monitoring in progress WP9: VA LEX1 and LEX3 platforms started WP10: TNA visiting grants in progress WP11: 1st periodic progress report delivered
MS3	Project mid-term	1 - JSI	24	WP2: data structuring finished (deliverable), data collection and best practices in progress WP3: common models and protocols finished (deliverable), linking of resources in progress WP4: all tasks in progress WP5: analytics for lexicography and crowdsourcing and gamification tasks finished (deliverables), remaining tasks in progress WP7: copyright and legal issues, data seal of compliance finished (deliverables), remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 2nd periodic progress report delivered
MS4	Project pre-final	1 - JSI	42	WP2: data collection and best practices in progress WP3: validation and quality assurance started, remaining tasks in progress WP4: Word Sense Disambiguation and Entity Linking finished (deliverable), remaining tasks in progress WP5: acquisition of lexical data from

**Schedule of relevant Milestones**

<b>Milestone number<sup>18</sup></b>	<b>Milestone title</b>	<b>Lead beneficiary</b>	<b>Due Date (in months)</b>	<b>Means of verification</b>
				corpora finished (deliverable), enrichment task in progress WP6: training and education concluded (deliverables) WP7: remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 3rd periodic progress report delivered
MS5	Project impact	1 - JSI	48	All results of the project as defined in the work plan are available (deliverables)

<b>Work package number</b> <sup>9</sup>	WP10	<b>Lead beneficiary</b> <sup>10</sup>	1 - JSI
<b>Work package title</b>	Coordination and Management		
<b>Start month</b>	1	<b>End month</b>	48

### Objectives

The objectives of this work package are to:

- initially define and implement the management structures and procedures for the project, and to coordinate and manage the project according to them. This includes administrative and financial issues, as well as project monitoring and assessment.
- formulate and implement a quality assurance procedure including a plan for assessing deliverables and the overall progress of the project.
- perform risk management including identification and assessment of risks over the entire duration of the project. This includes the definition of contingency plans to react to and tackle identified risks.

### Description of work and role of partners

**WP10 - Coordination and Management** [Months: 1-48]

JSI, LC, IVDNT, UNIROMA1, NUI GALWAY, OEAW, BCDH, UCPH

Task: T10.1 Administrative and financial management (M1-M48), leader: JSI

This task is responsible for setting up the initial project contract and to manage financial and administrative issues over the whole project duration.

Task: T10.2 Project coordination (M1-M48), leader: JSI

This task subsumes the general project coordination including coordination of the communication with the European Commission, management of the various boards defined in Section 3.2 of the proposal and handling of progress reports and deliverables. Moreover, this task has the overall responsibility of the management of the project and of its results. It also ensures that the required self-assessments and reports are being provided. Furthermore, project coordination will also be in charge of defining a risk management strategy for the project, including a contingency matrix which identifies risks, their probabilities, and corrective actions to be taken and which will be updated regularly.

### Participation per Partner

Partner number and short name	WP10 effort
1 - JSI	19.50
2 - LC	1.00
3 - IVDNT	1.00
4 - UNIROMA1	1.00
5 - NUI GALWAY	1.00
6 - OEAW	1.00
7 - BCDH	1.00
14 - UCPH	0.50
<b>Total</b>	<b>26.00</b>

**List of deliverables**

<b>Deliverable Number<sup>14</sup></b>	<b>Deliverable Title</b>	<b>Lead beneficiary</b>	<b>Type<sup>15</sup></b>	<b>Dissemination level<sup>16</sup></b>	<b>Due Date (in months)<sup>17</sup></b>
D10.1	Project manual and quality assurance plan	1 - JSI	Report	Confidential, only for members of the consortium (including the Commission Services)	12
D10.2	Data management plan	1 - JSI	Report	Public	24

**Description of deliverables**

D10.1 Project manual and quality assurance plan (M12)  
Detailed report on management procedures and Risk Management section which will be monitored throughout the project.

D10.2 Data management plan (M24)  
Preparation of pre-defined templates conformant with FAIR Data principles.

D10.1 : Project manual and quality assurance plan [12]  
Detailed report on management procedures and Risk Management section which will be monitored throughout the project.

D10.2 : Data management plan [24]  
Preparation of pre-defined templates conformant with FAIR Data principles.

**Schedule of relevant Milestones**

<b>Milestone number<sup>18</sup></b>	<b>Milestone title</b>	<b>Lead beneficiary</b>	<b>Due Date (in months)</b>	<b>Means of verification</b>
MS1	Project inception	1 - JSI	3	WP2, WP5: partners fully resourcing project WP8: Website established WP9: VA LEX2 platform started WP10: TNA selection procedure established WP11: Project satisfactorily commenced
MS2	Project establishment	1 - JSI	12	WP2: user needs established (deliverable), data collection, data structuring and best practices in progress WP3: linking of resources started, common models and protocols in progress WP4: Word Sense Disambiguation and Entity Linking started WP5: all tasks in progress WP6: skills analysis finished (deliverable), training materials and measures started WP7: all tasks in progress WP8: communication tools set up,

Schedule of relevant Milestones

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
				dissemination plan finished, dissemination and monitoring in progress WP9: VA LEX1 and LEX3 platforms started WP10: TNA visiting grants in progress WP11: 1st periodic progress report delivered
MS3	Project mid-term	1 - JSI	24	WP2: data structuring finished (deliverable), data collection and best practices in progress WP3: common models and protocols finished (deliverable), linking of resources in progress WP4: all tasks in progress WP5: analytics for lexicography and crowdsourcing and gamification tasks finished (deliverables), remaining tasks in progress WP7: copyright and legal issues, data seal of compliance finished (deliverables), remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 2nd periodic progress report delivered
MS4	Project pre-final	1 - JSI	42	WP2: data collection and best practices in progress WP3: validation and quality assurance started, remaining tasks in progress WP4: Word Sense Disambiguation and Entity Linking finished (deliverable), remaining tasks in progress WP5: acquisition of lexical data from corpora finished (deliverable), enrichment task in progress WP6: training and education concluded (deliverables) WP7: remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress



**Schedule of relevant Milestones**

<b>Milestone number<sup>18</sup></b>	<b>Milestone title</b>	<b>Lead beneficiary</b>	<b>Due Date (in months)</b>	<b>Means of verification</b>
				WP11: 3rd periodic progress report delivered
MS5	Project impact	1 - JSI	48	All results of the project as defined in the work plan are available (deliverables)

<b>Work package number</b> <sup>9</sup>	WP11	<b>Lead beneficiary</b> <sup>10</sup>	1 - JSI
<b>Work package title</b>	Ethics requirements		
<b>Start month</b>	1	<b>End month</b>	48

### Objectives

The objective is to ensure compliance with the 'ethics requirements' set out in this work package.

### Description of work and role of partners

**WP11 - Ethics requirements** [Months: 1-48]

**JSI**

This work package sets out the 'ethics requirements' that the project must comply with.

### List of deliverables

<b>Deliverable Number</b> <sup>14</sup>	<b>Deliverable Title</b>	<b>Lead beneficiary</b>	<b>Type</b> <sup>15</sup>	<b>Dissemination level</b> <sup>16</sup>	<b>Due Date (in months)</b> <sup>17</sup>
D11.1	H - Requirement No. 1	1 - JSI	Ethics	Confidential, only for members of the consortium (including the Commission Services)	48
D11.2	POPD - Requirement No. 2	1 - JSI	Ethics	Confidential, only for members of the consortium (including the Commission Services)	48
D11.3	POPD - Requirement No. 3	1 - JSI	Ethics	Confidential, only for members of the consortium (including the Commission Services)	48
D11.4	POPD - Requirement No. 4	1 - JSI	Ethics	Confidential, only for members of the consortium (including the Commission Services)	48

### Description of deliverables

The 'ethics requirements' that the project must comply with are included as deliverables in this work package.

D11.1 : H - Requirement No. 1 [48]

2.9. Copies of ethics approvals for the research with humans (if required) must be retained and made available to the EC if requested.

D11.2 : POPD - Requirement No. 2 [48]

4.1. Copies of opinion or confirmation by the competent Institutional Data Protection Officer and/or authorisation or notification by the National Data Protection Authority (which ever applies according to the Data Protection Directive (EC Directive 95/46) and the national law and after 25 May 2018, according to the new General Data Protection Regulation (Regulation EU 2016/679) must be retained and made available to the EC if requested.

D11.3 : POPD - Requirement No. 3 [48]

4.2. The applicant must specify which personal information will be collected on participants and confirm that all data collection and processing will be carried out according to the new General Data Protection Regulation (Regulation EU 2016/679) This must be supported by an Opinion of the project Data Protection Officer (if appointed). This must be retained and provided to the European Commission on request.

D11.4 : POPD - Requirement No. 4 [48]

4.5. The applicant must specify whether an institutional or organisational representatives' comments and/or opinions will be associated with their personal identity or only with their institution or organisation. All participants must be informed of this when they are recruited.

**Schedule of relevant Milestones**

Milestone number <sup>18</sup>	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS4	Project pre-final	1 - JSI	42	WP2: data collection and best practices in progress WP3: validation and quality assurance started, remaining tasks in progress WP4: Word Sense Disambiguation and Entity Linking finished (deliverable), remaining tasks in progress WP5: acquisition of lexical data from corpora finished (deliverable), enrichment task in progress WP6: training and education concluded (deliverables) WP7: remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 3rd periodic progress report delivered
MS5	Project impact	1 - JSI	48	All results of the project as defined in the work plan are available (deliverables)

### 1.3.4. WT4 List of milestones

Milestone number <sup>18</sup>	Milestone title	WP number <sup>9</sup>	Lead beneficiary	Due Date (in months) <sup>17</sup>	Means of verification
MS1	Project inception	WP1, WP10, WP4, WP7, WP8, WP9	1 - JSI	3	WP2, WP5: partners fully resourcing project WP8: Website established WP9: VA LEX2 platform started WP10: TNA selection procedure established WP11: Project satisfactorily commenced
MS2	Project establishment	WP1, WP10, WP2, WP3, WP4, WP5, WP6, WP7, WP8, WP9	1 - JSI	12	WP2: user needs established (deliverable), data collection, data structuring and best practices in progress WP3: linking of resources started, common models and protocols in progress WP4: Word Sense Disambiguation and Entity Linking started WP5: all tasks in progress WP6: skills analysis finished (deliverable), training materials and measures started WP7: all tasks in progress WP8: communication tools set up, dissemination plan finished, dissemination and monitoring in progress WP9: VA LEX1 and LEX3 platforms started WP10: TNA visiting grants in progress WP11: 1st periodic progress report delivered
MS3	Project mid-term	WP1, WP10, WP2, WP3, WP4, WP5, WP6, WP7, WP8, WP9	1 - JSI	24	WP2: data structuring finished (deliverable), data collection and best practices in progress WP3: common models and protocols finished (deliverable), linking of resources in progress WP4: all tasks in progress WP5: analytics for lexicography and crowdsourcing and gamification tasks finished (deliverables), remaining tasks in progress WP7: copyright and legal issues, data seal of compliance finished (deliverables), remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting

Milestone number <sup>18</sup>	Milestone title	WP number <sup>9</sup>	Lead beneficiary	Due Date (in months) <sup>17</sup>	Means of verification
					grants in progress WP11: 2nd periodic progress report delivered
MS4	Project pre-final	WP1, WP10, WP11, WP2, WP3, WP4, WP5, WP6, WP7, WP8, WP9	1 - JSI	42	WP2: data collection and best practices in progress WP3: validation and quality assurance started, remaining tasks in progress WP4: Word Sense Disambiguation and Entity Linking finished (deliverable), remaining tasks in progress WP5: acquisition of lexical data from corpora finished (deliverable), enrichment task in progress WP6: training and education concluded (deliverables) WP7: remaining tasks in progress WP8: remaining tasks in progress WP9: LEX1, LEX2, LEX3 infrastructures working WP10: TNA calls and visiting grants in progress WP11: 3rd periodic progress report delivered
MS5	Project impact	WP1, WP10, WP11, WP2, WP3, WP5, WP6, WP7, WP8, WP9	1 - JSI	48	All results of the project as defined in the work plan are available (deliverables)

### 1.3.5. WT5 Critical Implementation risks and mitigation actions

Risk number	Description of risk	WP Number	Proposed risk-mitigation measures
1	IPR limitations for existing lexicographic resources. (MEDIUM)	WP1, WP2, WP3, WP4	The success of the project depends on linking existing lexicographic data with open source data. If solving IPR issues will prove more difficult than expected, the consortium will assess the situation and re-allocate more resources to this task. Another mitigation measure will be to concentrate on existing open access resources and re-assess various settings in respect to IPR.
2	Issues ("bugs") with the early releases of the LEX1, LEX2, LEX3 infrastructures, resulting in negative publicity, frustration by early adopters and in the platform being abandoned. (LOW)	WP8	Extensive and regular testing and validation will detect potential problems early in the project. The first releases will not be critical for the wider acceptance. Coordinator will make sure that testers and evaluators are satisfied before the general release of services.
3	Unforeseen technical problems may not be resolved with the assigned resources. (LOW)	WP1, WP2, WP3, WP4, WP8	The partners are committed to invest a certain amount of additional own resources, since most of the addressed topics are also of high personal (and institutional) interest. In case this is not sufficient, the situation will be assessed by the Coordination of the project, in collaboration with the involved WP leaders to decide about adequate re-planning actions that ensure the overall project result.
4	Technical KPIs planned will not be met. (LOW)	WP10	Consortium will apply additional measures in development or resource allocation based on the early detected shortcomings in the development.
5	Losing a critical partner at a crucial point in the project. (LOW)	WP10	With 17 partners the consortium has been constructed with a level of redundant expertise. The most critical skills are available from at least two partners. In the rarest of cases, the missing contributions will be assessed and either some of the missing contributions will be assigned to other partners and/or a new partner with adequate competences will be identified. The consortium members have a sufficient professional network to identify an adequate new partner. The occurrence of this risk in each case requires a re-planning of the project.
6	Researchers might leave. (MEDIUM)	WP10	All work will be regularly documented and stored. Moreover, more than a single researcher per organisation will possess the necessary skills.
7	Underestimated workload in Work packages (LOW)	WP10	Person months can be reassigned from one WP to another.
8	Failure to meet Milestones (LOW)	WP10	The consortium will "freeze" certain developments so that other activities can continue on time, in order to reduce the impact of this risk. Tolerance levels will be taken into account in such decisions.

1.3.6. WT6 Summary of project effort in person-months

	WP1	WP2	WP3	WP4	WP5	WP6	WP7	WP8	WP9	WP10	WP11	Total Person/Months per Participant
1 - JSI	9	3	8	11	2	16	4	52	1	19.50		125.50
2 - LC	5	4	6	25	12	4	17	90	0	1		164
3 - IVDNT	20	4	2	6	4	7	3	0	1	1		48
4 - UNIROMA1	0	10	30	17	0	0	2	0	0	1		60
5 - NUI GALWAY	2	24	0	6	5	2	3	0	0	1		43
6 - OEAW	3	16	3	4	4	5	22	0	1	1		59
7 - BCDH	11	0	4	0	24	3	2	0	1	1		46
8 - MTANYTI	5	4	0	2	7	3	5	0	1	0		27
9 - IBL	11	6	5	2	0	0	2	0	1	0		27
10 - FCSH-UNL	3	2	0	0	5	0	2	0	0	0		12
11 - K Dictionaries	7	6	5	8	0	0	2	0	1	0		29
12 - CNR	2	7	0	0	4	0	4	0	1	0		18
13 - DSL	7	7	3	0	0	2	2	0	1	0		22
14 - UCPH	0	4	4	0	0	4	2	0	3	0.50		17.50
15 - UT	6	0	0	0	3	0	1	8	1	0		19
16 - EKI	9	0	0	3	0	0	2	0	1	0		15
17 - RAE	3	0	0	0	0	0	2	0	1	0		6
· RAEG	0	0	0	0	0	0	0	0	0	0	0	0
<b>Total Person/Months</b>	103	97	70	84	70	46	77	150	15	26		738

### 1.3.7. WT7 Tentative schedule of project reviews

<b>Review number <sup>19</sup></b>	<b>Tentative timing</b>	<b>Planned venue of review</b>	<b>Comments, if any</b>
RV1	26		



1.3.8. WT8 Summary of transnational / virtual access provision per installation

Access provider short name	Short name of infrastructure	Installation		Installation country code <sup>21</sup>	Type of access <sup>22</sup>	Unit of access	Unit cost (€)	Min. quantity of access to be provided	Access costs <sup>23</sup>		Estimated number of users	Estimated number of projects
		number <sup>20</sup>	Short name						On the basis of UC	As actual costs		
1 - JSI	LEX1	1	LEX1	SI	VA					115000		
1 - JSI	ELEXIS-SI	1	ELEXIS-SI	SI	TA-ac	Day		42.0		7875	3	3
2 - LC	LEX2	1	LEX2	CZ	VA					193750		
3 - IVDNT	ELEXIS-NL	1	ELEXIS-NL	NL	TA-ac	Day		42.0		8750	3	3
6 - OEAW	ELEXIS-AT	1	ELEXIS-AT	AT	TA-ac	Day		42.0		6875	3	3
7 - BCDH	ELEXIS-RS	1	ELEXIS-RS	RS	TA-ac	Day		42.0		6300	3	3
8 - MTANYTI	ELEXIS-HU	1	ELEXIS-HU	HU	TA-ac	Day		42.0		3500	3	3
9 - IBL	ELEXIS-BG	1	ELEXIS-BG	BG	TA-ac	Day		42.0		5715	3	3
11 - K Dictionaries	ELEXIS-IL	1	ELEXIS-IL	IL	TA-ac	Day		42.0		6250	3	3
13 - DSL	ELEXIS-DK	1	ELEXIS-DK	DK	TA-ac	Day		42.0		9250	3	3
15 - UT	LEX3	1	LEX3	DE	VA					41250		
15 - UT	ELEXIS-DE	1	ELEXIS-DE	DE	TA-ac	Day		42.0		7375	3	3
16 - EKI	ELEXIS-EE	1	ELEXIS-EE	EE	TA-ac	Day		42.0		3750	3	3
17 - RAE	ELEXIS-SP	1	ELEXIS-SP	ES	TA-ac	Day		42.0		7000	3	3

### **1. Project number**

The project number has been assigned by the Commission as the unique identifier for your project. It cannot be changed. The project number **should appear on each page of the grant agreement preparation documents (part A and part B)** to prevent errors during its handling.

### **2. Project acronym**

Use the project acronym as given in the submitted proposal. It can generally not be changed. The same acronym **should appear on each page of the grant agreement preparation documents (part A and part B)** to prevent errors during its handling.

### **3. Project title**

Use the title (preferably no longer than 200 characters) as indicated in the submitted proposal. Minor corrections are possible if agreed during the preparation of the grant agreement.

### **4. Starting date**

Unless a specific (fixed) starting date is duly justified and agreed upon during the preparation of the Grant Agreement, the project will start on the first day of the month following the entry into force of the Grant Agreement (NB : entry into force = signature by the Commission). Please note that if a fixed starting date is used, you will be required to provide a written justification.

### **5. Duration**

Insert the duration of the project in full months.

### **6. Call (part) identifier**

The Call (part) identifier is the reference number given in the call or part of the call you were addressing, as indicated in the publication of the call in the Official Journal of the European Union. You have to use the identifier given by the Commission in the letter inviting to prepare the grant agreement.

### **7. Abstract**

### **8. Project Entry Month**

The month at which the participant joined the consortium, month 1 marking the start date of the project, and all other start dates being relative to this start date.

### **9. Work Package number**

Work package number: WP1, WP2, WP3, ..., WPn

### **10. Lead beneficiary**

This must be one of the beneficiaries in the grant (not a third party) - Number of the beneficiary leading the work in this work package

### **11. Person-months per work package**

The total number of person-months allocated to each work package.

### **12. Start month**

Relative start date for the work in the specific work packages, month 1 marking the start date of the project, and all other start dates being relative to this start date.

### **13. End month**

Relative end date, month 1 marking the start date of the project, and all end dates being relative to this start date.

### **14. Deliverable number**

Deliverable numbers: D1 - Dn

### **15. Type**

Please indicate the type of the deliverable using one of the following codes:

- R Document, report
- DEM Demonstrator, pilot, prototype
- DEC Websites, patent filings, videos, etc.
- OTHER
- ETHICS Ethics requirement
- ORDP Open Research Data Pilot

### **16. Dissemination level**

Please indicate the dissemination level using one of the following codes:

- PU Public
- CO Confidential, only for members of the consortium (including the Commission Services)
- EU-RES Classified Information: RESTREINT UE (Commission Decision 2005/444/EC)
- EU-CON Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC)
- EU-SEC Classified Information: SECRET UE (Commission Decision 2005/444/EC)

**17. Delivery date for Deliverable**

Month in which the deliverables will be available, month 1 marking the start date of the project, and all delivery dates being relative to this start date.

**18. Milestone number**

Milestone number: MS1, MS2, ..., MSn

**19. Review number**

Review number: RV1, RV2, ..., RVn

**20. Installation Number**

Number progressively the installations of a same infrastructure. An installation is a part of an infrastructure that could be used independently from the rest.

**21. Installation country**

Code of the country where the installation is located or IO if the access provider (the beneficiary or linked third party) is an international organization, an ERIC or a similar legal entity.

**22. Type of access**

- VA if virtual access,
- TA-uc if trans-national access with access costs declared on the basis of unit cost,
- TA-ac if trans-national access with access costs declared as actual costs, and
- TA-cb if trans-national access with access costs declared as a combination of actual costs and costs on the basis of unit cost.

**23. Access costs**

Cost of the access provided under the project. For virtual access fill only the second column. For trans-national access fill one of the two columns or both according to the way access costs are declared. Trans-national access costs on the basis of unit cost will result from the unit cost by the quantity of access to be provided.

Date	Section	Change
11.9.2017	5.1.	Inclusion of third country Partner ethical statement.
4.10.2017	All	Removal of duplicated information present in Participant Portal.
10.10.2017	4.2	Third party (for <i>Real Academia Española</i> ) added
22.10.2017	3.4	Table 3.4b: six tables added, changes in existing three tables
21.11.2017	3.4	Table 3.4b: direct costs explained for JSI, LC, TCDH
4. 12. 2017	4.2.	Subcontracting removed for RAE.

1	Excellence	4
1.1	Objectives	4
1.2	Relation to the work programme	5
1.2.1	Networking activities	6
1.2.2	Trans-national access or virtual access activities	7
1.2.3	Joint research activities	8
1.3	Concept and methodology	8
1.3.1	Concept	8
1.3.2	Methodology	10
1.4	Ambition	22
1.4.1	Progressing the state-of-the-art in lexicography	22
1.4.2	Progressing the state-of-the-art in NLP, SW and AI	24
2	Impact	24
2.1	Expected impacts	24
2.2	Measures to maximise impact	26
2.2.1	Dissemination and exploitation of results	26
2.2.2	Communication activities	33
3	Implementation	35
3.1	Work plan — Work packages, deliverables	35
3.1.1	Dependency between work packages	36
3.1.2	Timing of Work Packages (Gantt Chart) and milestones	36
3.2	Management structure, milestones and procedures	39
3.3	Consortium as a whole	44
3.4	Resources to be committed	47
3.4.1	Table 3.4b: ‘Other direct cost’ items (travel, equipment, other goods and services)	48
3.5	References and bibliography	53
4	Section 4: Members of the consortium	54
4.1	4.1. Participants (applicants)	54
1.	“Jožef Stefan” Institute, Ljubljana, Slovenia (JSI)	54
2.	Lexical Computing CZ s.r.o. (LC)	58
3.	Instituut voor de Nederlandse Taal (Dutch Language Institute, INT)	61
4.	Sapienza University of Rome, UNIROMA1	63
5.	National University of Ireland Galway (NUIG)	66
6.	Austrian Academy of Sciences (OEAW)	69
7.	Centar za digitalne humanističke nauke (Belgrade Center for Digital Humanities, BCDH)	72
8.	RILMTA Research Institute for Linguistics Hungarian Academy of Sciences	74
9.	Institute for Bulgarian Language “Prof. Lyubomir Andreychin” (IBL)	77
10.	Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa - Portugal	79

11.	K Dictionaries (KD)	82
12.	CNR-ILC - Istituto di Linguistica Computazionale "A. Zampolli" - CNR - Italy	84
13.	DSL – The Society for Danish Language and Literature	87
14.	Centre for Language Technology at University of Copenhagen (UCPH)	89
15.	Trier Center for Digital Humanities (TCDH)/Trier University - Germany	91
16.	Institute of the Estonian Language	95
17.	Real Academia Española (RAE)	97
4.2	4.2. Third parties involved in the project (including use of third party resources)	99
5	Section 5: Ethics and Security	100
5.1	5.1 Ethics	100
5.2	5.2 Security	100

# 1 EXCELLENCE

---

## 1.1 OBJECTIVES

Reliable and accurate information on word meaning and usage is of crucial importance in today’s information society. The most consolidated and refined knowledge on word meanings can traditionally be found in dictionaries – monolingual, bilingual or multilingual.

Dictionaries are not only vast, systematic inventories of information on words, they are also important as cultural and historical artefacts. In each and every European country, elaborate efforts are put into the development of lexicographic resources describing the language(s) of the community. Although confronted with similar problems relating to technologies for producing and making these resources available, cooperation on a larger European scale has long been limited.

Consequently, the lexicographic landscape in Europe is currently rather heterogeneous. On the one hand, it is characterised by stand-alone lexicographic resources, which are typically encoded in incompatible data structures due to the isolation of efforts, prohibiting reuse of this valuable data in other fields, such as natural language processing, linked open data and the Semantic Web, as well as in the context of digital humanities. On the other hand, there is a significant variation in the level of expertise and resources available to lexicographers across Europe. This forms a major obstacle to more ambitious, innovative, transnational, data-driven approaches to dictionaries, both as tools and objects of research.

In 2013, the European lexicographic community was brought together for the first time in the European Network of e-Lexicography (ENeL) COST action ([www.elexicography.eu](http://www.elexicography.eu)). This initiative was set up to improve the access for the general public to scholarly dictionaries and make them more widely known to a larger audience. This networking initiative, which will end in October 2017, started with 34 members from 20 countries but currently counts 285 members from 31 countries. In the context of this network, a clear need has emerged for a broader and more systematic exchange of expertise, for the establishment of common standards and solutions for the development and integration of lexicographical resources, and for broadening the scope of application of these high quality resources to a larger community, including the Semantic Web, artificial intelligence, NLP and digital humanities.

ELEXIS aims to develop an infrastructure which will:

Objective 1	foster <b>cooperation and knowledge exchange</b> between different research communities in lexicography in order to bridge the gap between lesser-resourced languages and those with advanced e-lexicographic experience;
Objective 2	establish <b>common standards and solutions</b> for the development of lexicographic resources;
Objective 3	develop strategies, tools and standards for <b>extracting, structuring and linking</b> of lexicographic resources;
Objective 4	<b>enable access</b> to standards, methods, lexicographic data and tools for scientific communities, industries and other stakeholders;

Objective 5	promote an <b>open access culture</b> in lexicography, in line with the European Commission <i>Recommendation on access to and preservation of scientific information</i> .
-------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------

ELEXIS is based on the conviction that lowering the barrier for retrieving and analysing multilingual lexicographic data across Europe cannot be accomplished in the long term without lowering the barrier for providing lexicographic data to research infrastructures. As a result, the following impacts will be achieved:

- efficient (open) access to high quality lexicographic data for researchers, institutions and stakeholders from different fields;
- a common platform for building, sharing and exploiting knowledge and expertise between lexicography and computational linguistics which will facilitate cross-disciplinary fertilisation and a wider sharing of information, knowledge and technologies across and within these fields. The platform will thus bridge the gap between lesser-resourced languages and those with advanced e-lexicographic and/or computational linguistic experience;
- the creation of a scalable, multilingual and multifunctional, language resource. By integrating and linking lexical content and interlinking it with other structured or unstructured data - corpora, multimodal resources, etc. - on any level of lexicographic description, a multilingual and multifunctional language resource will be created incrementally enriching the available information;
- a new type of lexicography that no longer views languages as isolated entities, but fully embraces the pan-European nature of much of the vocabularies of languages spoken in Europe;
- the inter- and multidisciplinary nature of lexical data will help researchers ask new questions and pursue new avenues of research.

## 1.2 RELATION TO THE WORK PROGRAMME

The ELEXIS infrastructure is proposed under the »Integrating Activities for Starting Communities« call, with **the field of lexicography** addressed in the proposal. Activities are directed at all types of lexicographic communities throughout Europe and globally. Lexicography is understood as the art or craft of compiling, writing and editing dictionaries for human users and their scholarly research. With the arrival of the digital age, the field has undergone significant changes, resulting in the emergence of e-lexicography. There is a general trend to move away from paper, which can be seen, on the one hand, in the numerous retrodigitisation efforts and, on the other hand, in the appearance of digitally-born online dictionaries.

Lexicographic communities differ significantly depending on the status of their language (official, minority, regional, etc.), number of speakers, scope of description (monolingual, multilingual, general, specialised, etc.), purpose (descriptive, normative), and other criteria. It is perhaps indicative of the field that parts of the lexicographic community involved in lexical description of individual languages (monolingual lexicography) or in language standardisation did not show the need for large-scale pan-European collaboration before the digital age. In Europe, organised collaboration was more or less limited to the European Association of Lexicography (EURALEX) which was established in 1987 and organises a biannual conference. Language standardisation bodies are organised in the EFNIL forum (European Federation of National Institutions for



Language) dedicated to the exchange of information about their work and to gathering and publishing information about language use and language policy within the European Union.

For languages with the culture of language standardisation and a relatively small number of speakers, it is common for national language institutes or similar institutions to conduct the description and standardisation work. Typically, they are concerned with developing their own isolated language description portals dedicated to individual languages. One of the missions of the European Network of e-Lexicography COST Action was to identify and make an inventory of existing portals and lexicographic resources. The work carried out in the four working groups in this action showed that an enhanced form of collaboration is needed, which has led to forming the ELEXIS proposal.

As an integration activity, ELEXIS will combine networking activities, access activities and foster joint research activities.

### 1.2.1 Networking activities

ELEXIS will **consolidate the results of the ENeL COST action** and strengthen the lexicographic community in Europe, bringing all lexicographic institutions and experts throughout Europe together. By lowering the bar for using, analysing and exchanging lexical data, a carefully planned set of training and outreach activities within ELEXIS will lead to closer and more productive interactions among researchers working with lexical data across Europe.

In order to educate a new generation of researchers who understand the full potential of digital research infrastructures to transform their research; who optimally exploit the existing state-of-the-art tools; and who are able to create open, standards-compliant lexical datasets that can be fed back into the infrastructures and shared with other researchers, ELEXIS will:

- create robust **documentation, guidelines, collections of best practices** and **multilingual lexical data** in order to promote clearly defined workflows for producing, describing and annotating lexicographic research outputs in accordance with international standards and interoperable formats;
- organise **workshops, seminars, and conferences** in order to support dynamic knowledge exchange across national borders, which will be about 1) harmonising access to resources by developing joint conceptual and metadata models for advanced data exchange, large-scale data mining and enhanced discoverability; and 2) preventing a duplication of services by fostering better coordination of activities between national and transnational efforts in the field of lexicography;
- provide **online training modules** with suggested ECTS credits that can be easily incorporated into existing curricula, in partnership with #dariahTeach, an Erasmus+ funded Strategic Partnership for developing, maintaining and delivering modular, extensible and localisable training materials in Digital Humanities and related fields;
- set the groundwork and launch a campaign for the establishment of an academically recognised and industry-supported **data seal of compliance for lexicographic data** (similar to <http://datasealofapproval.org/en/> but focusing exclusively on access to high-quality lexical data);
- establish an **international forum** with both academic and industry stakeholders to explore the **socio-economic challenges** of lexicographic data as (linked) **open data** and provide innovative solutions for bridging the gap which currently divides academic from commercial lexicography, both in terms of copyright issues and sustainability models;

- pursue a dynamic **communication** strategy using a dedicated website, newsletters and social media in order to **disseminate project outputs** to a wide circle of stakeholders within academia, the industry as well as the general public;
- increase the wider social impact of the digital research infrastructure by **supporting innovative interdisciplinary approaches and partnerships** in analysing, visualising and mapping lexical data in cooperation with various stakeholders from the educational and the cultural heritage sectors as well as the media.

## 1.2.2 Trans-national access or virtual access activities

### *Virtual Access*

The infrastructure will provide access to lexicographic data and tools. Lexicographic data that will be produced or collected under the project will be made available to lexicographers and to the wider research community, to enable the creation of novel lexicographic resources, both for human users and for machines.

ELEXIS will provide virtual access to:

- existing **modern or historical lexicographic data** from national language centres dedicated to description of individual languages (the list of available lexicographic resources on p. 21)
- new **lexicographic data linked** by using tools developed in the project (ELEXIS matrix dictionary)
- **conversion tools** harmonising existing lexicographic data through common standardised data models
- **tools enabling direct linking** between two or more lexicographic resources or via BabelNet or other semantic network
- **word sense disambiguation** and **text enrichment tools** enabling the development of new lexicographic resources from corpora
- **extraction tools** producing dictionary content from corpora in an automated process
- online tools supporting lexicographic **workflow, crowdsourcing and gamification**

ELEXIS will build on top of two ESFRI infrastructures, CLARIN and DARIAH, and will also use the existing Linked Data (<http://linkeddata.org/>) and Linguistic Linked Open Data (<http://www.linguistic-lod.org/>) clouds, developed (among others) by two of the ELEXIS partners. Through technical links the individual lexicographic resources will be linked together to create an open lexicographic network of linked data.

### *Trans-National Access*

During the lifetime of the project ELEXIS will organise trans-national calls enabling researchers (a) to work with data with restricted access at host institutions; (b) to gain knowledge and expertise in close contact with lexicographers and experts in NLP and artificial intelligence. One of the reasons for limited accessibility of lexicographic data outside institutions which are the creators and copyright holders of the data is the effort needed for their compilation, which necessitates tighter control over access and availability of raw data. Trans-national activities represent one of the mechanisms of ELEXIS to enable access to such content to researchers from other institutions or countries. However, the results of research conducted in trans-national

activities will be available under open access licences according to the rules of the call enabling the international community to familiarise itself with previously inaccessible resources.

### 1.2.3 Joint research activities

Research activities in ELEXIS will focus on:

#### *A. tools and methods for enabling linked lexicographic resources*

Current lexicographic resources, both modern and historical, have different levels of structure and are not equally suitable for application in advanced NLP technologies. The project will develop strategies, tools and standards for extracting, structuring and linking the high quality semantic data from lexicographic resources and make them available to the Linked (Open) Data family.

We will also work on interlinking lexical content with other structured or unstructured data – corpora, multimodal resources, etc. – on any level of lexicographic description: semantic, syntactic, collocational, phraseological, etymological, translation equivalents, examples of usage, etc.

By creating an integrated, linked and interlinked resource, a huge amount of high quality lexical data will not only become available to the linguistic, NLP and Semantic Web community, it will also facilitate cutting edge research in digital humanities.

#### *B. tools and methods to support innovative e-lexicography*

ELEXIS will support novel lexicography by providing lexicographers with tools and methods that help them create new resources. Using machine learning, data mining and information extraction techniques proto-dictionary content will be produced in an automated way. The automatically extracted data can then be used as a starting point for further processing either in a more traditional lexicographic workflow or through crowdsourcing platforms, making it easier to create new resources. This novel approach can be applied to any language for which there is data available on the web. This is particularly important for under-resourced languages with out-dated or non-existent language descriptions, enabling researchers and the general public to learn about semantic, grammatical or other aspects of lexicographic description benefiting from the technology derived from language communities with advanced lexicographic descriptions.

## 1.3 CONCEPT AND METHODOLOGY

### 1.3.1 Concept

As a new infrastructure, ELEXIS will bring together research communities and consortium partners working in different fields, in order to support the community working in the emerging field of e-lexicography. In particular, ELEXIS will build on the existing expertise and knowledge of partners in the fields of **lexicography**, **computational linguistics** and **artificial intelligence** in an interdisciplinary effort to make existing lexicographic resources available on a significantly higher level compared to their availability as stand-alone resources, which is the current state of affairs.

These resources are in fact results of long-term projects in which literally thousands of person years were and continue to be dedicated to their compilation in national and regional projects, and in most cases they represent the most consolidated and refined knowledge on word meanings in individual languages. A tremendous effort is needed for their compilation, and this implies the necessity to control the contents in

order to ensure both the continuation of consistent language description and maximum quality of the results. Furthermore, and resulting from current isolation of efforts, these resources are typically encoded in incompatible data structures. Both issues contribute to the fact that the data from these resources is lost for extensive, interoperable and generally accessible computer use.

On the other hand, the language technology (LT) community, for their part, created an overwhelming number of different types of lexical resources over the last thirty years, which are used for natural language processing tasks. These include corpora, lexicons, glossaries (used in machine translation), machine-readable dictionaries, lexical databases, and many others. One of the important issues that will be addressed by ELEXIS is the fact that the impressive results of the LT community have rarely found their way into the practical work of creating lexicographic resources in the past. This can be largely attributed to the lack of a common platform for building, sharing and exploiting knowledge and expertise between computational linguistics and lexicography, which is one of the goals of the proposed infrastructure.

#### *A. Supporting lexicographic process and language description*

To support the lexicographic process and to contribute to lexicographically-oriented language description ELEXIS will:

- develop methods and tools for the automatic processing and extraction of data from corpora and other (multimodal) resources for lexicographic purposes;
- develop methods and tools for the inclusion of extracted data into interlinked (open) lexicographic data;
- develop methods, guidelines and tools enabling the use of crowdsourcing and citizen science in the lexicographic process;
- elaborate on the guidelines and solutions for handling copyright and authorship protection to enable inclusion of extracted data into the lexicographic workflow.

#### *B. Supporting natural language processing*

To support the natural language processing community, several steps are needed to make existing lexicographic resources globally available. ELEXIS will:

- develop methods, guidelines and tools for harmonisation of dictionary formats, building on the existing standards within the lexicographic and NLP community;
- develop methods and tools for automatic segmentation and identification of dictionary structure, enabling interlinking of dictionary content;
- develop methods and tools for interlinking, maintenance, reuse, sharing and distribution of existing lexicographic resources;
- define evaluation and validation protocols and procedures (lexicographic data seal of compliance);
- elaborate on the guidelines and solutions for handling copyright and authorship protection to enable open access to lexicographic data in LOD framework.

Therefore, in contrast with previous more NLP-oriented efforts, ELEXIS will develop methods and tools to produce collections of structured proto-lexicographic data in an automated process, using machine learning,

data mining and information extraction techniques, where the extracted data can be used as a starting point for further processing either in the traditional lexicographic process or through crowdsourcing platforms. To some extent, it could also be used for immediate visualisation for the general public interested in lexical behaviour of words.

### 1.3.2 Methodology

Lexicography as a field has a long tradition of refining semantic description of individual languages in comprehensive monolingual dictionaries, or performing detailed contrastive analysis between two or more languages in bilingual and multilingual dictionaries. However, these resources are currently not used within existing and emerging language technologies. They are almost completely absent in linked (open) data clouds and Semantic Web technologies, and are in fact “digitally invisible”.

In the last decade the new field of **e-lexicography** emerged, which can be seen in initiatives such as the ENEL COST action (<http://www.elexicography.eu/>), the eLex conference series (<https://elex.link/>), or the Globalex workshop at LREC 2016 (<http://ailab.ijs.si/globalex/>). Globalex is the first initiative which includes all continental lexicographic associations: EURALEX, ASIALEX, AFRILEX, AUSTRALEX and the Dictionary Society of North America, as well as eLex. The field of e-lexicography is dedicated to creating digitally-born dictionaries defined as lexical resources intended for human users but intentionally moving away from the paper medium and exploring the almost infinite possibilities of the new digital environment, with a view to take human-oriented lexical description to entirely different levels. In this context, machine learning, data mining and other computational techniques are starting to find their way into lexicography. Combining both traditional lexicographic knowledge and expertise with computational linguistics, while engaging also wider language communities in the process, creates huge potential for the development of the field.

#### *Lexicography and »semantic bottleneck«*

Lexicographic resources contain quality information about general vocabulary and more difficult types of language phenomena such as highly polysemous words or semantically opaque multi-word expressions (idioms, phraseology), which are rather inconsistently covered in LT-oriented resources. These phenomena represent a bottleneck in achieving precision and computational efficiency of NLP applications. This can be seen also from efforts such as PARSEME COST action (<http://typo.uni-konstanz.de/parseme/>) which is devoted to the role of multi-word expressions in parsing. Word sense disambiguation as part of content analytics, text understanding and computer reasoning remains another complex task for computational processing of text, and is still largely unsolved, especially for languages other than English. Typically, resources such as Wikipedia, Wiktionary, wordnets or framenets are used for word sense disambiguation tasks, collected in the (L)L(O)D cloud (<http://linkeddata.org/>, <http://www.linguistic-lod.org/>). Knowledge bases and complementary applications such as BabelNet (<http://babelnet.org/>), Babelfy (<http://babelfy.org/>), Cyc (<http://sw.opencyc.org/>) or wikifiers (<http://www.wikifier.org/>) have been developed to enrich text processing with semantic information. ELEXIS proposes enriching the existing linked data clouds and knowledge bases with data available in existing and new lexicographic resources, which are currently not used for solving these tasks.

#### *Standards in lexicography and NLP*

There are several reasons for the negligible incorporation of lexicographic data in LT so far. The first is almost non-existent interoperability and use of common standards in lexicography. In past decades there were several important efforts to harmonise and standardise linguistic resources, including lexicographic

resources. These include first initiatives such as EAGLES/ISLE, Multext(-East), PAROLE, SIMPLE, CONCEDE etc. in the 1990s. From these efforts, standards emerged such as

- *Text Encoding Initiative* (TEI - <http://www.tei-c.org/>),
- *Lexical Markup Framework* (LMF - <http://www.lexicalmarkupframework.org/>),

and others, most of them under the umbrella of the *Terminology and other language and content resources* ISO/TC 37 standard. The standardisation process was much more successful with resources directly dedicated to computer use, such as corpora, lexicons, lexical databases, wordnets, ontologies etc., but standards were less successful in case of lexicographic resources initially intended for human users. Therefore, although a huge number of lexicographic resources exist for various languages, as results of long-term projects in which a huge amount of work was dedicated to their compilation, they are typically encoded in incompatible data structures, and the data from these resources is lost for extensive, interoperable and generally accessible computer use.

#### *Availability of lexicographic data*

Although early digitisation projects involving lexicographic resources date back to the 1980s (Boguraev and Briscoe 1989), or in case of English even the 1960s (Urdang 1966), and the 1990s saw massive digitisation of existing dictionaries, including works like the Oxford English Dictionary, general (open) access to lexicographic data is extremely limited. The main reason for this is the massive effort necessary to compile such resources, which are usually produced either by national language institutions, or by commercial companies in the case of “commercial languages” with a sufficient number of speakers. The effort needed consequently implies the necessity to control the contents, resulting in the need to resolve intellectual property right issues before this data can be included in open access infrastructures. The ELEXIS infrastructure will dedicate serious efforts to solve IPR issues related to lexicographic data and enable their integration as linked data.

In the last decade, initiatives promoting open access to the results of publicly funded projects (Open Research Data Pilot etc.) and the increasing wealth of (open) data available on the Web (Wikipedia, Wiktionary etc.), also instigated new trends within lexicography, particularly the move towards e-lexicography. This new trend is not yet supported by an infrastructure where quality semantic data from dictionaries could be linked, shared, distributed and stored on a massive scale. Therefore, the objective of contributing quality semantic data in the digital age means that the proposed project will make it possible for existing lexicographic resources to be included seamlessly into the Linked (Open) Data family (cf. Figure 2).

#### *Virtuous cycle of e-lexicography*

As was established in surveys conducted within the ENeL action (Tiberius et al. 2015), the results of the LT work are rarely used in lexicography, which is one of the important issues addressed by the ELEXIS infrastructure. This can be largely attributed also to the lack of an infrastructure enabling sharing knowledge and expertise between LT and lexicography. Ideally, the part of the virtuous cycle starting from NLP towards lexicography will produce proto-dictionary content in a completely automatic process with the use of machine learning, data mining and information extraction techniques focusing on massive amounts of data in various modalities available on the Web.

ELEXIS will develop methods and tools to produce such collections of structured data in an automated process where the data can be used:

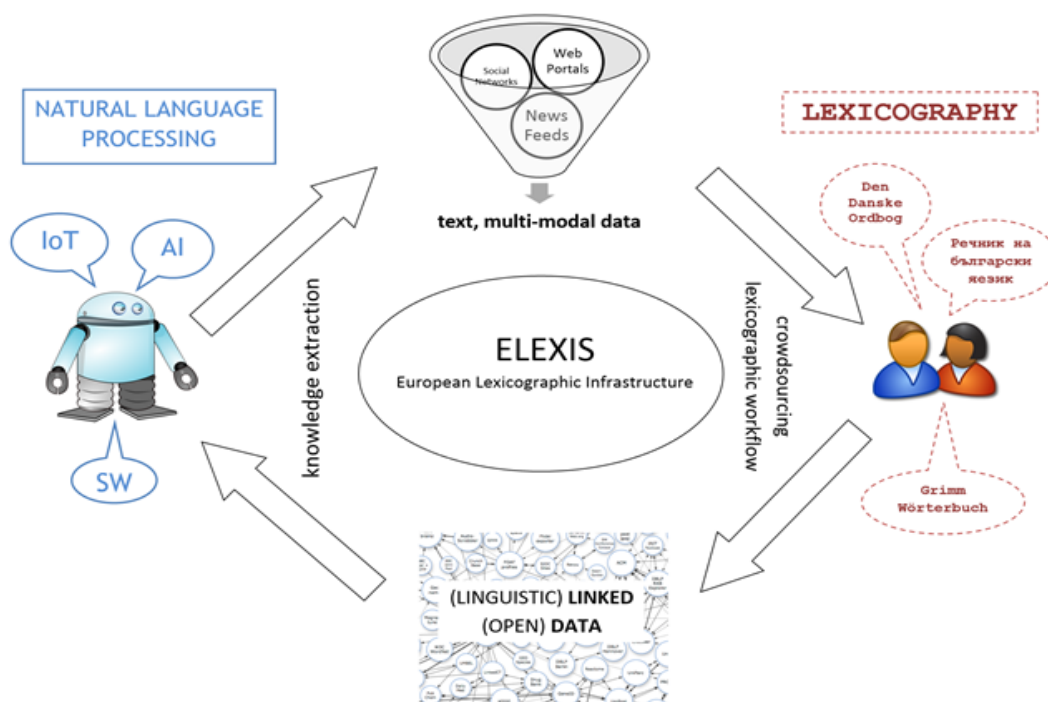
1. as a starting point for further processing of the collected material either by



- a. **traditional lexicographic process** or
- b. through **crowdsourcing** platforms;

2. for immediate visualisation for human users interested in lexical behaviour of words.

Examples of existing applications, services and tools are: (1) extraction of lexico-grammatical data with the Sketch Engine by Lexical Computing (<http://www.sketchengine.co.uk/>) or Kookkurrenzdatenbank CCDB (<http://corpora.ids-mannheim.de/ccdb/>), (2) definition extraction tools (cf. [RANLP 2009](#), International workshop on definition extraction), LT4eL project with local grammars to extract definition patterns (<http://www.lt4el.eu/>), (3) extraction of good (dictionary) examples: GDEX, a tool in the Sketch Engine, which is designed to help lexicographers identify dictionary examples by ranking sentences according to their dictionary potential, (4) multilingual data: services like Linguee (<http://www.linguee.com/>) using (free) parallel corpora, K Dictionaries (<http://www.kdictionaries-online.com/>) with the “English Multilingual Dictionary”, (5) other modalities: adding images in dictionary aggregators like Wordnik (<http://www.wordnik.com>) from Flickr, Wikimedia Commons, Europeana and other publicly available sources, audio from speech corpora, e.g. Gos speech corpus of Slovene (<http://www.korpus-gos.net/>), Forvo (<http://www.forvo.com/>) etc., video material from services such as Videolectures.net (<http://videolectures.net/>) with presentations as textual data linked with the relevant video segments, or



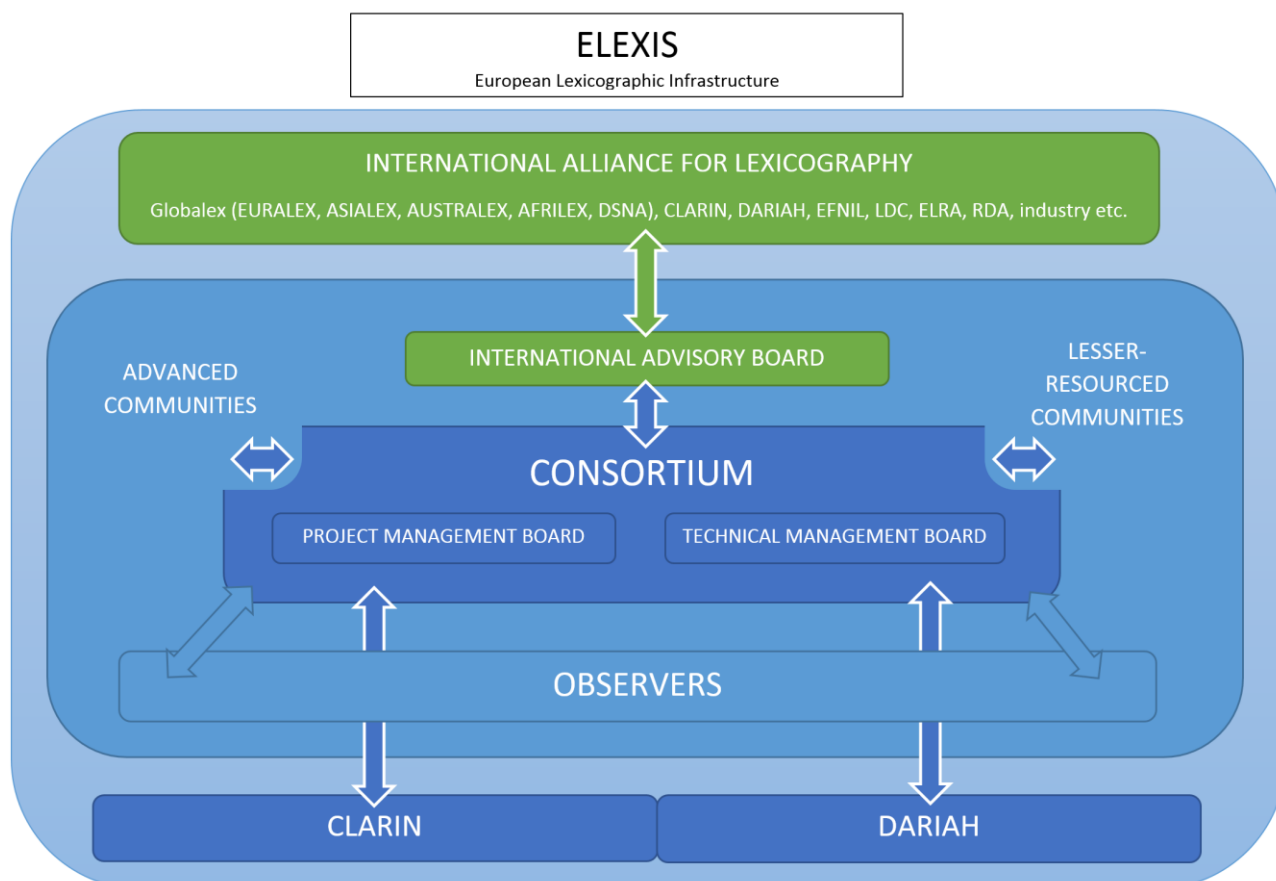
Picture 1: Virtuous cycle of e-lexicography

streaming as a complement to the static data, from news feed service (<http://newsfeed.ijs.si/>). One of the missions of ELEXIS is to enable an integration of (big) data in different modalities into the lexicographic process, prepared and visualised for human end users.

*Fostering cooperation and knowledge exchange in lexicography*

The objective of ELEXIS is to foster cooperation and knowledge exchange between different research communities in lexicography in order to bridge the gap between lesser-resourced languages and those with

advanced e-lexicographic experience, and one of the impacts of ELEXIS is defined as the emergence a new type of lexicography that no longer views languages as isolated entities, but fully embraces the pan-European nature of languages spoken in Europe. This ambition extends also to the global level. ELEXIS plans reflect this goal with an inclusive multi-layered organisation that aims at engaging different user groups with various levels of intensity during the project.



Picture 2: ELEXIS organisation

Elements of the structure consist of **ELEXIS consortium partners** as the core group responsible for the development of the infrastructure. As shown in Picture 2, another organisational layer are **observing institutions** that will be directly included in outreach and dissemination activities through various channels described in the Communication activities section. The central group of institutions that fall under the observer category are those producing quality lexicographic data and resources, filtered by the criteria of inclusion in the European Dictionary Portal developed by the European Network of e-Lexicography COST action: <http://www.dictionaryportal.eu/en/catalog/>. Typically but not exclusively, these institutions include (European) national language institutes, large dictionary publishers and other prominent producers of lexicographic data.

The broadest and also less defined user group consists of **lesser-resourced communities** and **advanced communities** working in lexicography. In practical terms, these include all researchers interested in lexicography who will gain access to newly developed data, tools and services through ELEXIS virtual access activities. The consortium itself reflects this division as it includes partners working with lesser-resourced languages (Serbian, Slovenian, Irish, etc.) and from advanced communities (English, German, Dutch, Italian etc.). The infrastructure will provide networking activities (training, online training material, conferences,



workshops etc.) to enable lesser-resourced languages to benefit from the results of the project, building on the expertise and available data in advanced communities.

ELEXIS consortium partners will be advised by an **International Advisory Board (IAB)**. Members of IAB will include appointed representatives of five continental societies and associations for lexicography (cf. letters of support in section 3.5) and top experts in both relevant fields, lexicography and NLP, from academia and industry (cf. list in section 3.3).

Through the International Advisory Board, and using its outreach and community building activities, ELEXIS will strive to form an **International Alliance for Lexicography**, which will include stakeholders from various fields. The alliance will be formed as a loose organisation dedicated to the advancement of lexicography in the digital age, with the ELEXIS International Advisory Board serving as the initial governing body. We expect the alliance to be joined by the five continental lexicographic associations who form the Globalex initiative, standardisation bodies organised in EFNIL (European Federation of National Institutions for Language), data providers such as LDC (Linguistic Data Consortium), ELRA (European Language Resources Association), RDA (Research Data Alliance), and representatives of both language technology (Language Technology Industry Association - LT Innovate) as well as lexicography and language learning industries (e.g. Oxford University Press, MacMillan Publishers - cf. letters of support). The aim of the alliance is to consolidate the field of lexicography and enable its transition to the digital environment, bringing together all stakeholders interested in language description and semantic data.

#### *Integrating existing infrastructures*

ELEXIS is ideally suited to serve as a hub of integration activities between CLARIN and DARIAH: dictionaries are essential language resources whose quality, reliability and coverage can be vastly improved by means of harmonising formats and optimising points of infrastructural access. At the same time, however, dictionaries are more than simply tools: they are also objects of humanities research in their own right. Humanities scholars study dictionaries in terms of their cultural and ideological values, or their role in language standardisation and nation-building, to name just a few different perspectives. European research infrastructures should cater equally to lexicographers, linguists and humanities scholars who write, use or study dictionaries. ELEXIS integrating access activities will help CLARIN and DARIAH achieve precisely that.

ELEXIS as a new infrastructure builds upon the existing tools and services of CLARIN and/or DARIAH with the goal of achieving something that neither infrastructure can at the moment provide on its own: a concerted pan-European effort aimed at combining and advancing the state-of-the-art in three distinct fields -- lexicography, NLP and digital humanities. CLARIN already has a leading role in providing language resource repositories, linguistic annotation pipelines and federated search facilities, whereas DARIAH is a leader in facilitating long-term access to and use of arts and humanities research data. By creating a common platform for building, sharing and exploiting high-quality, multilingual lexical data, ELEXIS will aim to serve as a catalyst for a closer cooperation between the two existing infrastructures. ELEXIS can succeed in this role because it has assembled a critical mass of eminent stakeholders from various disciplines who have both the technical and scholarly potential to 1) help lexicographers build better dictionaries using the most advanced NLP techniques; 2) provide NLP researchers with high-quality lexicographic data to test and improve their algorithms on; and 3) aid humanities scholars in accessing social, historical and cultural data contained in legacy dictionaries in order to develop new procedures and tools for analyzing, visualizing and interpreting large sets of lexical data.

Lexical data is essential for scholarly workflows in a wide range of disciplines. Through its carefully planned activities in Training and Education (WP5) as well as Dissemination and Community Building (WP7), ELEXIS will be in a position not only to promote its own research outputs and services, but also to identify and attract new users: lexicographers with little technical expertise, lesser-resourced language communities and traditional humanities scholars, for instance in literature, history and cultural studies, some of whom will engage with a digital research infrastructure for the first time through the ELEXIS platform. Dictionaries are scholarly tools par excellence. That is why Europe needs a strong and robust lexicographic infrastructure.

A dedicated task "Compatibility with CLARIN/DARIAH services" in WP6 will coordinate high-level contacts with the two infrastructures, while individual ELEXIS consortium members, who play leading roles in the two infrastructures, will work in parallel on maintaining productive relationships with the relevant working groups within CLARIN and DARIAH (WG Lexical Resources, WG Thesaurus Maintenance, WG Training and Education, WG Guidelines and Standards and WG Text and Data Analytics). This way, ELEXIS will align both its strategic goals and its concrete scholarly objectives with the ongoing work in the European infrastructure landscape, while providing new tools and services, as well as vast amounts of research data and new knowledge that will significantly expand current infrastructural offerings.

### *Relationship with CLARIN*

Data access services and standards used and available in **CLARIN** and **DARIAH** infrastructures will be explored. ELEXIS will therefore align efforts with complementary infrastructures and initiatives. CLARIN provides an interoperable research infrastructure for language resources and technology, and DARIAH is a research infrastructure for the arts and the humanities. ELEXIS will leverage and build upon the existing services with federated identity, persistent identifiers, content search or web service chaining, already provided by other infrastructures.

Many of the ELEXIS partners are either National Coordinators or members of CLARIN and DARIAH national consortia. However, trans-national or virtual access activities within ELEXIS will reach beyond the consortium itself. By focusing on the access needs of individual scholars as well as national language institutes, standardisation bodies and other stakeholders in lexicography, all of which have varying degrees of technological experience, ELEXIS will bring **new user bases** to both CLARIN and DARIAH, while making use of existing repositories, authentication procedures and other facilities.

Eight partners in the consortium are members of national CLARIN consortia:<sup>1</sup> JSI (Slovenia), INT (Netherlands), OEAW (Austria), RILMTA (Hungary), IBL (Bulgaria), CNR-ILC (Italy), UCPH (Denmark) and EKI (Estonia). Six of them are national coordinators (Slovenia, Austria, Hungary, Bulgaria, Italy and Denmark). Four of the partners are also hosting certified CLARIN B Centers with Data Seal of Approval. These will be used as repositories for providing access to the results of ELEXIS project of the »data type«, as shown in Picture :

❖ CLARIN.SI Language Technology Centre	B Center CLARIN.SI	Slovenia
❖ CLARIN Centre Vienna	B Center CLARIN-AT	Austria
❖ Instituut voor de Nederlandse Taal	B Center CLARIN-NL	Netherlands
❖ The CLARIN Centre at the University of Copenhagen	B Center CLARIN-DK	Denmark

<sup>1</sup> <https://www.clarin.eu/content/participating-consortia>

Task T6.4 in WP6 is dedicated to ensure compatibility between ELEXIS infrastructure and CLARIN/DARIAH services and tools. WP6 is lead by JSI, national coordinator of CLARIN.SI, and the task T6.4 is lead by UCPH, one of the founding members of European CLARIN with ample experience with technological, legal and other aspects of CLARIN infrastructure. As ELEXIS is a four-year project, the goal is to ensure life after project through CLARIN infrastructure. A subgroup of existing CLARIN partners will be formed in ELEXIS to work towards this goal.

#### *Relationship with DARIAH:*

Seven partners in the consortium are members of national DARIAH consortia:<sup>2</sup> JSI (Slovenia), NUIG (Ireland), OEAW (Austria), BCDH (Serbia), FCSH-UNL (Portugal), CNR-ILC (Italy) and UCPH (Denmark). Three of them are also national coordinators (Ireland, Austria, Serbia). OEAW is also co-leader of the DARIAH-ERIC Virtual Competency Centre “E-Infrastructure” while BCDH is co-leader of DARIAH-ERIC Working Groups “Lexical Resources” and “Training and Education”.

Both the Board of Directors of CLARIN-ERIC and of DARIAH-ERIC have expressed their support for this proposal.

#### *ELEXIS platform architecture*

Research and networking activities in ELEXIS will result in a platform which will consist of several sets of tools and services, as well as new data, in three distinct parts of the platform, or infrastructures.

**LEX1:** The first set of services and tools will be dedicated to automatic segmentation and structuring of content for dictionaries that are currently produced in digital environments but are typically encoded in their own custom data format. **ELEXIS conversion and alignment tools** will provide users of the infrastructure with the possibility to harmonise and convert their lexicographic resources to a uniform data format that allows their seamless integration in Linked Open Data. The existence of common data models and standards that are produced bottom-up from within the lexicographic community fostered by ELEXIS is a necessary condition for successful development of this segment of the platform. Standards will be developed and tested during the project on the data provided by the lexicographic partners and implemented in the newly-developed service.

To provide conceptual interoperability, services enabling linking of ELEXIS lexicographic resources will be developed and made available in the **ELEXIS linking tools** segment of the platform. This will provide the possibility to link lexical entries, senses and fundamental concepts in different lexical resources, using a semi-automatic approach. BabelNet, as an existing multilingual resource to provide cross-lingual linking, will be exploited for this purpose. Extensive linking of existing lexicographic resources by pivoting through BabelNet will enable the creation of what we call ELEXIS matrix dictionary – a universal repository of linked senses, meaning descriptions, etymological data, collocations, phraseology, translation equivalents, examples of usage and all other types of lexical information found in all types of existing lexicographic resources, monolingual, multilingual, modern, historical etc. Data from this new resource will be available through **ELEXIS matrix dictionary** RESTful Web service as part of the platform.

**LEX2:** Based on the the contribution of lexicographic data in addressing the “semantic bottleneck” a new infrastructure will be developed that will include **ELEXIS word sense disambiguation and entity linking tools** dedicated to semantic processing of corpus data, but including tasks such as sense clustering, domain labeling

---

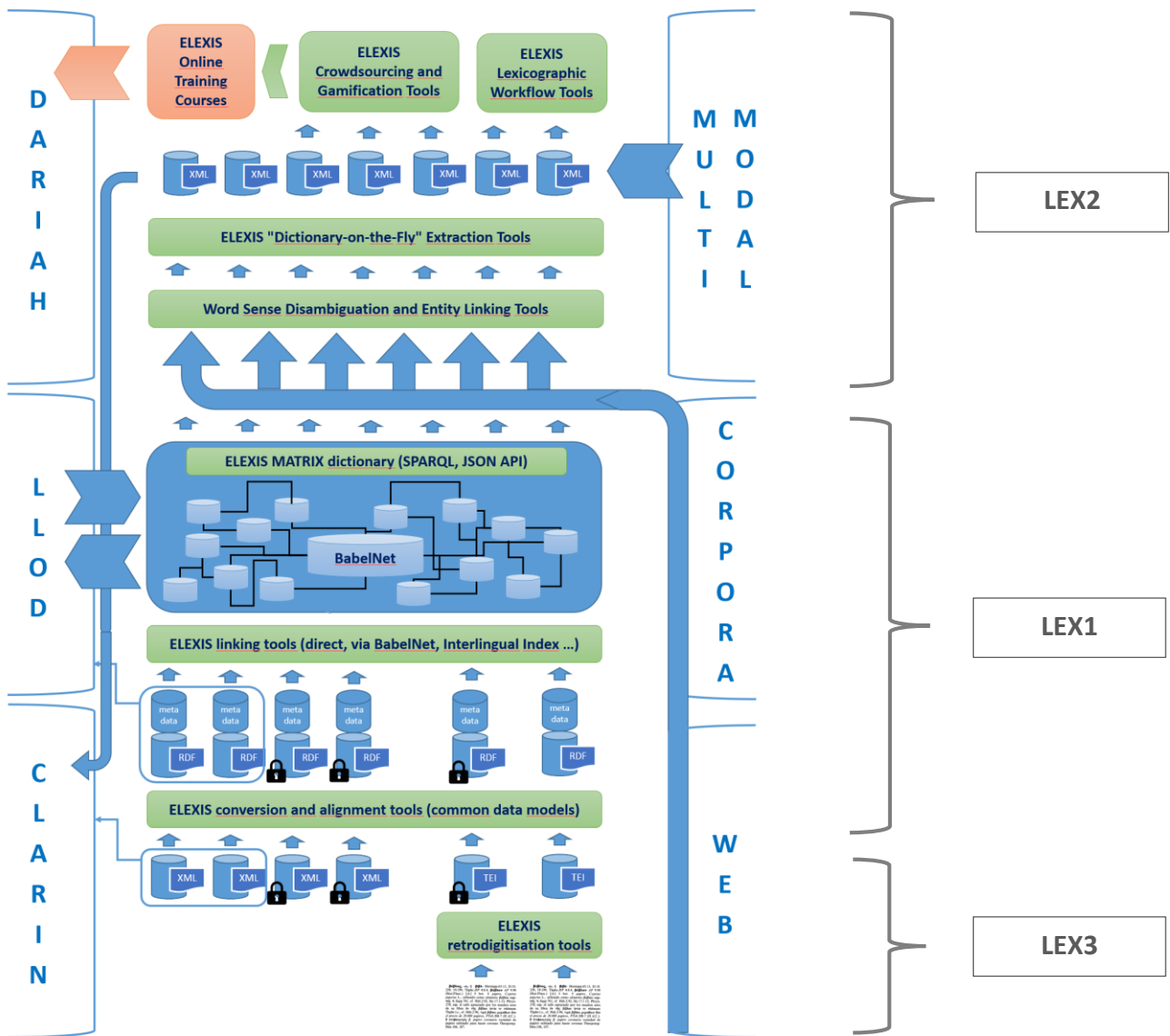
<sup>2</sup> <http://www.dariah.eu/about/our-partners/>

of text, diachronic distribution of senses (ranking senses by frequency of use over time) and similar. These tools will have an important impact on disambiguation and corpus analysis, and will open up the possibility to create lexicographic data from corpora in a fully automated process. This is included in the **ELEXIS »dictionary-on-the-fly«** segment of the platform. The service will be able to produce a proto-dictionary with sense distribution, extracted definitions, collocations, multi-word expressions, (good dictionary) examples, translation equivalents and data in other modalities.

To enable online lexicographic work on both existing and new (extracted) lexicographic data, two complementary sets of tools will be provided: **ELEXIS lexicographic workflow tools** and **ELEXIS crowdsourcing and gamification tools**. The first will include a user-friendly open source online dictionary writing system, with the aim to provide the central dictionary writing platform for new lexicography which also includes new possibilities of online collaboration. The other will provide tools for new techniques of dictionary creation, such as explicit or implicit crowdsourcing (gamification). Although similar tools are already in existence (Wiktionary, Urban Dictionary etc.), in ELEXIS the emphasis is on the inclusion of crowdsourcing and gamification techniques in the production of all types of lexicographic resources, also those that are traditionally considered as created exclusively by language description experts.

**LEX3:** The third set of services is dedicated to retrodigitised dictionaries in **ELEXIS retrodigitisation tools** part of the platform that will include (1) tools for automatic segmentation and structuring of content in retrodigitised dictionaries, and (2) an online generic, modular dictionary publication tool for retrodigitised dictionaries which also offers interfaces for the analysis and profiling of the underlying lexical data. The viewer will include different visualisation, geolocation and profiling tools that make it possible for end users to explore and navigate the dictionary content in novel ways that go beyond the dominant look-up paradigm.

Graphic representation of the ELEXIS platform architecture is presented in Picture 3:



Picture 3 ELEXIS platform

Tools and services: **green**, (lexicographic) data: **blue**, online training and education: **brown**, virtual access infrastructures: **grey**.

The platform will be built on the basis of technology that already exists in the consortium. We list some of the important existing tools that will be used in the development or directly included in the platform:

Software, Service	Short Description	Partner	Status	In
<b>NoSketchEngine</b>	Corpus query and corpus management software. <a href="https://www.sketchengine.co.uk/nosketch-engine/">https://www.sketchengine.co.uk/nosketch-engine/</a>	LC	open source	LEX2
<b>Sketch Engine</b>	A corpus management system designed for lexicographic purposes <a href="https://www.sketchengine.co.uk/">https://www.sketchengine.co.uk/</a>	LC	free trial	LEX2
<b>Lexonomy</b>	Online dictionary writing system. <a href="http://www.lexonomy.eu/_en/">http://www.lexonomy.eu/_en/</a>	LC	open source	LEX2
<b>Corpus Tools</b>	Tools available at <a href="http://corpus.tools">http://corpus.tools</a>	LC	open source	LEX2

	<p><b>Spiderling</b> is a web spider for linguistics. It can crawl text-rich parts of the web and collect a lot of data suitable for text corpora.</p> <p><b>JusText</b> is a HTML boilerplate removal tool. It can strip navigation links, headers, footers, etc. from HTML pages and leave just regular text containing full sentences</p> <p><b>Chared</b> is a tool for detecting the character encoding of a text in a known language. It contains models for a wide range of languages.</p> <p><b>Onion</b> (ONE Instance ONLY) is a de-duplicator for large collections of texts. It can measure the similarity of paragraphs or whole documents and drop duplicate ones based on the threshold you set</p> <p><b>Unitok</b> is a universal text tokeniser with specific settings for many languages. It can turn plain text into a sequence of newline-separated tokens (“vertical” format), while preserving XML-like tags containing metadata.</p> <p><b>wiki2corpus</b> is a script which downloads Wikipedia articles (for a given language) and outputs them in the form of prevertical which can be further processed by other corpus tools.</p>			
<b>BlackLab</b>	A corpus retrieval engine based on Apache Lucene <a href="http://inl.github.io/BlackLab/">http://inl.github.io/BlackLab/</a>	<b>INT</b>	open source	LEX2
<b>Event Registry and Enrycher</b>	Text and language processing, cross-lingual text matching. <a href="http://enrycher.ijs.si">http://enrycher.ijs.si</a> , <a href="http://eventregistry.org">http://eventregistry.org</a>	<b>JSI</b>	open source	LEX1, LEX2
<b>Ontogen, Textgarden, GLib</b>	Semantic representation services and tools. <a href="http://ontogen.ijs.si/">http://ontogen.ijs.si/</a> , <a href="http://ailab.ijs.si/dunja/textgarden/">http://ailab.ijs.si/dunja/textgarden/</a>	<b>JSI</b>	open source	LEX1, LEX2
<b>QMiner</b>	Big data analytics pipeline used for user modelling, learning analytics and personalisation	<b>JSI</b>	open source	LEX1, LEX2
<b>xling</b>	<a href="http://xling.ijs.si/">http://xling.ijs.si/</a> – a cross-lingual text matching and categorisation service covering 100 languages (based on the top 100 Wikipedia languages)	<b>JSI</b>	open access service	LEX1, LEX2
<b>wikifier</b>	<a href="http://wikifier.ijs.si">http://wikifier.ijs.si</a> - a web service for semantic annotation of text documents covering 100 languages (based on the top 100 Wikipedia languages)	<b>JSI</b>	open access service	LEX1, LEX2
<b>babelfy</b>	a tool with a unified, multilingual, graph-based approach to Entity Linking and Word Sense Disambiguation. <a href="http://babelfy.org/">http://babelfy.org/</a>	<b>UNIROMA1</b>	open access service	LEX1, LEX2

<b>MultiWiBi</b>	MultiWiBi (Multilingual Wikipedia Bitaxonomy) is an approach to the automatic creation of a bitaxonomy for Wikipedia in arbitrary language. <a href="http://wibitaxonomy.org/">http://wibitaxonomy.org/</a>	<b>UNIROMA1</b>	open access service	LEX1, LEX2
<b>NAISC</b>	a system for creating mappings between datasets including ontologies, dictionaries and other tabular and structured data <a href="https://github.com/jmccrae/naisc">https://github.com/jmccrae/naisc</a>	<b>NUIG</b>	open source	LEX1
<b>Yuzu</b>	a system to quickly integrate different datasets in various formats and make them available through a browseable HTML interface and SPARQL search <a href="https://github.com/jmccrae/yuzu">https://github.com/jmccrae/yuzu</a> Demo at <a href="http://linghub.org">http://linghub.org</a>	<b>NUIG</b>	open source	LEX1

### *ELEXIS lexicographic data*

ELEXIS will produce or enable work on several types of lexicographic resources. Initially, existing lexicographic resources that are available in the consortium will be harmonised and converted into new data sets with common standardised formats and data models. As one of the important objectives of ELEXIS is enabling access to lexicographic data and promoting an open access culture in lexicography, efforts will be made during the project to solve IPR issues currently preventing the inclusion of the converted and harmonised data into existing repositories in CLARIN infrastructure or/and Linguistic Linked Open Data cloud (LLOD), as represented in Picture 3. ELEXIS will work on discovering ways to overcome and provide both legal and technical solutions to enable immediate inclusion of existing resources under open licences. Another mechanism to overcome IPR issues will be the creation of a »matrix dictionary« with information based on existing resources and obtained via BabelNet through cross-lingual linking of dictionary data and meta-data. This resource will be available under the same licence as BabelNet (Creative Commons Attribution-Non Commercial-Share Alike 3.0 License).

On the NLP side of the virtuous cycle of e-lexicography, semantically disambiguated corpus data processed with corpus analysis tools enhanced by newly-available lexicographic data will enable the creation of a new type of lexicographic resource: dictionary-on-the-fly. In particular, lesser-resourced languages will benefit from these new types of resources since new methods and tools will significantly shift the lexicographer's starting point and reduce the time-consuming parts of lexicographic work. In principle, having enough web or corpus data in a particular language will be a sufficient condition for a dictionary of that language to be created, bridging the gap between lesser-resourced languages and those with advanced e-lexicographic experience. The extracted textual data will be enriched with multi-modal data and made available via online publishing tools and in existing CLARIN repositories, as shown in in Picture 3, under open licences.

Existing lexicographic data available in the ELEXIS consortium and used in the development or included in LEX1, LEX2 and LEX3 will be the following (for availability we use CLARIN categories – PUB: distributed publicly, ACA: distributed for research purposes, RES: require permission from the rights holder, UNK: unknown):

<b>partner</b>	<b>dictionary or lexicographic resource</b>	<b>language</b>	<b>availability</b>
<b>JSI</b>	Slovene Lexical Database	Slovene	PUB
<b>INT</b>	Dictionary of Contemporary Dutch (ANW)	Dutch	RES



<b>INT</b>	Dictionary of the Dutch Language (WNT)	Dutch	RES
<b>INT</b>	Dictionary of Old Dutch (ONW)	Old Dutch	RES
<b>INT</b>	Dictionary of Early Middle Dutch (VMNW)	Early Middle Dutch	RES
<b>INT</b>	Dictionary of Middle Dutch (MNW)	Middle Dutch	RES
<b>OEAW</b>	Dictionary of Bavarian Dialects of Austria	Austrian	UNK
<b>OEAW</b>	Dagaare-Cantonese-English Dictionary	Dagaare, Cantonese, English	PUB
<b>OEAW</b>	Hausa-English Dictionary	Hausa, English	UNK
<b>OEAW</b>	Database of Bavarian Dialects of Austria	Austrian Variants	UNK
<b>OEAW</b>	Russian Dialect Dictionary	Russian	UNK
<b>OEAW</b>	Tunico	Tunisian	PUB
<b>OEAW</b>	Viennese Historical Dictionaries Online (ViDi)	Austrian Variants	UNK
<b>BCHD</b>	Karadžić, Serbian Dictionary (1818, 1852)	Serbian	ACA
<b>BCHD</b>	Miklošić, Lexicon Palaeoslovenico-Graeco-Latinum (1862—1865)	Old Church Slavic	ACA
<b>BCHD</b>	Daničić, Dictionary of Serbian Literary Antiquity (1863-4)	Serbian (medieval)	ACA
<b>BCHD</b>	Bojanić & Trivunac, Dictionary of Dubrovnik Dialect	Serbian (dialect)	ACA
<b>BCHD</b>	Elezović, Dictionary of Kosovo-Metohija Dialect	Serbian (dialect)	ACA
<b>BCHD</b>	Zlatanović, Dictionary of Southern Serbian Dialects	Serbian (dialect)	ACA
<b>BCHD</b>	Žugić, Dictionary of Jablanica Region	Serbian (dialect)	ACA
<b>RILMTA</b>	Hungarian Concise Dictionary	Hungarian	ACA
<b>IBL</b>	Dictionary of synonyms	Bulgarian	UNK
<b>IBL</b>	Dictionary of antonyms	Bulgarian	UNK
<b>IBL</b>	Dictionary of new words	Bulgarian	ACA
<b>IBL</b>	Dictionary of Bulgarian	Bulgarian	ACA
<b>KD</b>	K English Multilingual Dictionary	English multilingual	RES
<b>KD</b>	Global French Multilingual + L2-French	French multilingual	RES
<b>KD</b>	Random House Webster's College Dictionary	English	RES
<b>DSL</b>	The Danish Dictionary	Danish	RES
<b>DSL</b>	Dictionary of the Danish Language	Danish	RES
<b>DSL</b>	Moths Dictionary	Danish	RES
<b>DSL</b>	Old Danish Dictionary	Old Danish	RES
<b>DSL</b>	Danish Thesaurus	Danish	RES
<b>UHP</b>	Dictionary of Danish Insular Dialects	Danish	ACA
<b>UHP</b>	Dictionary of Old Norse Prose	Old Norse	ACA
<b>TCDH</b>	The German Dictionary by Jacob and Wilhelm Grimm (first edition)	German	RES
<b>TCDH</b>	Rhenish Dictionary	German (dialect)	RES
<b>TCDH</b>	Palatinate Dictionary	German (dialect)	RES
<b>TCDH</b>	Dictionary of the German-Lorraine Dialects	German (dialect)	RES
<b>TCDH</b>	Dictionary of the Alsatian Dialects	German (dialect)	RES
<b>TCDH</b>	Grammatical-Critical Dictionary of the High-German Idiom (Adelung, second edition)	German	RES



<b>TCDH</b>	Middle High German Dictionary (Benecke, Müller, Zarncke)	Middle High German	RES
<b>TCDH</b>	Middle High German Dictionary (Lexer)	Middle High German	RES
<b>EKI</b>	The Dictionary of Standard Estonian ÕS 2013	Estonian	ACA
<b>EKI</b>	The Explanatory Dictionary of the Estonian Language	Estonian	ACA
<b>EKI</b>	The Dictionary of Foreign Words	Estonian	ACA
<b>EKI</b>	The Estonian Etymological Dictionary	Estonian	ACA
<b>EKI</b>	The Basic Estonian Dictionary	Estonian	ACA
<b>EKI</b>	The Estonian-Russian Dictionary	Estonian-Russian	ACA
<b>EKI</b>	The Russian-Estonian Dictionary	Russian-Estonian	ACA
<b>RAE</b>	Diccionario de la lengua española, 22nd Ed. (2001)	Spanish	RES

## 1.4 AMBITION

Through a continuous circular process in the virtuous cycle of e-lexicography, the ambition of ELEXIS is to enable all research communities working in the field of language description and standardisation to benefit from rapid development in ICT technologies, by introducing levels of automation in the lexicographic process that are currently not conceivable. Combining expertise from the fields of lexicography and natural language processing, semantic web and artificial intelligence will enable significant progress beyond the state-of-the-art in lexicography, and it will also significantly improve existing lexical resources for NLP.

### 1.4.1 Progressing the state-of-the-art in lexicography

#### *Linking existing lexicographic data*

ELEXIS will build on the infrastructures defined in other projects, especially CLARIN and DARIAH, which allow language resources (including lexical resources) to be shared. This will allow the partners within the network to easily share their resources, yet this does not necessarily lead to any interoperability. In order to motivate interoperability, we will enable partners and other stakeholders to encode their data with common concepts from models such as the BabelNet and other Semantic Web models, such as DBpedia. Moreover, to ensure that there is integration at even the most basic level we will define a minimal **common data model** capturing the basic concepts of a lexicographic resource such as entries (single-word, multi-word), senses, syntactic frames, etymologies etc. and linguistic relationships such as synonymy/antonymy, translation, domain/region/register classification, relatedness, etc. that will be compatible with existing models used in the community, including TEI, LMF and OntoLex-Lemon. In addition, each partner will be able to form modules defining specific extensions of the common core to specify particular elements of their lexicographic resources. This will be implemented by providing a set of round-trip conversions that will be able to take data in different formats and make it available in the core formats, without loss of information. The converted data will be available in RDF, facilitating linking and publishing on the Web as linked data. We aim to do this for all resources with an open license.

A key goal of the ELEXIS project is to enable individual collaborators to link their existing dictionaries and thus to create a **huge multilingual registry that connects dictionaries across common concepts**. This is vital for encouraging reuse and collaboration between partners and in ELEXIS in general.

#### *Integrating and enriching lexicographic data*

Moreover, the above procedure allows the integration of datasets into a single database. In particular, we note that the use of SPARQL and other graph-based and NoSQL databases has been shown to enable interesting and deep analyses across multiple datasets. In particular, this analysis will be used in the project to enable **lexicographic resources to be connected to other forms of linguistic data**, such as text corpora and typological databases, and also multimodal data, such as speech corpora and databases, image/video repositories and similar, creating novel resources offering more and better information to end users.

The linking of data will lead to a virtuous cycle where integrating existing data will provide further information to a system for multilingual and **multimodal information extraction**, which can then further feedback into the linking systems. We will develop a system of integration based on the following principles: lemma and gloss-based similarity matching entries based on their similarity (McCrae et al. 2016); graph-based similarities matching elements based on their synonyms and hypernyms; random-walk based algorithms for resource interlinking (Pilehvar and Navigli 2014); cross-lingual similarities based on context-aware machine translations. All of the approaches will be combined using state-of-the-art statistical learning to provide high quality alignments that can be used to help in the automatic and semi-automatic linking of dictionaries.

### *Supporting new lexicography*

In addition to linking, integrating and enriching existing data, an important technological goal of the project is to support novel lexicography by providing lexicographers with tools that help them to create new resources. In particular, we will provide an **integrated platform** that allows lexicographers to see results coming from information extraction and corpus information, as well as crowdsourcing. This will enable lexicographers to make more detailed and consistent analyses of words in context. However, it is important to note that from the point of view of human-oriented lexicography, faced with abundance of data, the emphasis is on knowing what not to say and on how to say it efficiently. Therefore methods and tools for visualisation and presentation of lexicographic data are also extremely important in this context, and will receive due attention.

### *Bridging the gap between lesser-resourced and advanced communities*

The significant variation in the level of expertise and resources available to lexicographers across Europe is a major obstacle to more ambitious, transnational, data-driven approaches to dictionaries as both tools and objects of research. By focusing on defining clear and reproducible workflows, developing sharable conceptual models for lexical data, establishing robust data services and providing a range of training, educational and outreach activities, ELEXIS will **foster cooperation and knowledge exchange** between different research communities in order to bridge the gap between lesser-resourced languages and those with advanced e-lexicographic assets.

The role of a research infrastructure is not only to enable access to and exchange of raw primary data, but also to leverage technologies in order to generate and propagate new knowledge. Yet in order to establish and deliver a portfolio of tools and services aligned with the inter- and multidisciplinary nature of lexical data that can help researchers ask new questions and pursue new avenues of research, a truly European digital research infrastructure in the field of e-lexicography has to make sure that it is **equally open to all languages**, all researchers, all institutions and all stakeholders, regardless of their level of technical expertise. ELEXIS is based on the conviction that lowering the barrier for retrieving and analysing multilingual lexicographic data across Europe cannot be accomplished in the long term without lowering the barrier for providing lexicographic data to research infrastructures.

**1.4.2 Progressing the state-of-the-art in NLP, SW and AI**

Linguistic Linked Open Data (LLOD) is an initiative to break the data silos of linguistic data and thus encourage NLP applications that can use data from multiple languages, modalities (e.g., lexicon, corpora, etc.) and develop novel algorithms. However, as of today, the LLOD is not populated by many lexicographic resources, due to the lack of a dedicated infrastructure for resource interlinking and of effective ontology alignment algorithms, which depend on multilingual semantic similarity, entity linking and word sense disambiguation. ELEXIS will revolutionise the LLOD scenario, by providing key innovation across many research domains, and radically changing the development and application of lexico-semantic resources. As a result, the so-called knowledge acquisition bottleneck will be relieved and important applications in NLP and AI, including machine translation, machine reading and intelligent digital assistance, will be considerably enhanced, thanks to their ability to scale to wide coverage in multiple languages. The novel linking and cross-lingual mapping algorithms developed in this project will find usage not only for lexicographic data but also lead to generic techniques for data integration, that will enable datasets to be more easily integrated and thus eliminate complex ‘data wrangling’ techniques that currently plague big data applications. Similarly, the techniques for data validation and quality assurance that will be developed in this project will not only be useful for lexicographic data but also be applicable to Semantic Web, linked data and big data in general, ensuring that the data economy is more usable and thus enabling automatic agents to work in the data mining ecosystem with little human input.

## 2 IMPACT

### 2.1 EXPECTED IMPACTS

Specific impact	ELEXIS contribution
<b>Providing efficient access to quality lexicographic data</b>	<p>Through ELEXIS lexicographic communities working on scholarly dictionaries, language description and standardisation will gain access to:</p> <ul style="list-style-type: none"> <li>○ modern and retrodigitised historical lexicographic data,</li> <li>○ large amounts of linked and integrated semantic data and</li> <li>○ extracted structured data from text corpora and multimodal resources,</li> <li>○ online training materials in DARIAH services,</li> <li>○ networking events (meetings, conferences, workshops).</li> </ul> <p>Computational linguistics and language resources communities will gain access to currently inaccessible data from quality lexicographic resources and interlinked semantic data, as well as extracted data from corpora and multimodal resources;</p> <p>Digital humanities communities will gain user-friendly and efficient access to modern and historical lexicographic resources as cultural and historical artefacts, supporting research in a wide area of humanities disciplines such as history, religion, gender studies, literature and education.</p>

---

<b>Establishing inter- infrastructure synergies and optimisation</b>	<p>Currently isolated European language infrastructures working on lexical description of individual languages in national language institutes and standardisation bodies will be joined in one pan-European infrastructure.</p> <p>Close links and synergies will be established between CLARIN and DARIAH, with ELEXIS working on top of existing services as a new user community.</p>
<b>Enabling the use of new technology and data in industry</b>	<p>Industrial partners in ELEXIS will be able to take the role of intermediaries between research and industry in language technology and language learning, as well as lexicography and lexical content publishing in general. Interest of industry is visible from participating partners and from the letters of interest by important stakeholders in the field;</p> <p>Information from quality lexicographic resources and interlinked semantic data will be opened up and made available for use in commercial scenarios, based on ELEXIS work on IPR issues currently hindering the accessibility of the data.</p> <p>Lexicographic data will be evaluated by industry-supported data seal of compliance.</p>
<b>Facilitating inclusion of innovative lexicography in research and education</b>	<p>Online training courses on innovative e-lexicography with suggested ECTS produced by education partners (from universities) will be incorporated into existing curricula.</p> <p>Language teaching and language learning communities will be able to develop and use new improved training materials, based on the (open) access to lexica interlinked on a large scale.</p> <p>Previously unaccessible lexicographic data will be made available for research through virtual access platforms and through visiting grants in trans-national access.</p>
<b>Encouraging cross- disciplinary fertilisations in academia and industry</b>	<p>Both computational linguistics and lexicography will be able to achieve a higher level of language description and text processing in a virtuous cycle of cross-disciplinary exchange of knowledge and data;</p> <p>Research or study of lexica in linguistic studies and related disciplines will be enabled by massive interlinking of previously isolated lexicographic resources, which can lead to new discoveries, particularly in the semantic domain.</p> <p>In humanities disciplines, such as history, religion, gender studies, literature and education, new resources and services can be used for cross-lingual studies, based on interlinked and integrated semantic data;</p> <p>artificial intelligence systems will be able to make use of lexicographic data in repositories, interlinked semantic data and extracted data from multilingual and multimodal resources.</p>
<b>Enabling massive integration of</b>	<p>Stand-alone modern and historical lexicographic resources available as isolated incompatible data will be linked, integrated and enriched on</p>

---

<b>knowledge-based resources</b>	<p>different levels. A scalable, multilingual and multifunctional, language resource will be created by:</p> <ul style="list-style-type: none"> <li>- <b>linking resources:</b> this means providing links between different elements of dictionary entries (lemmas/headwords, senses, definitions, multi-word expressions, etymologies, etc.) enabling any dictionary (element) to be linked with all other dictionaries (or dictionary elements). Result: a growing network of existing dictionaries linked across common concepts via a huge (multilingual) index.</li> <li>- <b>integrating resources:</b> this means taking information from individual resources and putting them together in a new resource / aligning them to create a combined resource. Result: any combination of existing (linked) resources resulting in a new resource available for immediate use or as a starting point for creating a novel individual lexicographic resource.</li> <li>- <b>enriching resources</b> with multimodal data (image, sound, video), and unstructured text (corpora, news feeds, social media etc.) Result: a portal with cross-lingual, cross-media information on word usage.</li> </ul> <p>Ultimate goal is the creation of a universal (integrated and enriched) registry/network of semantic relations used as a semantic intermediary language for global knowledge exchange, focused on difficult polysemous vocabulary (single-word and multi-word), modern and historical; the realisation of a universal lexicographic metastructure; a matrix dictionary spanning across languages and time.</p>
----------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<b>Enhancing past investments from EU Structural and Investment Funds</b>	<p>Language resources used for natural language processing which were funded also by European Structural and Investment Funds, particularly in the case of lesser-resourced languages (e.g. Slovenia - <a href="http://eng.slovenscina.eu/">http://eng.slovenscina.eu/</a>), will be enhanced by linking them with results of nationally funded (long-term) lexicographic projects, thus providing synergy between the two types of funding.</p>
---------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

## 2.2 MEASURES TO MAXIMISE IMPACT

### 2.2.1 Dissemination and exploitation of results

#### *Dissemination activities*

From the beginning, each project partner will be committed to creating a high level of publicity for the project by communicating the launch and objectives of the project through their own communication channels – e.g. press releases, newsletters and news-related media, website, etc. – to generate broad public awareness of ELEXIS activities. A project website will be created, hosted and maintained by the partner responsible for dissemination, on behalf of the consortium. It will be designed, structured and implemented in the first two months of the project. Its primary aim is to provide general purpose information about the project and its

objectives, summarise the major project results to third parties, provide up-to-date event and news information, and provide a space for download of public documents. An internal communication/document sharing platform with restricted access will also be made available to the consortium and associated partners, and to the European Commission. This Web platform will be permanently updated and improved in line with the project results.

The dissemination plan recognises the fact that various communities and types of users (professional, semi-professional and general public) will benefit on account of the scalable project outcomes. The dissemination plan will ensure that the results are exploited at partner's level, at consortium level and by external communities. The activities will not be limited geographically and will address users world-wide because the ELEXIS outcomes can realistically aspire to set universal standards for lexicography.

In order to reach different target groups, multi-channel dissemination actions will be carried out, with information adjusted to audience level of needs and involvement. We identify several major target groups:

#### Professional large-scale lexicography

This group includes commercial and non-commercial entities undertaking large-scale lexicographic projects carried out by professional specialised teams from these areas:

- national language institutes (including consortium members)
- academia, universities, research institutions outside the ELEXIS consortium
- language standardisation bodies and their umbrella organisation EFNIL (European Federation of National Institutions for Language).
- industry (publishing houses, also software developers and language industry – in connection with large lexicographic projects)

#### Professional small-scale lexicography

This group includes entities undertaking small-scale lexicographic projects carried out on a highly professional level either for research purposes or to address the needs of a small well-defined community. The following groups may fall into this category:

- individual researchers (from the field of lexicography, also language studies, translation studies or the sister field of digital humanities, as well as natural language processing in connection with lexicography)
- trainers and students (interested in the educational aspects of the ELEXIS projects such as learning material, training events)
- professionals and practitioners (language professionals, translators, proofreaders and others who use or produce linguistic resources in their daily professional life)
- freelance terminologists

#### Spontaneous and small-scale lexicography

This group includes an enormous number of small projects often carried out without expertise in lexicography to address very specific needs of highly-specialised or very small professional or general public communities. A typical example would be a highly specialised domain-specific glossary. The following groups may fall into these categories:

- professional organisations, associations and authorities, non-profit organisations
- general public

Each group will be targeted with tailor-made messages that address the needs of the community to maximise the impact of any such activity. This will build and foster a loyal community and encourage organic dissemination – a vital prerequisite for the effective exploitation of the project results by the widest possible community.

The following table summarises the dissemination policies that ELEXIS participants will undertake in correspondence to the target audience:

Target group	Dissemination approach	Activities
<b>Professional large-scale lexicography</b>		
National language institutes	This target group will exploit ELEXIS services (LEX1, LEX2, LEX3) in their entirety receiving special treatment in dissemination activities. The group is relatively small but influential on national or regional level. Individual institutions will be approached and offered observer status in ELEXIS.	<ul style="list-style-type: none"> <li>• Web site</li> <li>• Scientific papers</li> <li>• Targeted workshops, scientific events</li> <li>• Online training courses, training events</li> <li>• Brochures, publicity material</li> <li>• International Lexicographic Alliance (long-term community building)</li> <li>• Measuring user satisfaction and usability</li> </ul>
Academia, universities, research institutions	This target group is larger and more heterogeneous, and will be approached through general dissemination activities. Many of institutions are existing users of the Sketch Engine tool and will be approached also through this channel.	<ul style="list-style-type: none"> <li>• Web site</li> <li>• Scientific papers</li> <li>• Targeted workshops, scientific events</li> <li>• Online training courses, training events</li> <li>• Brochures, publicity material</li> <li>• International Lexicographic Alliance (long-term community building)</li> <li>• Measuring user satisfaction and usability</li> </ul>
Language standardisation bodies	This target group is organised in the European Federation of National Institutions for Language (EFNIL). In ELEXIS activities we will use this channel to disseminate the results in this target group. Some ELEXIS partners are members of EFNIL (OEAW, RILMTA, IBL, EKI), and will help with dissemination. We will offer this group to join ELEXIS as observers.	<ul style="list-style-type: none"> <li>• Web site</li> <li>• Brochures, publicity material</li> <li>• Online training courses, training events</li> <li>• International Lexicographic Alliance (long-term community building)</li> </ul>



Industry	Dissemination will focus on large dictionary publishers, content providers, and other businesses interested in lexicography. The two industrial partners in ELEXIS (LC, KD) have a large network of business partners. Also this group will be invited to become observers in ELEXIS.	<ul style="list-style-type: none"> <li>• Web site</li> <li>• Targeted workshops, training events</li> <li>• Online training courses</li> <li>• Brochures, publicity material</li> <li>• International Lexicographic Alliance (long-term community building)</li> <li>• Measuring user satisfaction and usability</li> </ul>
<b>Professional small-scale lexicography</b>		
Individual researchers Trainers and students Professionals and practitioners Freelance terminologists	Dissemination to this group will focus on disseminating the solutions in LEX2 and LEX3 infrastructures. This will be done through European and international conferences/ workshops, scientific newsletters, magazines, websites, etc. Special effort will be made in relation to individuals working on lesser-resourced languages.	<ul style="list-style-type: none"> <li>• Web site</li> <li>• Social media, mailing lists</li> <li>• Scientific papers</li> <li>• Targeted workshops, scientific events</li> <li>• Online training courses, training events</li> </ul>
<b>Spontaneous and small-scale lexicography</b>		
Professional organisations, associations and authorities, non-profit organisations and general public	Dissemination to this group will focus on disseminating the solution in LEX2 infrastructure that are suitable for non-professionals: online dictionary writing systems, gaming and crowdsourcing software etc.	<ul style="list-style-type: none"> <li>• Web site</li> <li>• Social media, mailing lists</li> <li>• Online training courses</li> </ul>

### Exploitation of results

ELEXIS will measure impact created through the innovations implemented in the LEX platforms and accompanying education and training activities, in particular:

- LEX1: tools for automatic segmentation and structuring of lexicographic content, conversion and alignment tools, linking tools with the matrix dictionary (through BabelNet);
- LEX2: tools for semantic processing of corpus data, and corpus analysis tools, with the possibility to create lexicographic data from corpora in a fully automated process («dictionary-on-the-fly»), integration and enrichment tools;
- LEX3: tools for automatic segmentation and structuring of content in retro-digitised dictionaries, an online generic, modular dictionary publication tool for retrodigitised dictionaries;
- Online tutorials and instruction manuals for ELEXIS services, workshops and summer schools, TNA visiting grants.



With the aim of defining the expectations of the project in a measurable way, we provide the table below with impact criteria, metrics and benchmarks to measure the success of target outcomes for different categories of exploitation and dissemination.

<b>OBJECTIVE 1</b>		
<b>Foster cooperation and knowledge exchange between different research communities in lexicography</b>		
<b>Target outcome</b>	<b>Indicator</b>	<b>Target</b>
Research visits (Trans-National Access)	Yearly number of visits accomplished (from M6-M48)	= 10
Use of online training modules (#dariahTeach, Videolectures.net etc.)	Number of users in mid-term (M24)	<= 500 (not satisfactory) >= 500 (satisfactory)
Use of online training modules (#dariahTeach, Videolectures.net etc.)	Number of users at the end of the project (M48)	<= 2000 (not satisfactory) >= 2000 (satisfactory)
Attendance at ELEXIS training events	Number of registrations per event	<= 40 (not satisfactory) >= 40 (satisfactory)
Attendance at ELEXIS community building events	Number of registrations per event	<= 100 (not satisfactory) >= 100 (satisfactory)
Forming International Alliance for Lexicography (institutes, universities, industry)	Number of institutions in the alliance (end of project)	<= 70 (not satisfactory) >= 70 (satisfactory)
Social media engagement	Number of followers (end of project)	<= 5000 (not satisfactory) >= 5000 (satisfactory)
Promotional activities	Web site analytics (after M24) - daily visits	<= 50 (not satisfactory) >= 50 (satisfactory)

<b>OBJECTIVE 2</b>		
<b>Establish common standards and solutions for the development of lexicographic resources</b>		
<b>Target outcome</b>	<b>Indicator</b>	<b>Target</b>
Publication of standardisation documentation for lexicographic resources	Developed standard included in ISO or OASIS (end of project)	>= 1 new standard
Use of Data Seal of Compliance	Number of lexicographic resources with the Seal (end of project)	< 100 (not satisfactory) > 100 (satisfactory)
Use of conversion tools to harmonise the different data formats of lexicographic resources (Virtual Access)	Number of resources using API pipeline in LEX1 and LEX3 platforms Number of languages (X)	< 500 (not satisfactory) > 500 (satisfactory)  X < 10 (not satisfactory)

		10 < X < 20 (satisfactory) X > 20 ( very satisfactory)
Scientific output on standardisation in lexicography	Number of scientific journal articles Number of scientific conference papers	< 3 (not satisfactory) > 3 (satisfactory) < 15 (not satisfactory) > 15 (satisfactory)

<b>OBJECTIVE 3</b> <b>Develop strategies, tools and standards for the extracting, structuring and linking of lexicographic resources</b>		
<b>Target outcome</b>	<b>Indicator</b>	<b>Target</b>
Publication of documentation on best practices for lexicography	% of lexicographic resources using suggested solutions (end of project)	< 1/3 (not satisfactory) > 1/3 (satisfactory)
Use of linking tools for lexicographic resources (Virtual Access)	Number of resources using API pipeline in LEX1 and LEX3 platforms (end of project) Number of languages (X)	< 500 (not satisfactory) > 500 (satisfactory)  X < 10 (not satisfactory) 10 < X < 20 (satisfactory) X > 20 ( very satisfactory)
Scientific output on linking in lexicography	Number of scientific journal articles Number of scientific conference papers	< 5 (not satisfactory) > 5 (satisfactory) < 25 (not satisfactory) > 25 (satisfactory)

<b>OBJECTIVE 4:</b> <b>Enable access to standards, methods, lexicographic data and tools for scientific communities, industries and other stakeholders</b>		
<b>Target outcome</b>	<b>Indicator</b>	<b>Target</b>
Use of tools for semantic processing of corpus data, and corpus analysis tools (Virtual Access)	Number of users of LEX2 platform (M24)	users > 10.000 requests per day > 100.000
Use of tools for semantic processing of corpus data, and corpus analysis tools (Virtual Access)	Number of users of LEX2 platform (M48)	users > 50.000 requests per day > 500.000

Creation of “Dictionary-on-the-fly” lexicographic data for different languages	Number of languages	< 30 (not satisfactory) > 30 (satisfactory)
Scientific output on automatic creation of lexicographic content (dictionary-on-the-fly)	Number of scientific journal articles	< 5 (not satisfactory) > 5 (satisfactory)
	Number of scientific conference papers	< 25 (not satisfactory) > 25 (satisfactory)

<b>OBJECTIVE 5</b>		
<b>promote an open access culture in lexicography</b>		
<b>Target outcome</b>	<b>Indicator</b>	<b>Target</b>
Publication of guidelines on solving IPR issues in lexicography	Number of lexicographic resources using solutions from guidelines (end of project)	< 500 (not satisfactory) > 500 (satisfactory)
Lexicographic resources (individual data sets) available under open access licenses	Number of lexicographic resources (end of project)	< 100 (not satisfactory) > 100 (satisfactory)
Meta-data from lexicographic resources available under open access licenses (Linguistic Linked Open Data)	Number of lexicographic resources (end of project)	< 500 (not satisfactory) > 500 (satisfactory)

#### *Data management plan*

During the project, ELEXIS participants will create a detailed data management plan (DMP), with pre-defined templates, based on the input of lexicographic data providers in the consortium. The plan will be conformant with FAIR Data principles, as one of the main objectives of ELEXIS is to make lexicographic data (F)indable, (A)ccessible, (I)nteroperable and (R)e-usable. For persistent identification and authorisation we will use existing services and repositories offered by CLARIN. Accessibility protocols for ELEXIS data will be defined and included in the DMP. Furthermore, (meta-)data interoperability is at the core of ELEXIS efforts and will also be part of the DMP templates.

All the papers and scientific publications will be free to access and will be available under the gold open access. The green open access will be implemented for the public deliverables and reports as well as the test data that can be further used for research and development. The privacy of individual data will be handled according to national and international directives and law - for the reference check the IPR section.

#### *Management of intellectual property rights (IPR)*

In order to properly manage the intellectual property and results generated in the course of the project, as well as pre-existing intellectual property, a consortium agreement will be signed by all partners. Among the issues this document will address are the following:

- Confidentiality among institutions involved in the project, as well as confidentiality policy towards external institutions;
- Ownership of future results, taking into account pre-existing intellectual property and each partner's contribution to the production of new or emergent intellectual property;
- Transfer of knowledge and results to those industrial partners interested in exploiting project results as well as to third parties.

Therefore, for the success of ELEXIS it is essential that all project partners agree on explicit rules concerning IP ownership, access rights to any Background and Foreground IP for the execution of the project and the protection of intellectual property rights (IPRs) and confidential information before the project starts. Such issues will be addressed in detail within the Consortium Agreement between all project partners.

The main purpose of the Consortium Agreement is to establish a legal framework for the project in order to minimize any internal issues within the ELEXIS consortium related to the work and its intellectual property. Specific provisions will be laid out in the Consortium Agreement so as to regulate questions like ownership (joint ownership, transfer of ownership), protection of knowledge, access rights (exclusion of specific pre-existing know-how, access rights required for carrying out the project or for use purposes, exclusivity, sublicensing, use of third party resources, etc

Open source: ELEXIS will be an active contributor to open-source software development, additionally encouraging technology uptake as a consequence. Various partners have gathered experiences for many years in open source development, which have become the technical foundation for both advanced research prototypes and robust commercial applications. The pre-existing tools are mostly open-source and they will be further developed in the project. The core technology of the project will be made openly available under an appropriate license that will be identified in the Consortium Agreement.

### 2.2.2 Communication activities

The ELEXIS communication strategy will be set up with a broad audience in mind, and will build an offline and online presence of the project with the aim to:

- disseminate the project goals and outcomes effectively,
- set up efficient tools for the communication with different stakeholders,
- build a loyal community of stakeholders to serve as intermediaries for reaching other communities,
- exploit synergies in liaisons and collaborations with other infrastructures,
- recruit new users.

We will introduce the the following communication channels:

**Project website:** the project partner OEAW will set up a website that will also serve as a repository of the materials accumulated and produced by ELEXIS. The website will contain generic information about the project, its partners and events (tools, documentation, datasets, events, promo materials, webinars, etc.). The project website will be maintained by JSI for at least three years after the duration of the project. In addition to the actual project website domain, platforms providing virtual access (LEX1, LEX2, LEX3) will be on separate domains, e.g. the Sketch Engine (<http://www.sketchengine.co.uk>) in LEX2.

**Digital promotion** and communication will include:

- Regular news items on the project website
- Videos, such as mini-interviews with experts from the field, mini-tutorials and hints & tips videos focused on deploying the infrastructure in various scenarios

- Regular social media posts for example on Twitter by the project on the ELEXIS account and by the consortium members on their institutional accounts or private accounts
- Sharing the above mentioned videos on social media like twitter: This posts will be designed to be appealing to a non-technical audience by including video content, currently the most shared content on the web
- Pictures and photo material as well as graphics representative for the ELEXIS project
- Presence of the project on research portals and within interest groups
- Twitter lists
- Newsletter updates to the existing subscribers
- Online questionnaire to assess the community's experience and opinion of the project's intermediate outcomes

**Videlectures.net:** this infrastructure developed by JSI for providing access to video lectures, slides or textual documents will be used. Videlectures.net is a multilingual service and provides also the facility for subtitling and (machine) translation of subtitles. It will be used for dissemination of scientific content from ELEXIS events.

**ELEXIS events:** A kick-off meeting will be held at the beginning of the project, and ELEXIS community fostering events in the second, third and fourth year of the project, training events.

**Third party events:** Dissemination will be carried out at joint events (e.g. workshops, conferences, project booth co-located at other international or national events) in order to reach out to other communities and enlarge the community of interest; the participation of single researchers with a talk or a poster with specific project results.

Besides the channels that will be set up by the ELEXIS project with the ELEXIS branding the project will use the communication and dissemination channels already established by the consortium partners (their website, their social media account, mailing lists) or other research infrastructures and initiatives (working groups, social media groups etc.). As one of the early tasks in the project we will update and adapt the dissemination and communication plan, outlined here in the proposal.

## 3 IMPLEMENTATION

---

### 3.1 WORK PLAN — WORK PACKAGES, DELIVERABLES

Work plan includes ten work packages:

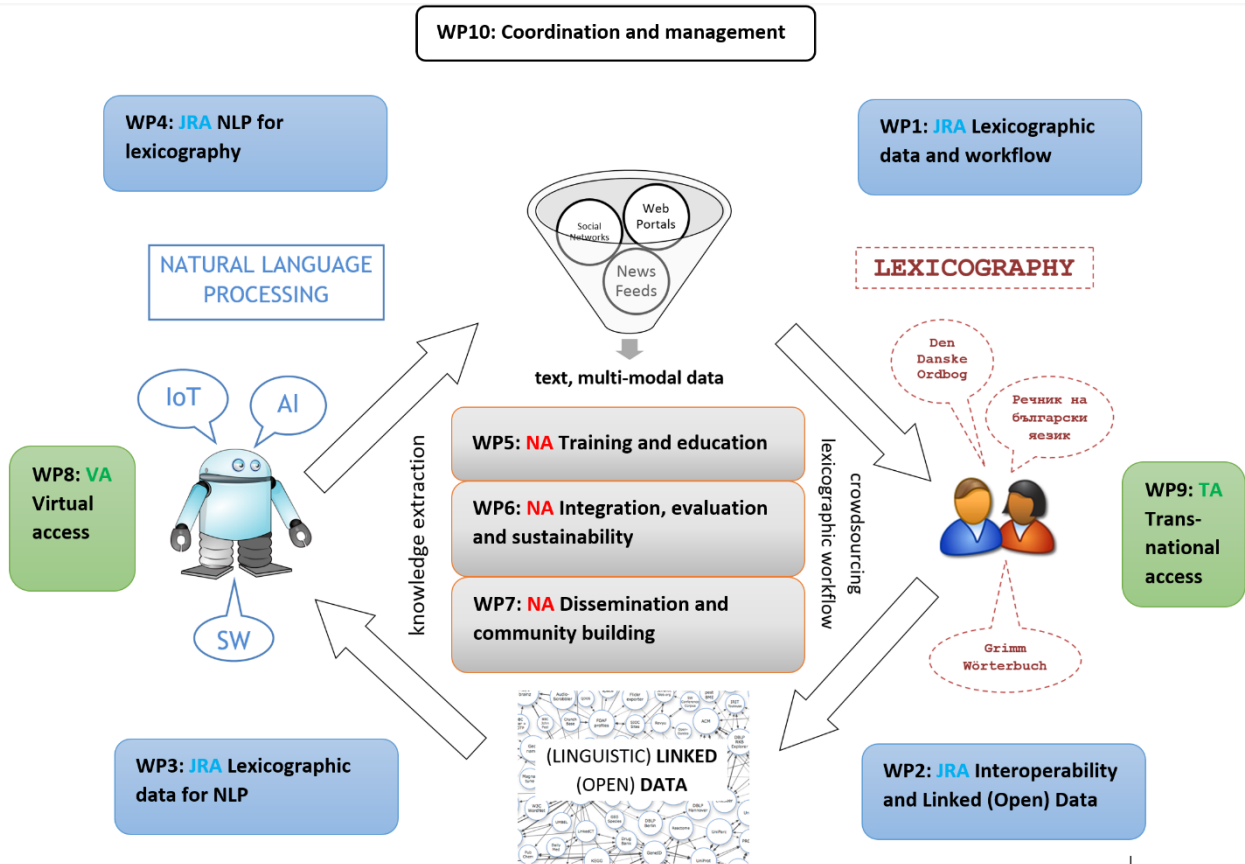
No	Name	Type
1	<b>JRA Lexicographic data and workflow</b>	Joint Research Activity
2	<b>JRA Interoperability and Linked (Open) Data</b>	Joint Research Activity
3	<b>JRA Data enrichment and analytics</b>	Joint Research Activity
4	<b>JRA Knowledge extraction for lexicography</b>	Joint Research Activity
5	<b>NA Training and Education</b>	Networking Activity
6	<b>NA Integration, evaluation and sustainability</b>	Networking Activity
7	<b>NA Dissemination and community building</b>	Networking Activity
8	<b>VA Virtual Access</b>	Virtual Access Activity
9	<b>TA Trans-National Access</b>	Trans-National Access Activity
10	<b>Coordination and Management</b>	

The first four work packages are defined as **joint research activities** covering the main points of the virtuous cycle of e-lexicography, as shown in Picture 4. The two work packages positioned at the top of the image (WP1, WP4), focus on the lexicographic part of the cycle. The two work packages at the bottom (WP2, WP3) focus on the NLP part.

The next three work packages cover ELEXIS **networking activities** ranging from dissemination, outreach and community building (WP5) to educational activities, development of modules and training material (WP6). As part of the networking efforts WP7 is dedicated to establishing interoperability of the technical architecture, solving important IPR issues and also securing interoperability with existing infrastructures (CLARIN and DARIAH) (WP7).

The two work packages dedicated to **virtual** (WP8) and **trans-national access** (WP9) cover core infrastructural activities providing online or on-site access to lexicographical resources, newly developed NLP tools and expertise to researchers from the field of lexicography, as well as to those from digital humanities, natural language processing, computational linguistics and artificial intelligence interested in working with lexicographic resources and tools. Finally, WP10 will coordinate and manage the project.

3.1.1 Dependency between work packages



Picture 4: Dependency between Work Packages

**Blue** = Joint Research Activity, **Grey** = Networking Activity, **Green** = Virtual Access / Trans-National Access

The virtuous cycle starts with **WP1** as the point of entrance for lexicographic data, harmonising the lexicographic landscape by producing standardised data formats. This data feeds into **WP2**, with linked lexicographic data as the product used in **WP3** for semantic processing of unstructured data. Which is in turn used in **WP4** to enable automatic creation of lexicographic content and better lexicographic description. Therefore, WP4 depends on the results of WP3 to conclude the cycle. These activities then further support WP1 starting another round of virtuous cycle.

The platform consisting of LEX1, LEX2 and LEX3 infrastructures in **WP8** (Virtual Access) depends on the results of Joint Research Activity work packages where the platform is developed on the basis of existing tools, services and expertise present in the consortium.

Networking activities in **WP5**, **WP6** and **WP7** also depend on the results of the research in Joint Research Activity work packages, and on the development of the platform in WP8, which is described in WP5 training materials, and disseminated in WP7. **WP9** (Trans-National Access) does not depend on other work packages.

3.1.2 Timing of Work Packages (Gantt Chart) and milestones

Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12 Q13 Q14 Q15 Q16  
 start finish M3 M6 M9 M12 M15 M18 M21 M24 M27 M30 M33 M36 M39 M42 M45 M48

	start	finish	M3	M6	M9	M12	M15	M18	M21	M24	M27	M30	M33	M36	M39	M42	M45	M48
<b>WP1: JRA - Lexicographic data and workflow</b>	<b>M1</b>	<b>M48</b>																
T1.1 User needs	M1	M12																
T1.2 Data collection and preparation	M1	M48																
T1.3 Best practices for lexicography	M3	M48																
T1.4 Data structuring for existing resources	M1	M24																
<b>WP2: JRA - Interoperability and Linked (Open) Data</b>	<b>M1</b>	<b>M48</b>																
T2.1 Common models and protocols for lexicon access	M1	M24																
T2.2 Semi-automatic linking of lexical resources	M12	M48																
T2.3 Cross-lingual mapping through shared conceptualisation	M18	M48																
T2.4 Validation and quality assurance for lexical resources	M36	M48																
<b>WP3: JRA - Lexicographic Data for NLP</b>	<b>M9</b>	<b>M48</b>																
T3.1 Word Sense Disambiguation and Entity Linking	M9	M42																
T3.2 Multilingual semantic parsing	M12	M48																
T3.3 Lexico-semantic analytics for NLP	M12	M48																
<b>WP4: JRA - NLP for Lexicography</b>	<b>M1</b>	<b>M48</b>																
T4.1 Lexico-semantic analytics for lexicography	M1	M24																
T4.2 Acquisition of lexicographic data from corpora	M1	M36																
T4.3 Enrichment of lexicographic resources	M6	M36																
T4.4 Crowdsourcing and gamification	M6	M18																
<b>WP5: NA - Training and Education</b>	<b>M3</b>	<b>M42</b>																
T5.1 Skills analysis	M3	M12																
T5.2 Training materials	M13	M42																
T5.3 Training measures	M13	M42																
<b>WP6: NA - Integration, evaluation and sustainability</b>	<b>M1</b>	<b>M48</b>																
T6.1 Impact assessment and reporting	M1	M48																
T6.2 Copyright and legal issues	M1	M24																
T6.3 Data seal of compliance	M1	M24																
T6.4 Compatibility with CLARIN/DARIAH services	M1	M48																
<b>WP7: NA - Dissemination and community building</b>	<b>M1</b>	<b>M48</b>																
T7.1 Dissemination and communication plan	M1	M14																
T7.2 Setting up communications tools	M1	M6																





## 3.2 MANAGEMENT STRUCTURE, MILESTONES AND PROCEDURES

### *ELEXIS consortium management structure*

Governance structure in ELEXIS consortium will include two main decision making bodies: the Project Management Board and the Technical Management Board.

The **Project Management Board (PMB)** is chaired by the Project Coordinator. It is responsible for the formal decisions and the strategic management of the project, and includes the Project Coordinator (PC), the Project Manager (PM) and one designated senior representative from each partner, forming an 18-member body altogether. PMB members are responsible to take binding decisions on behalf of the partner they represent. PMB meetings are held at least every six months (physically or virtually) in order to monitor and guide project. The tasks and responsibilities of the PMB include:

- Management and monitoring of project development according to the work plan.
- Guidance of the project with respect to external development and potential collaborations.
- Review of the general scientific and technical program and the project outcomes.
- Monitoring of the quality assurance plan and review and approval of the risk assessment.
- Conflict resolution management.
- Review and approval of financial issues.
- Supervision and advice with respect to issues of intellectual property management, dissemination, exploitation and promotion activities, assessment of ethical questions.

The **Technical Management Board (TMB)** is in charge of the implementation of the scientific and technical work, and monitors, on a daily basis, the progress of the project in accordance to the objectives, deliverables and milestones defined in the project work plan. The board includes the Project Coordinator and the Work Package Leaders (WPL) and meets at least quarterly (physically or virtually). The tasks and responsibilities of the TMB include:

- Definition of the scientific and technical roadmap for the project, in particular the coordination of the integration of work package results according to the work plan.
- Review and monitoring of the work plan and the general project progress.
- Supervision of the technology transfer process towards use case partners and potential adopters.
- Support to the PMB in issues of risk assessment, contingency planning, conflict resolution and revisions of the work plan.

The **Project Coordinator (PC)** has the overall responsibility for intermediation between the consortium and the European Commission, as well as for the contractual obligations defined in the Grant Agreement. The PC ensures the consistency of the overall work plan in close cooperation with the Work Package Leaders. His duties are to monitor the scientific and technical progress on a day-to-day basis and advise the PMB should any deviations from the project plan be necessary. **Simon Krek** will fill this role in ELEXIS.

The **Project Manager (PM)** provides support to the PMB with respect to various administrative, financial and organisational issues. The main responsibilities related to this role are the management of all the procedures related to the accounting and reporting towards the European Commission, the coordination of the administrative and financial relationships between project partners, and controlling that project activities are performed within budget constraints. The PM will be appointed by JSI, as coordinator of ELEXIS.

**Work Package Leaders (WPL)** coordinate and manage the corresponding work package activities on a day-to-day basis. They are responsible for the delivery of work-package-relevant results foreseen in the work plan and ensure that scientific and technical results of the work package are in line with the overall vision of the project and integrated with other work packages. WPLs report the current status of the work and potential deviations to the TMB and are involved in the quality assurance process when it comes to deliverables of their work package and in reporting, progress monitoring, risk management and contingency planning. Work package leaders will be: **Carole Tiberius** (WP1), **John McCrae** (WP2), **Roberto Navigli** (WP3), **Miloš Jakubiček** (WP4 and WP8), **Toma Tasovac** (WP5), **Simon Krek** (WP6 and WP10), **Tanja Wissik** (WP7) and **Bolette Sandford Pedersen** (WP9).

**International Advisory Board (IAB)** will be in advisory role and will consist of appointed representatives of five continental societies and associations for lexicography and experts in both relevant fields, lexicography and NLP, from academia and industry. The following organisation and experts agreed to participate in IAB:

**Organisations:**

Acronym	Organisation	Web site
<b>AFRILEX</b>	African Association for Lexicography	<a href="http://afrilex.africanlanguages.com/">http://afrilex.africanlanguages.com/</a>
<b>ASIALEX</b>	The Asian Association for Lexicography	<a href="http://www.asialex.org/">http://www.asialex.org/</a>
<b>AUSTRALEX</b>	Australasian Association for Lexicography	<a href="http://www.adelaide.edu.au/australex/">http://www.adelaide.edu.au/australex/</a>
<b>EURALEX</b>	European Association for Lexicography	<a href="http://euralex.org/">http://euralex.org/</a>
<b>DSNA</b>	Dictionary Society of North America	<a href="http://www.dictionariesociety.com/">http://www.dictionariesociety.com/</a>

**Experts:**

Member	Institution	Short bio
<b>Michael Rundell</b>	MacMillan Publishers	Michael Rundell was Managing Editor at Longman, Editorial Director of the DANTE project, and has been Editor-in-Chief of Macmillan Dictionary since 1997. He has been involved in developing several corpora (including the British National Corpus), and been at the forefront of applying computational techniques to the analysis of corpus data and the compilation of dictionary text. <a href="http://michaelrundell.com/">http://michaelrundell.com/</a>
<b>Ryan McDonald</b>	Google	Ryan McDonald is a Research Scientist at Google. He received a Ph.D. from the University of Pennsylvania and a Hon. B.Sc. from the University of Toronto. Ryan's thesis focused on the problem of syntactic dependency parsing. His work allowed complex linguistic constructions to be modeled in a direct and tractable way, which enabled parsers that are both efficient and accurate. In 2008 he wrote a book on the subject entitled Dependency Parsing. Since joining Google, Ryan has continued to work on syntactic analysis, in particular, extending statistical models learned on resource rich languages, like English, to resource poor languages. <a href="https://ryanmcd.github.io/">https://ryanmcd.github.io/</a>
<b>Christiane Fellbaum</b>	Princeton University	Christiane Fellbaum is a senior research scholar in the Computer Science Department at Princeton University. She is the co-developer and current director of the WordNet project, and the co-founder and co-president of

---

		the Global WordNet Association. Her honors include the Wolfgang Paul-Prize of the German Humboldt Foundation (2001) and the Antonio Zampolli Prize (2006). Her research Interests include theoretical linguistics, computational and corpus linguistics, and natural language processing. <a href="https://www.cs.princeton.edu/people/profile/fellbaum">https://www.cs.princeton.edu/people/profile/fellbaum</a>
<b>Piek Vossen</b>	Vrije Universiteit Amsterdam	Piek Vossen is full Professor of Computational Lexicology at the VU University Amsterdam, Head of the Computational Lexicology & Terminology Lab (CLTL), co-founder and co-president of the Global WordNet Association (GWA) and partner in the Centre for Digital Humanities Amsterdam. In 2013 he won the prestigious Spinoza Award of the Netherlands Organisation for Scientific Research (NWO) and in 2015 he has been honoured by the Dutch Royal House as a “Knight in the Order of the Dutch Lion”. He is involved in the many national and international projects, among which: Acquilex, EuroWordNet, Meaning, Cornetto, DutchSemCor, KYOTO and GlobalWordNetGrid. <a href="http://vossen.info/">http://vossen.info/</a>
<b>Gregory Grefenstette</b>	Florida Institute for Human and Machine Cognition	Gregory Grefenstette has been a scientific director at Exalead (2008-2013), senior research scientist at CEA LIST (2004-2008), principal research scientist at the Clairvoyance Corporation (2001-2004) and principal scientist at the Xerox Research Centre Europe (1993-2001). His research Interests include personal semantics, lifelogging, natural language processing, information retrieval, cross-language information retrieval and image annotation. <a href="https://www.lri.fr/~ggrefens/">https://www.lri.fr/~ggrefens/</a>
<b>Robert Lew</b>	Adam Mickiewicz University in Poznań	Robert Lew is a Professor at the Department of Lexicography and Lexicology at Adam Mickiewicz University in Poznań, Poland. He is the reviews editor of the International Journal of Lexicography (Oxford University Press) and member of EURALEX board. He has been involved as lexicographer, editor, or advisor in a number of major lexicographic projects for a number of publishers (including Harper-Collins, Pearson-Longman, Cambridge University Press and Macmillan).
<b>Wendalyn Nichols</b>	Cambridge University Press	Wendalyn Nichols is the publishing manager for lexical content in the English Language Teaching division of Cambridge University Press. She leads the online Cambridge Dictionary content strategy, data licensing, and new product development. A former teacher of ESL/ELT, Wendalyn became a lexicographer at the beginning of the “corpus revolution.” She has held senior roles in educational, trade, and custom publishing with Longman, Random House, and McMurry, in product commissioning and development, marketing, and operations, overseeing the transition from print to digital of a wide range of educational and consumer products. She is the publications committee chair for the Dictionary Society of North America and is a member of the advisory board for the graduate

---

program in linguistics at Montclair State University. She lives and works in New York.

**Philippe Climent** IDM Philippe Climent is the founder and head of IDM, with a computer science background. IDM team developed a dictionary writing system, DPS, which has been widely adopted by the dictionary industry and actors of lexicography. Currently IDM spans out in the USA, UK/Ireland and Mainland Europe, as well as the People’s Republic of China.

*Roles and boards in the project management*

Level	Management	Composition	Principal responsibilities
<b>Overall project</b>	Project Management Board (PMB)	Project Coordinator (PC), Project Manager (PM) and one representative of each partner	Strategic guidance of the project.
			Main decisions about project coordination, planning, direction and management.
			Conflict resolution and risk management.
			Quality assurance and progress monitoring.
	Technical Management Board (TMB)	Project Coordinator (PC), all Work Package Leaders (WPL)	Realisation of the technical directions and strategies of the project.
			Cross-work-package coordination.
			Daily project management, in particular quality assurance and progress monitoring.
	International Advisory Board (IAB)	Representatives of lexicographic associations, experts	
<b>Work package</b>	Work package	Representatives from partners as defined in Section 1.3	Coordination of the corresponding work package.
			Report on technical work progress in the work package.

*Intra-project communication*

Effective intra-project communication among partners is a key enabler for successful collaborative projects. Communication includes physical meetings to guarantee in-depth knowledge and idea exchange, integration between work packages and overall alignment and progress monitoring. The PMB and the TMB will meet at least twice and four times per year, respectively, while urgent matters can be addressed through email

exchange or phone conferences. In addition to this, the project will be supported by the following instruments:

- A shared folder repository for exchanging and storing all the documents relevant for the project. This repository will be structured following the natural structure of the project directories for the different work packages with internal sub-directories for the deliverables. The repository will also contain a directory for the documents and presentations related to the meetings.
- Four types of mailing lists will be available for communication and exchange of documents:
  - a project-level mailing list for consortium-internal discussions related to all WPs;
  - a PMB mailing list supporting the communication among members;
  - Work package mailing lists for coordination and internal discussion in each work package;
  - administrative mailing list for administrative and reporting tasks.
- A wiki will be available and used for collaborative tasks, such as the generation of periodical administrative reports, the definition of glossaries and project-internal roadmaps, etc.
- The project Web site covering all public outcomes of the project will be used as a means to access deliverables, publications, presentations, demonstrators and showcases, promotion material and to remain up-to-date with respect to dissemination and community building activities.

### *Decision-Making and Conflict Resolution*

Decision making is based on the consensus principle at every level of the project. At the work package level, consensus should be reached verbally through phone conferences and email exchange. At the TMB level, consensus could also be reached verbally or, if necessary, by majority voting. At the PMB level, the decisions require voting and a third-party majority to be taken. In this case, each partner will have one vote and the vote of the PC will act as a casting vote in the case of a tie. After consulting with the EC, the PMB will make the following decisions:

- Approval of budgets and timing and content of work plans
- Approval of major changes in the work to be delivered in the project
- Changes in the consortium and amendment of partners
- Proposed changes to the grant and consortium agreement
- Suspension or termination of the project or parts of it
- Actions to be taken in case of misconduct of a partner

All project participants are in charge of identification of any conflicts. Any disagreement should be reported to the work package leaders or the Project Coordinator who will pursue the mediation of the conflict or escalate it to higher levels, if necessary. The conflict resolution process, together with the decisions made, is recorded in a written report and logs of the communication between the parties involved. Both decision making and conflict resolution mechanisms are subject to the Consortium Agreement. These practices have proven effective in various European research projects over the past decade and build on the long-standing expertise of the partners in similar settings.

### *Quality Assurance and Progress Monitoring*

The project will install a quality assurance procedure that is the result of extensive experiences of the consortium in running collaborative research projects. A central component thereof applies to contractual deliverables. These undergo a three-step quality check before submitted to the European Commission. First, the deliverables are checked by a quality assessor. This is a project member external to the team which produced the deliverable. Reviewers affiliated to organisations outside the project consortium can be appointed as well upon approval by the TMB and in accordance with the IPR of the corresponding deliverables. The result of this first step is a review document consisting of recommendations to be

implemented by the authoring team. As subsequent steps, the deliverables are evaluated by the work package leader. The quality assurance procedure is initiated four weeks before the final submission deadline. The three steps are scheduled as follows:

- Four weeks before the deadline: the detailed review of technical quality of the deliverable is reported by the quality assessor.
- Three weeks before the deadline: the deliverable is revised by the authors according to the review report. The work package leader ensures that the deliverable implements the reviewer's comments and returns it to the authoring team if deemed appropriate.
- One week before the deadline: a final quality check is undertaken by the Project Coordinator before submitting it to the EC.

Project progress will be continuously monitored, and where discrepancies between plans and progress are observed (or predicted), corrective actions will be initiated. In particular, the PMB and TMB will carry out risk assessment at their regular meetings. This involves identifying project risks, assessing their probability and the nature of the consequences, should the risk be incurred. If the risk level is judged to be high, changes in project planning may be necessary. Decisions on any necessary re-planning of detailed tasks at the work package level will be made by the Work Package Leader (WPL), in consultation with all partners involved in the work package. Results should be reported to the Technical Management Board (TMB) and Project Coordinator (PC). Project level changes will be the responsibility of the Project Management Board (PMB), except in the case of major changes listed below. In addition to any reviews arising from regular risk assessment, the detailed project plan will be reviewed at least once per year and revised if necessary.

Certain types of re-planning may require the approval of the Commission, according to the terms of the Contract. It will be the responsibility of the Project Coordinator (PC) to contact the Commission regarding such matters. Project re-planning which results in changes deemed to be major must be handled by the Project Management Board (PMB), using voting procedures. Changes will be deemed to be major if any one partner protests about a proposed change, or if the change involves:

- Modifications to the Consortium Agreement or to the management structures and principles here described;
- Changes in project policy concerning ethical issues;
- Problems with the performance of any partner, or the desire of a partner to leave the consortium;
- Re-allocation of budget between work packages and/or partners.

Implementation of major changes may necessitate a change in the overall project plan, detailed project plans or the work breakdown structure of the project. As explained above, the management structure of the project essentially follows the work breakdown structure of the project. The management structure can, therefore, adapt to changes in the work breakdown structure. Further detail will be elaborated in the Consortium Agreement

### 3.3 CONSORTIUM AS A WHOLE

ELEXIS consortium contains 17 partners from 13 EU countries (Austria, Bulgaria, Czech Republic, Denmark, Estonia, Germany, Hungary, Ireland, Italy, Netherlands, Portugal, Slovenia and Spain), and two partners from Associated Countries – Serbia as an EU candidate country and Israel. The extent of the consortium reflects ELEXIS goal of bringing together research communities working in different fields, from lexicography to natural language processing and digital humanities, in an interdisciplinary effort to support lexicography. The consortium covers all the elements of the virtuous cycle of e-lexicography, and as such it consists of different types of partners fulfilling their role in different parts of the cycle:



- partners with lexicographic data and/or expertise
- partners with computational linguistics data and/or expertise
- partners with expertise in standardisation
- digital humanities partners
- technology partners.

**Partners with lexicographic data and expertise** cover the lexicographic side of the virtuous cycle, bringing quality lexicographic data and lexicographic expertise in the consortium. Eleven ELEXIS partners fall into this category, and they are also the ones who provide the 11 trans-national infrastructures described in WP9. Six of them are in the position of being the designated national language description institutions for various European languages: INT (Dutch), IBL (Bulgarian), DSL (Danish), RILMTA (Hungarian), EKI (Estonian) and RAE (Spanish). Other lexicographic partners are involved either in retro-digitisation and research on historical dictionaries (BCDH, TCDH), multi-lingual lexicography (KD), description of regional variants (OEAW), or production of novel types of dictionaries for human users (JSI).

**Partners with computational linguistics data and expertise** work as an intermediary between lexicographic and technology partners, as they use lexicographic data to produce computational linguistics resources, such as corpora, different types of computational lexicons (morphological, semantic), WordNets and similar. Many of them are also involved in CLARIN infrastructure and are leading institutions in national CLARIN consortia. We define 11 partners as providers of computational linguistics data and expertise, some of them also combine lexicographic and computational linguistics knowledge within the same institution (JSI, INT, OEAW, RILMTA, IBL, KD) or work on the same language in combination, e.g. Danish in case of DSL and UCPH. An important partner in this category is Lexical Computing with the Sketch Engine tool as a leading corpus management system which will be used as the core for the LEX2 infrastructure.

**Partners with expertise in standardisation** will work together on standards, both on the lexicographic and NLP side, which are one of the crucial building blocks of ELEXIS. Standardisation partners are already active in standardisation efforts in case of existing standards, such as Text Encoding Initiative (JSI, BCDH, INT), Lexical Markup Framework (CNR-ILC) or Ontolex-Lemon (NUIG, OEAW). Some of the partners are also directly involved in standardisation bodies, such as International Standards Organisation (FCSH-UNL - TC37 family of standards) or OASIS (LC). They will use their expertise in the process of defining common data models for lexicographic resources collected in the consortium.

**Digital humanities partners** are primarily responsible for educational aspects of the project, such as training materials, online training modules, and for the establishment of the data seal of compliance for lexicographic data. Most of them are members of DARIAH infrastructure and will bring expertise, tools and services from DARIAH, such as #dariahTeach learning platform. Digital Humanities partners include OEAW, BCDH and NUIG, as the national coordinators of DARIAH in Austria, Serbia and Ireland respectively. TCDH as another DH partner is involved primarily in retro-digitisation and will be responsible for LEX3 infrastructure. JSI in its DH role will provide learning platforms, such as free and open access educational video lectures repository Videolectures.net and ExplorEdu, a system of open and available web services and mobile applications for automatic identification of open educational resources (OER).

**Technology partners** have expertise in machine learning, data mining, artificial intelligence, Semantic Web, and other field in computer science, and are responsible for the NLP side of the virtuous cycle of e-lexicography. ELEXIS will combine expertise from a large research group working in the areas of machine learning, data mining, language technologies, semantic technologies and sensor networks at JSI, and from the Linguistic Computing Laboratory at UNIROMA1 working on knowledge acquisition, ontology learning and



population, word sense disambiguation and induction, semantic parsing and semantic information extraction. Technology partners include also those working on extensive corpora for various languages (LC) and/or Linked Data (NUIG, OEAW).

It is important to emphasise that the consortium includes two **industrial partners** (LC, KD) who will be responsible for the industrial/commercial involvement in the project and dissemination of results in the commercial (dictionary publishing) environment, as part of the exploitation plan.

The design of the areas and the associated work packages has been arranged carefully to ensure maximum efficiency of input from each partner while ensuring a suitable distribution of responsibilities. We have ensured that we have concentrated on covering all of the aspects required for successful innovation, development and exploitation of the project. The competences of the partners are described below. Many of the partners have proven experience in collaborating within EU projects, and most of them closely collaborated in the European Network of e-Lexicography COST action in the past four years. Some partners have also collaborated together before in other European projects. As a result, members of the consortium have established good personal relationships and a strong team ethic.

The following table summarises partners' competences:

	<i>Partners with lexicographic data and/or expertise</i>	<i>Partners with computational linguistics data and/or expertise</i>	<i>Partners with expertise in standardisation</i>	<i>Digital humanities partners</i>	<i>Technology partners</i>
1/JSI	X	X	X	X	X
2/LC		X	X		X
3/INT	X	X	X		
4/UNIROMA1		X			X
5/NUIG		X	X	X	X
6/OEAW	X	X	X	X	X
7/BCDH	X			X	
8/RILMTA	X	X			
9/IBL	X	X			
10/FCSH-UNL			X		
11/KD	X	X			
12/CNR-ILC		X	X		
13/DSL	X				
14/UCPH		X			
15/TCDH	X			X	
16/EKI	X				
17/RAE	X				

The table highlights that the project will be a truly synergistic experience, utilising the best approaches from the whole consortium while promoting cross-fertilisation of ideas. However, a main feature of the project is that it is fundamentally multi-disciplinary. The interdisciplinary nature will ensure that the ideas that are usually dealt with in isolated communities can be presented and seeded into other domains.

### 3.4 RESOURCES TO BE COMMITTED

The total number of requested person-months for the ELEXIS project is 738 in 10 work packages, of which four (WP1, WP2, WP3, WP4) are defined as Joint Research Activities, three as Networking Activities (WP5, WP6, WP7) and two as VA/TNA Activities (WP8, WP9). WP10 is dedicated to management. The distribution of staff effort in the four types of activities is as follows (also shown on p. 91):

<b>Type of Activity</b>	<b>Person/month percentage</b>	<b>Number of person-months</b>
Joint Research Activities	47,97%	354
Networking Activities	26,15%	193
Virtual and Trans-National Access	22,36%	165
Management	3,52%	26

The distribution of budget in the consortium reflects the responsibilities of partners as the coordinating partner (JSI) and work package leaders (LC, INT, UNIROMA1, NUIG, OEAW, BCDH and UCPH). In case of the coordinator the share also includes reimbursement of travel and subsistence for trans-national access which will be handled by the coordinator. The overall cost of travel and subsistence is shown in the Table 3.4b below and in the supporting document with estimations of the VA/TNA access costs in ELEXIS (ELEXIS\_Estimation\_of\_the\_Access\_Costs).

Table of costs:

3.4.1 Table 3.4b: 'Other direct cost' items (travel, equipment, other goods and services)

The share of other direct costs for travel, equipment, and 'goods and services' exceeds 15% of the personnel costs in case of the three partners which provide **virtual access** (JSI, LC, TCDH), and some of the partners that provide **trans-national access** (BCDH, MTANYTI, IBL, K DICTIONARIES, EKI, RAE) in ELEXIS. We justify the share of other direct cost below in the tables, one per partner:

Participant Number/ Short Name: 1/JSI	Cost (€)	Justification
<b>Travel &amp; subsistence for trans-national access</b>	6,800 €	For trans-national access in ELEXIS we use actual cost to declare access cost. These costs cover travel and subsistence costs for the three expected visiting grants at JSI – ELEXIS-SI infrastructure.
<b>Other Travel</b>	42,595 €	These costs will be used for JSI personnel to travel to project meetings, conferences and other events relevant for the project.
<b>Equipment</b>	0 €	
<b>Other goods and services</b>	52,000 €	The costs are split as follows: <ul style="list-style-type: none"> <li>- €40,000 – costs related to property rights (for lexicographic resources), translation costs etc.</li> <li>- €9,000 – dissemination and propagation costs</li> <li>- €3,000 – conference fees for presenting project-related research</li> </ul>
<b>Estimated direct eligible costs of providing virtual access</b>	20,000 €	Maintenance costs for LEX1 part of ELEXIS platform
<b>Total</b>	121.395 €	

Participant Number/ Short Name: 2/LC	Cost (€)	Justification
<b>Travel &amp; subsistence for trans-national access</b>	0 €	
<b>Other Travel</b>	33,065 €	These costs will be used for LC personnel to travel to project meetings, conferences and other events relevant for the project.
<b>Equipment</b>	0 €	
<b>Other goods and services</b>	15,000 €	10,000 € – dissemination and propagation costs 5,000 € – conference fees for presenting project-related research
<b>Estimated direct eligible costs of providing virtual access</b>	45,000 €	Costs for LEX2 part of ELEXIS platform: 25,000 € – Server housing in a data centre 20,000 € – Storage capacities and maintenance
<b>Total</b>	93,065 €	

Participant Number/ Short Name: 15/TCDH	Cost (€)	Justification
<b>Travel &amp; subsistence for trans-national access</b>	7,380 €	For trans-national access in ELEXIS we use actual cost to declare access cost. These costs cover travel and subsistence costs for the three expected visiting grants at TCDH – ELEXIS-DE infrastructure.
<b>Other Travel</b>	11,159 €	These costs will be used for TCDH personnel to travel to project meetings, conferences and other events relevant for the project.
<b>Equipment</b>	0 €	
<b>Other goods and services</b>	2,000 €	Conference fees for presenting project-related research
<b>Estimated direct eligible costs of providing virtual access</b>	5,000 €	Maintenance cost for LEX3 part of ELEXIS platform.
<b>Total</b>	25.539 €	

<b>Participant Number/ Short Name: 7/BCDH</b>	<b>Cost (€)</b>	<b>Justification</b>
<b>Travel &amp; subsistence for trans-national access</b>	6,120 €	For trans-national access in ELEXIS we use actual cost to declare access cost. These costs cover travel and subsistence costs for the three expected visiting grants at BCDH – ELEXIS-RS infrastructure.
<b>Other Travel</b>	28.528 €	These costs will be used for BCDH personnel to travel to project meetings, conferences and other events relevant for the project.
<b>Equipment</b>	0 €	
<b>Other goods and services</b>	2,000 €	These are dissemination and propagation costs, conference fees for presenting project-related research.
<b>Total</b>	36.648 €	

<b>Participant Number/ Short Name: 8/MTANYTI</b>	<b>Cost (€)</b>	<b>Justification</b>
<b>Travel &amp; subsistence for trans-national access</b>	6.120 €	For trans-national access in ELEXIS we use actual cost to declare access cost. These costs cover travel and subsistence costs for the three expected visiting grants at MTANYTI – ELEXIS-HU infrastructure.
<b>Other Travel</b>	10.796 €	These costs will be used for MTANYTI personnel to travel to project meetings, conferences and other events relevant for the project.
<b>Equipment</b>	0 €	
<b>Other goods and services</b>	1.000 €	These are dissemination and propagation costs, conference fees for presenting project-related research.
<b>Total</b>	17.916 €	

<b>Participant Number/ Short Name: 9/IBL</b>	<b>Cost (€)</b>	<b>Justification</b>
<b>Travel &amp; subsistence for trans-national access</b>	6.120 €	For trans-national access in ELEXIS we use actual cost to declare access cost. These costs cover travel and subsistence costs for the three expected visiting grants at IBL – ELEXIS-BG infrastructure.
<b>Other Travel</b>	15.292 €	These costs will be used for IBL personnel to travel to project meetings, conferences and other events relevant for the project.
<b>Equipment</b>	0 €	
<b>Other goods and services</b>	2.000 €	These are dissemination and propagation costs, conference fees for presenting project-related research.
<b>Total</b>	23.412 €	

<b>Participant Number/ Short Name: 11/K Dictionaries</b>	<b>Cost (€)</b>	<b>Justification</b>
<b>Travel &amp; subsistence for trans-national access</b>	6.624 €	For trans-national access in ELEXIS we use actual cost to declare access cost. These costs cover travel and subsistence costs for the three expected visiting grants at K Dictionaries – ELEXIS-IL infrastructure.
<b>Other Travel</b>	20.750 €	These costs will be used for K Dictionaries personnel to travel to project meetings, conferences and other events relevant for the project.
<b>Equipment</b>	0 €	
<b>Other goods and services</b>	1.000 €	These are dissemination and propagation costs, conference fees for presenting project-related research.
<b>Total</b>	28.374 €	

<b>Participant Number/ Short Name: 16/EKI</b>	<b>Cost (€)</b>	<b>Justification</b>
<b>Travel &amp; subsistence for trans-national access</b>	6.308 €	For trans-national access in ELEXIS we use actual cost to declare access cost. These costs cover travel and subsistence costs for the three expected visiting grants at EKI – ELEXIS-EE infrastructure.
<b>Other Travel</b>	5.750 €	These costs will be used for EKI personnel to travel to project meetings, conferences and other events relevant for the project.
<b>Equipment</b>	0 €	
<b>Other goods and services</b>	1.000 €	These are dissemination and propagation costs, conference fees for presenting project-related research.
<b>Total</b>	13.058 €	

<b>Participant Number/ Short Name: 17/RAE</b>	<b>Cost (€)</b>	<b>Justification</b>
<b>Travel &amp; subsistence for trans-national access</b>	7.080 €	For trans-national access in ELEXIS we use actual cost to declare access cost. These costs cover travel and subsistence costs for the three expected visiting grants at RAE – ELEXIS-SP infrastructure.
<b>Other Travel</b>	4.040 €	These costs will be used for RAE personnel to travel to project meetings, conferences and other events relevant for the project.
<b>Equipment</b>	0 €	
<b>Other goods and services</b>	1.000 €	These are dissemination and propagation costs, conference fees for presenting project-related research.
<b>Total</b>	12.120 €	

### 3.5 REFERENCES AND BIBLIOGRAPHY

- Aguado-de-Cea, G., Montiel-Ponsoda, E., Kernerman, I. & Ordan, N. (2016) From dictionaries to cross-lingual lexical resources. *META Forum*, Lisbon. [http://www.meta-net.eu/events/meta-forum-2016/slides/29\\_cea.pdf](http://www.meta-net.eu/events/meta-forum-2016/slides/29_cea.pdf).
- Boguraev B. and Briscoe T. (Eds.). (1989) *Computational Lexicography for Natural Language Processing*. Longman Publishing Group, White Plains, NY, USA.
- Costa, R. (2013). Terminology and Specialised Lexicography: two complementary domains. In *Lexicographica. International Annual of Lexicography*. Volume 29, Issue 1. Berlin, New York: De Gruyter, pp. 29–42. DOI: 10.1515/lexi-2013-0004.
- Declerck, T. & Mörth, K. (2016). Towards a Sense-based Access to Related Online Lexical Resources. In: T. Margalidze & G. Meladze (Eds.): *Proceedings of the XVII EURALEX International Congress*. Tbilisi, Georgia, pp. 660-668.
- Eszter S., Benyeda, I. Z., Koczka, P. & Ludányi, Z. (2015). Automatic creation of bilingual dictionaries for Finno-Ugric languages. In: *Proceedings of the First International Workshop on Computational Linguistics for Uralic Languages*, Tromsø, pp. 119-131.
- Gantar, P., Kosem, I., Krek, S. (2016) Discovering Automated Lexicography: The Case of the Slovene Lexical Database. In *Int J Lexicography*. 29 (2): 200-225. DOI: <https://doi.org/10.1093/ijl/ecw014>.
- Kallas, J., Kilgarriff, A., Koppel, K., Kudritski, E., Langemets, M., Michelfeit, J., Tuulik, M. & Viks, Ü. (2015). Automatic generation of the Estonian Collocations Dictionary database. *Proceedings of the eLex 2015 conference*. 1–20.
- Khan, F., Diaz-Vera, J. E. & Monachini, M. (2016). Representing Polysemy and Diachronic Lexico- Semantic Data on the Semantic Web. *Proceedings of the Second International Workshop on Semantic Web for Scientific Heritage Co-Located with 13th Extended Semantic Web Conference (ESWC 2016)*, Heraklion, pp. 37-45.
- Kilgarriff, A., et al. (2014) The Sketch Engine: ten years on. In *Lexicography*, 1–30.
- McCrae, J. P., Aguado-de-Cea, G., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr D. & Wunner, T. (2012). Interchanging lexical resources on the Semantic Web. *Language Resources and Evaluation*, 46(6), 701-709.
- McCrae, J. P., Chiarcos, C., Bond, F., Cimiano, P., Declerck, T., de Melo, G., Gracia, J., Hellmann, S., Klimek, B., Moran, S., Osenova, P., Pareja-Lora, A. and Pool, J. (2016) “The Open Linguistics Working Group: Developing the Linguistic Linked Open Data Cloud.” 10th Language Resources and Evaluation Conference.
- Měchura, M. (2017) Introducing Lexonomy: an open-source dictionary writing and publishing system. (to appear). *Proceedings of eLex conference*, Leiden 2017.
- Pedersen, B. S, Nimb, S., Asmussen, J., Sørensen, N., Trap-Jensen, L., Lorentzen, H. (2009). DanNet – the challenge of compiling a WordNet for Danish by reusing a monolingual dictionary. *Language Resources and Evaluation*, Computational Linguistics Series, pp. 269-299.
- Pilehvar, M. T. and Navigli, R. A. Robust Approach to Aligning Heterogeneous Lexical Resources. (2014) *Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014)*, Baltimore, USA, June 22-27, pp. 468-478.
- Stoyanova I., Koeva S., Todorova M. & Leseva S. (2016). Semi-automatic Compilation of a Very Large Multiword Expression Dictionary for Bulgarian. In: *Proceedings of GLOBALEX 2016: Lexicographic Resources for Human Language Technology*, Workshop at LREC2016, Portorož, Slovenia, pp. 86–95.
- Tasovac, T. (2012). Potentials and challenges of WordNet-based pedagogical lexicography: The Transpoetika Dictionary. In S. Granger & M. Paquot (Eds.), *Electronic Lexicography*. Oxford: Oxford University Press, pp. 237-258.
- Urdang L. (1966). “The Systems Designs and Devices Used to Process: The Random House Dictionary of the English Language”. *Computers and the Humanities* 1 (2).



## 4 SECTION 4: MEMBERS OF THE CONSORTIUM

### 4.1 4.1. PARTICIPANTS (APPLICANTS)

#### 1. "Jožef Stefan" Institute, Ljubljana, Slovenia (JSI)

##### Role in the project (cf. Consortium as a whole)

- Coordinator
- Work package leader (WG6, WG10)
- Partner with lexicographic data and expertise
- Partner with computational linguistics data and expertise
- Partner with expertise in standardisation
- Digital humanities partner (e-learning, digital editions)
- Technology partner (machine learning, data mining, artificial intelligence)

##### Description of the legal entity

Jožef Stefan Institute (JSI) is the leading research institution for natural sciences in Slovenia having over 900 researchers within 25 departments working in the areas of computer science, physics, and chemistry and biology. Three units of the Institute will collaborate in the proposed project.

The **Artificial Intelligence Laboratory**, having approx. 40 researchers, is one of the largest European research groups working in the areas of machine learning, data mining, language technologies, semantic technologies and sensor networks. The key research direction is combining modern statistical data analytic techniques with more semantic/logic based knowledge representations and reasoning techniques with the purpose to progress in solving complex problems such as text understanding, large scale probabilistic reasoning, building broad coverage knowledge bases, and dealing with scale. The members have developed several software tools for multimodal data analysis, among others: the Text-Garden suite of text mining tools, the OntoGen system for ontology learning, the Document-Atlas for complex visualization, the AnswerArt system for semantic search over large textual databases, the Enrycher system for semantic enrichment of textual data, SearchPoint system for visual and contextualized Web browsing, XLing for cross-lingual matching and categorization across 100 languages, and Event Registry for global real-time media observatory.

The **Centre for Knowledge Transfer in Information Technologies** has approx. ten researchers and technical staff working in the areas of research results dissemination and eLearning. In particular, the centre is well known by portals: VideoLectures.NET with multimedia materials of numerous scientific events, on-line training materials, and collection of tutorials on different scientific fields; ScienceAtlas.ijs.si and IST-World.Org for analysis and visualization of large bibliographic and project databases. The centre is covering management, training and dissemination activities of several EU projects.

The **Department of Knowledge Technologies** has a staff of 35 researchers and 15 PhD students and external collaborators. The Department performs research in advanced information technologies aimed at acquiring, storing and managing knowledge to be used in the development of an information- and knowledge-based society. Established areas of our work include intelligent data analysis (machine learning, data mining, and knowledge discovery in databases), computational creativity, decision support and knowledge management.

The Department is also a recognised centre of research on language technologies, computational linguistics, corpus linguistic, and digital humanities. It has been involved in the compilation of the majority of existing Slovene mono- and multilingual corpora, the development of manually annotated corpora for training language annotation tools, the development of the tools, such as part-of-speech taggers, lemmatisers, parsers and named entity recognisers and in work on standardisation of linguistic encoding in the scope of TEI and ISO. The department is also the home of the Slovene research infrastructure CLARIN.SI, a member of the CLARIN ERIC.

### Short CVs of main team members

**Simon Krek** (PhD) (M) is a researcher and project manager in the Artificial Intelligence Laboratory. He received a PhD in linguistics from the University of Ljubljana. The topic of PhD was automatic extraction of lexicographic data from corpora. His main areas of research are lexicography and lexicogrammar, corpus linguistics, natural language processing, language technology infrastructure and computer-aided language learning and teaching. From 1994-2004 he worked as a dictionary editor at DZS publishing house in Ljubljana, leading a ten-year project of compiling a new English-Slovenian dictionary as the editor-in-chief. From 2008-2013 he coordinated a language resources and technologies project "Communication in Slovene" which produced large written and spoken corpora, web concordancers, natural language processing tools and resources for Slovene (tagger, parser, morphological lexicon, lexical database) and two language learning web portals. He participated in several EU projects (META-NET, XLIKE, XLIME). He is one of the initiators of the Slovene language infrastructure CLARIN.SI (<http://www.clarin.si>) and one of the representatives in the CLARIN.SI consortium from JSI. He is a member of EURALEX board, and the chair of the Innovative e-dictionaries Working Group in European Network of e-Lexicography COST action.

**Dunja Mladenić** (PhD) (F) is a researcher, project manager and head of the Artificial Intelligence Laboratory. She is an active researcher in the area of machine learning, text mining and semantic Web. She graduated in Computer Science from the University of Ljubljana and continued as a PhD student focused on Artificial Intelligence. She is the author and editor of several books and papers on machine learning, data/text mining and semantic technologies. She has experience in coordinating EU projects and acting on management board of several EU FP5, FP6 and FP7 projects. She was a program co-chair of ECML 2007, a general chair of ECMLPKDD 2009. Dunja Mladenić is the Slovenian representative in EC Enwise STRATA ETAN Expert Group, she serves as project evaluator and reviewer for various EC programmes.

**Marko Grobelnik** (M) is a researcher and project manager in the Artificial Intelligence Laboratory. He is an expert in the areas of analysis and knowledge discovery in large complex databases. He collaborates with major European and US academic institutions and consults industries such as British Telecom, Microsoft Research, Nature, New York Times, Bloomberg, and Accenture, and is the author of several books in the area of machine learning, data mining, text mining and semantic technologies and authors of many scientific papers. He is also W3C AC representative for IJS, CEO of the company Quintelligence and co-founder of the company Cycorp Europe. In terms of the past project experience, he has been technical coordinator for projects FP6 IST-World and FP7 VID1 and coordinator of FP7 project X-LIKE (Cross-Lingual Knowledge-Extraction); he was a member of project management board in several FP6 and FP7 Projects (SEKT, NEON, ACTIVE, COIN).

**Mitja Jermol** (M) is the holder of UNESCO Chair on Open Technologies for Open Educational Resources and Open Education and a head of the Centre for Knowledge Transfer at Jozef Stefan Institute in Slovenia. Center has setup and is being running one of the world's top open scientific video portal <http://videolectures.net>. Mitja's research is focused on using Artificial Intelligence in the context of Business Intelligence, Personalised Learning, Smart Cities and Factories of the Future. Mitja has been active in more than 20 H2020, FP7 and FP6 projects in these areas. Together with his team he has

initiated a nation-wide initiative on developing and deploying open education called “Opening Up Slovenia”. He is a member of the European Complex System Society, Open Education Consortium, Opencast and European current research information systems (CRIS) community. Mitja is a co-founder of the two institute spin-outs, Quintelligence and Cycorp RER. Mitja was also conference chair for “Open Course Ware Consortium” Global Conference 2014.

**Jan Rupnik** (PhD) (M) is a researcher at Artificial Intelligence Laboratory. His research interests include Machine Learning, Data Mining, Data Fusion, Cross-Lingual Text Mining, Predictive Analytics, Applications of Data Mining in different domains. Most of his research work is connected to the development of statistical methods that enable crossmodal data integration, with focus on scalability. He defended his PhD thesis titled “Multi-View Canonical Correlation Analysis” in 2016. Jan Rupnik has been involved in a number of EU FP7 projects, including SMART (Statistical Multilingual Analysis For Retrieval And Translation), XLIKE (Cross-lingual Knowledge Extraction), EURIDICE (The Intelligent Cargo Concept in the European Project), and SOPHOCLES (Self-Organised information PrOcessing, Criticality and Emergence in multilevel Systems).

**Tomaž Erjavec** (PhD) (M) is a senior scientific associate at the Department of Knowledge Technologies and was previously also employed at the University of Edinburgh and University of Tokyo and as a detached national expert at the EU Joint Research Centre, Ispra. He is an expert on language technologies, with a focus on the Slovene and other Slavic languages, multilingual applications, methods and standards for the compilation and annotation of language resources. He was a lecturer in the fields of Corpus linguistics and Human language technologies at University of Ljubljana, University of Nova Gorica, University of Graz and Jožef Stefan Postgraduate School. Erjavec has over 100 publications in journals and conference proceedings, is on the editorial board of Journal for Language Resources and Evaluation and Journal of Corpus Linguistics; he served on the Board of the European Chapter of the Association for Computational Linguistics and Council of the Text Encoding Initiative Consortium and was the founding president of the Slovenian Language Technologies Society. He is the official member of the Slovenian standards organisation SIST in ISO TC 37 SC4 “Language resource management” and the national coordinator of the CLARIN.SI research infrastructure for language resources and tools.

**Nikola Ljubešić** (PhD) (M) is a researcher in the Department of Knowledge Technologies and Assistant Professor at the Department of Information and Communication Sciences at the Faculty of Humanities and Social Sciences, University of Zagreb, Croatia, where he teaches courses in natural language processing on both undergraduate and graduate level. His research interests cover a broad spectrum of NLP problems including web corpora construction, terminology extraction, bilingual lexicon extraction from comparable corpora and machine translation.

**Iztok Kosem** (PhD) (M) is currently employed at the Faculty of Arts, University of Ljubljana, and as the director of Trojina, Institute for Applied Slovene Studies (if the proposal is approved, he would be employed at JSI). His main areas of research are lexicography and lexicogrammar, corpus linguistics, crowdsourcing, and computer-aided language learning and teaching. From 2000-2005 he worked as a lexicographer on an English-Slovenian dictionary at the DZS publishing house in Ljubljana, and after completing his PhD studies in the UK in 2010, he assumed the role of Assistant coordinator on the Communication in Slovene project that produced a number of language technologies and resources for Slovene. Currently Izток Kosem is a project leader of two national projects, Conceptualising the Hungarian-Slovenian dictionary and Šolar 2.0, and is also one of the editors of the Collocations dictionary of Slovene. Finally, he is Vice Chair of European Network of e-Lexicography COST action, as well as a Slovenian representative on the management committee.

**All the named persons are employed by the beneficiary or will be employed after project starts (Iztok Kosem) or have an equivalent relationship covering the eligibility of costs.**

## Relevant publications

- Gantar, P., Kosem, I., Krek, S. Discovering Automated Lexicography: The Case of the Slovene Lexical Database. In *Int J Lexicography* (2016) 29 (2): 200-225. DOI: <https://doi.org/10.1093/ijl/ecw014>.
- Mladenić, D., Grobelnik, M. Automatic text analysis by artificial intelligence. *Informatica*, ISSN 0350-5596, 2013, vol. 37, no. 1, pp. 27-33
- Rupnik, J., Muhič, A., Leban, G., Škraba, P., Fortuna, B., Grobelnik, M.. News across languages - cross-lingual document similarity and event tracking. *The journal of artificial intelligence research*, ISSN 1076-9757, 2016, vol. 55, pp. 283-316, doi: 10.1613/jair.4780.
- Erjavec, T.. MULTEXT-East : morphosyntactic resources for Central and Eastern European languages. *Language resources and evaluation*, ISSN 1574-020X, 2012, vol. 46, no. 1, str. 131-142, doi: 10.1007/s10579-011-9174-8.
- Ljubešić, N., Dobrovoljc, K., Fišer, D.. \*MWElex - MWE Lexica of Croatian, Slovene and Serbian extracted from parsed corpora. *Informatica*, ISSN 0350-5596, Sep. 2015, no. 3, vol. 39, str. 293-300.

## Previous projects/activities

### **XLime - Objective ICT-2013.4.1 Content analytics and language technologies,**

FP7 Collaborative project: cross-lingual cross-media knowledge extraction. The goal is to connect different modalities of real-time data with an emphasis on hundreds TV streams, social media data and e-commerce data. The research goal is to establish cross-media probabilistic semantic space where all the different data modalities could be mapped, compared and on the top of which new services could be provided. The project is challenging from the side of scale, velocity of data and data complexity.

### **X-Like (ICT-2011.4.2) - Language Technologies**

FP7 STREP: Cross Lingual Knowledge Extraction, coordinated by JSI-AI Lab. The goal of the X-Like project was to develop technology to monitor and aggregate knowledge that is currently spread across mainstream and social media, and to enable cross-lingual services for publishers, media monitoring and business intelligence. The aim was to combine scientific insights from several scientific areas to contribute in the area of cross-lingual text understanding. The developed technology will be language-agnostic, while within the project we will specifically address English, German, Spanish, Chinese and Hindi as major world languages and Catalan and Slovenian as minority languages.

### **RENDER (FP7 – 257790) - Content analytics**

RENDER project provided a comprehensive conceptual framework and technological infrastructure for enabling, supporting, managing and exploiting information diversity in Web-based environments. This was demonstrated through the usage of realistic data sources, including news streams covering over 5000 sources worldwide with 100,000 items per day, (micro) blog streams adding up to more than a million posts per day, a full data stream from Wikipedia, and the Linked Open Data Cloud; through open source extensions to popular collaboration and communication platforms such as MediaWiki.

### **TransLectures (Objective ICT-2011.4.2) - Language Technologies:**

FP7 STREP: Transcription and Translation of Video Lectures. The goal is to develop innovative, cost-effective tools for the automatic transcription and translation of online educational videos. Specifically, the project is developing tools for use on VideoLectures.NET, a collection of videos recorded at various

academic events set up by Jožef Stefan Institute, and for poliMedia, a lecture capture system designed and implemented at the Universitat Politecnica de Valencia.

### Communication in Slovene - Language Technologies, Lexicography. E-learning

Web site: <http://eng.slovenscina.eu/>. The project was financed by European Social Fund and Ministry of Education, Science and Sports. The most important results include language learning web portals, lexical data — lexical database with semantic, syntactic, collocational and other data, and Slovene lexicon Sloleks with inflectional paradigms for 100,000 Slovene words, text corpora of written and spoken language. These resources can be explored online by using web concordancers developed in the project. The project developed some of the basic language technology applications and resources needed for computer processing of text in the Slovene language, such as tagger, parser etc.

#### Description of infrastructure

Pre-existing tools and services:

- Event Registry and Enrycher <http://enrycher.ijs.si>, <http://eventregistry.org>
- Ontogen, Textgarden, GLib <http://ontogen.ijs.si/>, <http://ailab.ijs.si/dunja/textgarden/>
- QMiner <https://github.com/qminer/qminer>
- xling <http://xling.ijs.si/>
- wikifier <http://wikifier.ijs.si>
- Videolectures.net <http://videolectures.net>

Lexicographic data:

- Slovene Lexical Database <http://eng.slovenscina.eu/spletni-slovar/leksikalna-baza>

## 2. Lexical Computing CZ s.r.o. (LC)

#### Role in the project (cf. Consortium as a whole)

- Work package leader (WG4, WG8)
- Partner with computational linguistics data and expertise
- Partner with expertise in standardisation
- Technology partner (machine learning, data mining, artificial intelligence)

#### Description of the legal entity

**Lexical Computing** is a research company founded by Dr. Adam Kilgarriff in 2003. It works at the intersection of corpus and computational linguistics, and is committed to an empiricist approach to the study of language, in which corpora play a central role: for a very wide range of linguistic questions, if a suitable corpus is available, it will help our understanding. Its strap line is ‘corpora for all’.

It has a leading corpus query tool, the Sketch Engine, incorporating ‘word sketches’, one page corpus-driven summaries of word’s grammatical and collocational behaviour. The lead users for the Sketch Engine have been dictionary publishers and it is in day-to-day use for lexicography at Oxford University Press, Cambridge University Press, Collins, Macmillan, Cornelsen and the Instituut voor de Nederlandse Taal (INT, Dutch Language Institute) among others.

To be able to provide corpus services, Lexical Computing needs large text collections - text corpora. As of March 2017 we have corpora for over 90 languages, many of them with more than a billion of words. For the most part, these are collected from the web – Lexical Computing is a lead player in the ‘web as corpus’ initiative – and have involved collaborations with language experts for the languages in question.

### Short CVs of main team members

**Miloš Jakubíček** (M) is CEO of Lexical Computing, an NLP researcher and software engineer. His research interests are devoted mainly to two fields: effective processing of very large text corpora and parsing of morphologically rich languages. For the past six years he has been involved in the development of the Sketch Engine corpus management suite, working on fast indexation and querying of billion words corpora. He is Director of Lexical Computing CZ, s.r.o. responsible for the development of core products of the Lexical Computing group.

**Pavel Rychlý** (PhD) (M) is a software architect and assistant professor at Masaryk University in Brno and will function in the project as a leading researcher in the area of the building large corpora, corpus tools and their effective exploitation. As a leading programmer he will be helping other team members with the technical tasks. Dr. Pavel Rychlý works in the field of corpus processing and building for more than 15 years. In 2000, he received Ph.D. in Informatics from Masaryk University on the thesis of "Corpus Managers and their effective implementation". He is the author and the main developer of widely-used corpus systems Manatee, Bonito and Sketch Engine. His research interests are intelligent processing of very large corpora, lexical acquisition and natural language processing in general.

**Ondřej Matuška** (M) manages digital services and user experience at Lexical Computing. He has a strong background in English language teaching publishing and learner dictionary publishing as Macmillan Education sales manager for the Czech Republic. He is professionally interested the exploitation of text corpora and corpus tools for foreign language acquisition. Ondřej Matuška is also involved in the development of the new Sketch Engine user interface to facilitate the use of corpus generated information by non-academic audience such as translators, interpreters, teachers and students.

**Vojtěch Kovář** (PhD) (M) is a researcher in computational linguistics, especially interested in corpus processing, language modelling and syntactic analysis (mainly English and Czech), and their real-world applications in language technology. He is also the head of the Sketch Engine support team. He has been developing the front-ends of Sketch Engine for a long time as well as designing many of the underlying analytic algorithms. In 2014 he defended his PhD thesis on application-driven syntactic analysis at Masaryk University in Brno, also in relation to corpus analysis.

**David Filip** (PhD) (M) is Chair (Convener) of OASIS XLIFF OMOS TC; Secretary, Editor and Liaison Officer of OASIS XLIFF TC; a former Co-Chair and Editor for the W3C ITS 2.0 Recommendation; and co-moderator of the Interoperability and Standards WG at JIAMCATT. He has been also appointed as NSAI expert to ISO TC37 SC3 and SC5, ISO/IEC JTC1 WG9, WG10 and SC38. His specialties include open standards and process metadata, workflow and meta-workflow automation. David works as a Research Fellow at the ADAPT Research Centre, Trinity College Dublin, Ireland. Before 2011, he oversaw key research and change projects for Moravia’s worldwide operations. David held research scholarships at universities in Vienna, Hamburg and Geneva, and graduated in 2004 from Brno University with a PhD in Analytic Philosophy. David also holds master’s degrees in Philosophy, Art History, Theory of Art and German Philology.



**Michal Měchura (M)** is a software developer and language engineer. He specializes on building online dictionaries and lexical databases, designing multilingual websites and developing software for digital humanities. He works as a technical consultant for Fiontar in Dublin City University and for the New English-Irish Dictionary Project in Foras na Gaeilge. He is also a PhD student at the Natural Language Processing Centre in Masaryk University. He was involved in many Irish lexicographic projects including the National Terminology Database for Irish and the Dictionary and Language Library.

**All the named persons are employed by the beneficiary or will be employed after project starts or have an equivalent relationship covering the eligibility of costs (David Filip, Michal Měchura).**

#### Relevant publications

- Kilgarriff, A., et al. The Sketch Engine: ten years on. In *Lexicography* (2014): 1–30.
- Kilgarriff, A., Rychlý P., Smrž P., Tugwell D.. The Sketch Engine. In *Proceedings of EURALEX 2004*, Lorient, France; Pp. 105–116,
- Kovář, V., Baisa, V., Jakubíček M. Sketch Engine for Bilingual Lexicography. In: *International Journal of Lexicography*, July 2016, doi: 10.1093/ijl/ecw029
- Jakubíček, M., Kilgarriff A., Kovář, V., Rychlý P., Suchomel V. The TenTen Corpus Family. In *7th International Corpus Linguistics Conference CL 2013*

#### Previous projects/activities

**DIGILING** is an Erasmus+ project with the goal of meeting the increasing European job market demands for digitally skilled linguists, it started in 2016 and will finish in 2019, see

<http://www.digiling.eu/>

**PRESEMT** was an EU, FP7 project in the area of 'hybrid' Machine Translation, combining statistical and rule-based methods: see <http://www.presemt.eu>. It ran from 2010 till 2012.

**KELLY** was an EU Lifelong Learning Programme project, preparing wordlists for learners and word cards for nine languages, see <http://www.kellyproject.eu>. It ran from 2009 to 2011.

**DANTE** was a project funded largely by the Irish Government, has produced a very-high-quality lexical database for English: see <http://webdante.com>. It ran from 2007 to 2010.

#### Description of infrastructure

Lexical Computing operates at the moment a cluster of servers hosting the Sketch Engine services, together with a set of large disk array hosting the underlying textual data (about 300 TB as of 2017). Lexical Computing has a long-term experience with providing services and support covered by a Service Level Agreement to its customers, both in academia and industry.

Pre-existing tools and services:

- |                  |                                                                                                               |
|------------------|---------------------------------------------------------------------------------------------------------------|
| • NoSketchEngine | <a href="https://www.sketchengine.co.uk/nosketch-engine/">https://www.sketchengine.co.uk/nosketch-engine/</a> |
| • Sketch Engine  | <a href="https://www.sketchengine.co.uk/">https://www.sketchengine.co.uk/</a>                                 |
| • Lexonomy       | <a href="http://www.lexonomy.eu/_en/">http://www.lexonomy.eu/_en/</a>                                         |
| • Corpus Tools   | <a href="http://corpus.tools">http://corpus.tools</a>                                                         |

### 3. Instituut voor de Nederlandse Taal (Dutch Language Institute, INT)

#### Role in the project (cf. Consortium as a whole)

- Work package leader (WG1)
- Partner with lexicographic data and expertise
- Partner with computational linguistics data and expertise
- Partner with expertise in standardisation

#### Description of the legal entity

**The Instituut voor de Nederlandse Taal (INT)** is a research institute financially supported by the governments of the Netherlands and (Flemish) Belgium. It takes a central position in the Dutch-speaking world (including Vlaanderen, Suriname and the Netherlands Antilles) as a developer, keeper and distributor of sustainable language resources through the centuries, using reliable scholarly methods.

There are approximately 30 employees working at the INT. The institute has a long-standing experience in building corpora and dictionaries, in linguistic processing and enrichment of language resources, and in developing linguistic and lexicographical retrieval applications. It is known as the institute where the large scholarly dictionaries of the Dutch Language have been compiled: i.e. the *Oudnederlands Woordenboek* (ONW, Dictionary of Old Dutch; ca. 475 – 1200), the *Vroegmiddelnederlands Woordenboek* (VMNW, Dictionary of Early Middle Dutch; 1200 – 1300), the *Middelnederlandsch Woordenboek* (MNW, Dictionary of Middle Dutch; ca. 1250 – 1550), the *Woordenboek der Nederlandsche Taal* (WNT, Dictionary of the Dutch Language; ca. 1500 – 1976). Currently the institute is compiling a new dictionary of contemporary Dutch, the *Algemeen Nederlands Woordenboek* (ANW, Dictionary of contemporary Dutch).

The institute also has an official role as a repository for digital language resources for the Dutch language previously hosted at the 'TST-Centrale' (Agency for Human Language Technology). As such it has ample experience in acquisition, management, and maintenance of linguistic resources and tools, and making these available for education, research and development.

The institute has participated in the EC funded projects PAROLE, SIMPLE, TELRI, ENABLER, ELAN, IMPACT, SUCCEED and Transcriptorium. INT is a member of the CLARIN common languages resource infrastructure network and is one of the CLARIN centers in the Netherlands with a specific focus on resources that are relevant to the lexicological study of the Dutch language and on resources relevant for research in and development of language and speech technology. INT is also a board member of the IMPACT Center of Competence (<http://www.digitisation.eu/>).

#### Short CVs of main team members

**Carole Tiberius** (PhD) (F) is a computational linguist. After degrees in translation and computational linguistics, she obtained a PhD on multilingual lexical knowledge representation from the University of Brighton. Before joining INT, she worked as a linguist on typological databases at the University of Surrey. At INT she is involved in lexicographic projects (mainly *Algemeen Nederlands Woordenboek*) and the maintenance of externally created bilingual lexical resources. From 2011-2015, she has been leader of the ICT track of the Language Portal project, an online portal for the grammars of Dutch and Frisian. Currently, she is Vice-Chair of the Innovative e-dictionaries Working Group in the European Network of e-Lexicography COST action.



**Bob Boelhouwer** (PhD) (M) is a computational linguist. He holds a PhD in Cognitive Science entitled From letter strings to phonemes: the role of orthographic context in phonological recoding from Nijmegen University. At INT, he is responsible for the repository of digital language resources for the Dutch language previously hosted at the 'TST-Centrale' (Agency for Human Language Technology). Currently, he is Vice-Chair of the Integrated Interface to European Dictionary Content Working Group in the European Network of e-Lexicography COST action.

**Katrien Depuydt** (F) is senior researcher at INT. She has worked on two major historical dictionaries and has led several projects on corpus building, computational lexica, tools for lexicon building and retrieval systems for corpora and dictionaries. She has been involved in several European (IMPACT, SUCCEED, TRANSCRIPTORIUM) and national projects, leading work packages on lexicon development and corpus processing.

**Jesse de Does** (PhD) (M) is a computational linguist. He holds a PhD in applied mathematics and a master's degree in Slavic Linguistics, and has many years of experience in language processing and retrieval applications both for historical and modern lexicographic data. He is an expert in TEI.

**Jan Niestadt** (M) is software development team leader at INT. He has many years of experience developing high-quality, (re)usable software for lexicography, including the INL-DWS. He is also the creator and main developer of Lucene-based corpus search system BlackLab. His research interest lies in intelligent processing and searching in large data collections.

**Tanneke Schoonheim** (PhD) (F) is a linguist and a lexicographer. After studying Dutch Language and Literature at the University of Leiden, she obtained a PhD in Arts with a dissertation on historical onomastics. She has been working at INT since 1986, first as a lexicographer of the Early Middle Dutch dictionary, later as chief-editor of the Old Dutch dictionary and the Etymological dictionary of Dutch. Now she is chief-editor of the Algemeen Nederlands Woordenboek. Besides her work on various lexicographic dictionaries, she has published widely on Old Dutch, Early Middle Dutch, lexicography and onomastics. She is a member of the EURALEX board as well as STSM manager of the European Network of e-Lexicography COST action.

**All the named persons are employed by the beneficiary.**

#### Relevant publications

- Boelhouwer, B., Sijens, H. & Dykstra, A. (Forthcoming). Dictionary portals. In P. A. Fuertes-Olivera (Ed.), *The Routledge Handbook of Lexicography*.
- de Does, J., Niestadt, J. & Depuydt, K. (Forthcoming). Creating research environments with BlackLab. In J. Odijk & A. van Hessen (Eds.), *CLARIN in the Low countries: a Research Infrastructure for the Humanities*.
- Klosa, A. & Tiberius, C. (2016). Der lexikografische Prozess. In A. Klosa & C. Müller-Spitzer (Eds.), *Internetlexikografie. Ein Kompendium*. De Gruyter, pp. 65-110.
- Schoonheim, T & Tempelaars, R. (2010), Dutch Lexicography in Progress, The Algemeen Nederlands Woordenboek (ANW). In: A. Dykstra & T. Schoonheim (eds.), *Proceedings of the XIV Euralex International Congress. Ljouwert, Fryske Akademy/Afûk*.
- Tiberius, C., Niestadt, J. & Schoonheim, T. (2014). The INL Dictionary Writing System. In I. Kosem & M. Rundell (Eds.), *Slovenščina 2.0: Lexicography*, 2(2), 72–93.

#### Previous projects/activities

**IMPACT** (<http://www.impact-project.eu/>) EU-FP7. This project has significantly improved access to historical text by removing the barriers that stand in the way of the mass digitisation of the European cultural heritage. It ran from 2008-2012.

**SUCCEED** (<http://www.succeed-project.eu/>) EU-FP7-ICT. The Support Action Centre of Competence in Digitisation (Succeed) promoted the take up and validation of research results in mass digitisation, with a focus on textual content. It ran from 2013-2014.

**CLARIN-NL** (2009-2013). Within CLARIN, the institute was involved in various subprojects, i.e. TICCLops, DUELME-LMF, GTB-WFT, VU-DNC and NAMESCAPE.

**European Network of e-Lexicography** (ENeL COST Action IS1305). INT is grant holder of the ENeL COST Action. The action finishes in October 2017.

**Historical Dutch Dictionaries online:** The following historical dictionaries of Dutch are available online through the Integrated Language Database at <http://gtb.inl.nl>:

- *Dictionary of Old Dutch* ([ONW](#)) 500-1200
- *Dictionary of Early Middle Dutch* ([VMNW](#)) 1200-1300
- *Dictionary of Middle Dutch* ([MNW](#)) 1250-1550
- *Dictionary of the Dutch Language* ([WNT](#)) 1500-1976
- *Dictionary of Frisian Language* ([WFT](#)) 1800-1975

Together they describe the Dutch vocabulary from ca. 500 to 1976 and the Frisian vocabulary from 1800 to 1975.

#### Description of infrastructure

INT has an official role as a repository for digital language resources for the Dutch language previously hosted at the 'TST-Centrale' (Agency for Human Language Technology). As such it has ample experience in acquisition, management, and maintenance of linguistic resources and tools, and making these available. This work is supported by a service desk.

INT is a CLARIN center in the Netherlands with a specific focus on resources that are relevant to the lexicological study of the Dutch language and on resources relevant for research in and development of language and speech technology.

Pre-existing tools and services:

- BlackLab (<http://inl.github.io/BlackLab/>)

Lexicographic data:

- Algemeen Nederlands Woordenboek; Dictionary of Contemporary Dutch
- Woordenboek der Nederlandse Taal
- Oudnederlands Woordenboek
- Vroegmiddelnederlands Woordenboek
- Middelnederlands Woordenboek

#### 4. Sapienza University of Rome, UNIROMA1

Role in the project (cf. Consortium as a whole)

- Work package leader (WG3)
- Partner with computational linguistics data and expertise
- Technology partner (machine learning, data mining, artificial intelligence)

**Description of the legal entity**

**The Sapienza University of Rome** is a seven-century-old university in the heart of Rome. It is one of the largest universities in Europe, with around 150,000 students. Its Faculty of Information Engineering, Informatics and Statistics (that includes the Department of Computer Science) is one of the youngest, most energetic and scientifically active environments at Sapienza. Its Department of Computer Science is a modern, well-equipped research institution with a top-class faculty and a strong Ph.D. program. The Department comprises 44 faculty members (among whom 5 ERC grant holders), 15 postdocs and around 30 Ph.D. students.

All Sapienza researchers participating in this project are members of the Linguistic Computing Laboratory (LCL) of the Computer Science Department. LCL comprises 12 members (1 full and 1 associate professor, 6 Ph.D. students, 4 research fellows) and pursues active, state-of-the-art research in a number of areas of Natural Language Processing (NLP) such as large-scale knowledge acquisition, ontology learning and population, Word Sense Disambiguation and Induction, semantic parsing and semantic Information Extraction. In recent years, LCL freely released many linguistic resources (such as multilingual lexical knowledge bases and annotated corpora), as well as several web applications (including popular terminology and glossary extraction systems) in the domain of language processing. These resources can be freely accessed from the group website at <http://lcl.uniroma1.it>.

LCL has participated in several national and international projects, including MultiJEDI (2011-2016), on multilingual word sense disambiguation, LIDER (2013-2016), on the creation of a Linguistic Linked Data cloud, INTEROP (2003-2007), a NoE on enterprise interoperability, LC3 (2007-2010), a national project on the development of ontology-based storytelling in the cultural heritage domain and CHAT (2006-2009), a national project on e-learning and distance learning multimedia applications. Currently, prof. Navigli is the head of the MOUSSE ERC Starting Grant (2017-2022) on language-independent multilingual semantic representations.

**Short CVs of main team members**

**Prof. Roberto Navigli** (PhD) (M) is an Associate Professor in the Department of Computer Science of the Sapienza University of Rome. He was awarded the Marco Cadoli 2007 Italian National Prize for the best doctoral thesis in Artificial Intelligence. In 2013 he received the Marco Somalvico AI\*IA prize for the best young Italian researcher in AI. His research lies in the field of Natural Language Processing (including multilingual word sense disambiguation and induction, multilingual entity linking, semantic parsing, large-scale knowledge acquisition, gamification, ontology learning from scratch, open information extraction and relation extraction).

He is the recipient of two ERC Grants: a Starting Grant 2010 in computer science on the topic of multilingual joint word sense disambiguation and a Consolidator Grant 2016 on the topic of multilingual representations of sentences. He was also a co-PI of a Google Focused Research Award on Natural Language Understanding and a partner of the LIDER EU project.

He has served as an area chair of ACL, WWW, and \*SEM, and a senior program committee member of IJCAI. Currently he is an Associate Editor of the Artificial Intelligence Journal, a member of the editorial board of the Journal of Natural Language Engineering, a guest editor of the Journal of Web Semantics, and a former editorial board member of Computational Linguistics. He was invited as speaker at

popular conferences such as RANLP, SOFSEM and CICLING and workshops on NLP- and Semantic Web-related topics.

**Prof. Paola Velardi** (PhD) (F) received her Degree in Electrical Engineering from the Sapienza University of Rome in 1978. Since 2001 she is full professor. Her main interests are in the areas of natural language processing, machine learning, ontology learning and the semantic web. She participated as partner, responsible or consultant in many international EC projects and national projects. She published more than 120 publications on international, high-rated, journals and conferences. She is reviewer of the main conferences in the area of language processing and artificial intelligence (ACL, EACL, IJCAI etc.) and IEEE journals.

**Claudio Delli Bovi** (M) is a PhD student in the Department of Computer Science of the Sapienza University of Rome, and a member of the Linguistic Computing Laboratory research group, under the supervision of Prof. Roberto Navigli. He obtained a MSc degree (cum laude) in Computer Science Engineering from the Sapienza University of Rome, where he specialized in Artificial Intelligence and Natural Language Processing. His research interests lie in the areas of Semantic Information Extraction, syntactic-semantic modeling for Sense Disambiguation and semantically-enriched Machine Translation.

**Alessandro Raganato** (M) is a PhD student in the Department of Computer Science of the Sapienza University of Rome, and a member of the LCL research group, under the supervision of Prof. Roberto Navigli. He obtained a MSc degree (cum laude) in Computer Science from the Sapienza University of Rome, where he specialized in Artificial Intelligence. His research interests lie in the field of multilingual Natural Language Processing and include Part-Of-Speech Tagging, Morphological Analysis, Word Sense Disambiguation, Named Entity Recognition and Entity Linking.

**All the named persons are employed by the beneficiary or have an equivalent relationship covering the eligibility of costs (Claudio Delli Bovi, Alessandro Raganato).**

#### Relevant publications

- Camacho-Collados, J., Pilehvar, M. T. & Navigli, R. (2016). Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities. *Artificial Intelligence*, 240, 36-64. NASARI multilingual semantic vector representations available at <http://lcl.uniroma1.it/nasari>
- Flati, T., Vannella, D., Pasini, T. & Navigli, R. (2016). MultiWiBi: The multilingual Wikipedia bitaxonomy project. *Artificial Intelligence*, 241, 66-102. The multilingual Wikipedia bitaxonomy is available at <http://wibitaxonomy.org>
- Navigli, R. & Ponzetto, S. (2012). BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. *Artificial Intelligence*, 193: 217-250. This paper led to the creation of the largest multilingual encyclopedic dictionary and semantic network, namely BabelNet ([babelnet.org](http://babelnet.org)).
- Moro, A., Raganato, A., Navigli R. (2014). Entity Linking meets Word Sense Disambiguation: a Unified Approach. *Transactions of the Association for Computational Linguistics (TACL)*, 2, 231-244. This paper led to the creation of Babelfy, a word sense disambiguation and entity linking system which works in 271 languages ([babelify.org](http://babelify.org)).
- Jurgens, D. & Navigli, R. (2014). It's All Fun and Games until Someone Annotates: Video Games with a Purpose for Linguistic Annotation. *Transactions of the Association for Computational Linguistics (TACL)*, 2, 449-464.

#### Previous projects/activities

<p><b>LIDER</b> (<a href="http://lider-project.eu">http://lider-project.eu</a>), EU-FP7: Development of best practices and a reference architecture for the publication of linguistic linked data.</p> <p><b>MultiJEDI</b> (<a href="http://multijedi.org">http://multijedi.org</a>), ERC Starting Grant: multilingual Word Sense Disambiguation, automatic creation of a multilingual encyclopedic dictionary (BabelNet) and creation of multilingual sense representations like NASARI.</p>
<p><b>Description of infrastructure</b></p>
<p>Pre-existing tools and services:</p> <ul style="list-style-type: none"> <li>• <b>babelfy</b> (<a href="http://babelfy.org/">http://babelfy.org/</a>)</li> <li>• <b>MultiWiBi</b> (<a href="http://wibitaxonomy.org/">http://wibitaxonomy.org/</a>)</li> </ul>

**5. National University of Ireland Galway (NUIG)**

<p><b>Role in the project (cf. Consortium as a whole)</b></p>
<ul style="list-style-type: none"> <li>• Work package leader (WG2)</li> <li>• Partner with computational linguistics data and expertise</li> <li>• Partner with expertise in standardisation</li> <li>• Technology partner (machine learning, data mining, artificial intelligence)</li> </ul>
<p><b>Description of the legal entity</b></p>
<p><b>The National University of Ireland, Galway</b> will participate in ELEXIS through two research institutes: The Insight Centre for Data Analytics and the Moore Institute.</p> <p>The <b>Insight Centre</b> at NUIG was founded in 2003 as DERI (Digital Enterprise Research Institute) and incorporated into the nationwide Insight Centre for Data Analytics (<a href="http://insight-centre.org/">http://insight-centre.org/</a>) in July 2013. Its most recent peer review, conducted by the main funder Science Foundation Ireland (SFI), concluded that the science and technology application of the institute were “impressive” and “world class”. Today, Insight at NUIG hosts in excess of 100 members and has established itself as a top player worldwide in the areas of Semantic Web and Linked Data. The institute has acquired direct research awards in excess of €60 million from SFI, Enterprise Ireland (EI), and EU framework programmes. It has developed and successfully implemented a research strategy around the goal of “Enabling Networked Knowledge”, which aims at capitalizing on knowledge as the fuel for the digital service economy, by linking information and exploiting the resulting knowledge graphs as the basis for economic productivity. The institute performs fundamental and applied research in a range of research areas to enable this, including data streams and sensor networks, knowledge discovery, natural language processing, social semantics and social network analysis, among others. se cases have been implemented with more than 100 industry and public partners in collaborative projects and the centre has been very successful in acquiring EU funding. To date the institute has participated in 8 IPs, 17 STREPs, 4 NOEs, 2 CA, 2 CIP PSPs, 1 MC, 2 COST Actions, and 4 other EC funding programmes, with total funding of €17.58 million since 2003. Additionally, the institute has been successful in attracting over €42.42M from National funding sources, including from EI, SFI, IRCSET and direct industrial collaborations.</p>

**The Moore Institute** is at the forefront of inquiry into humanities, cultures and society and is the co-lead of DARIAH Ireland. The institute was founded in 2000 and has earned major national and international grants and led projects across the disciplines in areas from archaeology to ancient history and medieval studies, and from the early modern period to the present day, including politics, gender, and performance.

### Short CVs of main team members

**John P. McCrae** (PhD) (M) is a Lecturer Above-the-Bar at the National University of Ireland Galway. He completed his PhD at the National Institute of Informatics, Tokyo, while contributing to the BioCaster system for detecting disease outbreaks by reading texts in East Asian languages. From 2009-2015 he was at Bielefeld University, where he played a leading role in the development of the lemon (Lexicon Model for Ontologies), a major contribution to the representation of semantics relative to natural language, which is now being used by most relevant research groups, notably in recent efforts by the Global WordNet Association to standardize an interlingual index of concepts. Secondly, out of the work on this topic he has been instrumental in creating the topic of linguistic linked open data as a major research theme, which has been supported by over a dozen workshops and events and was a major theme of the previous Language Resource and Evaluation Conference (LREC). This topic led to the LIDER project, which used linguistic linked open data as an enabler for content analytics in enterprise. Since August 2015, he has been working at the Insight Centre for Data Analytics, where his work has focussed on linguistic linked data, reproducible research, ontologies and NLP for minority languages.

**Paul Buitelaar** (PhD) (M) is Senior Lecturer at the National University of Ireland Galway (NUIG), vice-director of the Insight Centre for Data Analytics at NUIG and head of the Insight Unit for Natural Language Processing. His main research interests are in the development and use of Natural Language Processing methods and solutions for semantic-based information access. He has been involved in a large number of national and international funded projects in this area, including the coordination of the Monnet FP7 project on cross-lingual ontology-based information access with use cases in public and financial services. In recent years, his team developed the Saffron framework for knowledge extraction, which has been used in industry applications with companies such as Irish Times, Kennys and HP. Another line of recent research has been on the definition and implementation of lemon, a vocabulary for Linguistic Linked Data, which was developed in the context of the EU funded project Monnet and has been applied in EU projects such as LIDER, EuroSentiment and MixedEmotions among others.

**Mihael Arcan** (PhD) (M) is a post-doctoral researcher in the Unit for Natural Language Processing at the Insight Centre for Data Analytics, where his main research topic focuses on term translation with statistical machine translation (SMT). He studied German Language at the University of Ljubljana, Slovenia, where he worked on his Diploma thesis "Named Entity Recognition for German and Slovene" under the supervision of Stojan Bračič and Špela Vintar. In 2009 he obtained a Masters degree in Computational linguistics at the Ruhr University in Bochum, Germany on extraction of semantic relations from the Slovene national corpus. After this he worked for Lionbridge as a developer of language technologies for Slovene, and for the Slovenian Project "Communication in Slovene". At Insight@NUI Galway he worked in the MONNET (Multilingual Ontologies for Networked Knowledge) project, a European Project with the goal to provide standards and technology to facilitate multilingual access to Semantic Web knowledge resources. During the EUROSENTIMENT project and currently at the MixedEmotions project he is focusing on the generation of multilingual variants of automatically generated lexicons.

**Justin Tonra** (PhD) (M) is Lecturer in English (Digital Humanities) at the National University of Ireland Galway, where he teaches digital humanities, book history, textual studies, and literary studies in a number of historical periods and genres. He is Humanities Director of the Digital Arts and Humanities



PhD Programme and NUI Galway, Co-National Coordinator of DARIAH-IE, and a Member of the Executive Committee of the Moore Institute for Research in the Humanities and Social Studies. He has worked on a number of prominent digital humanities projects: the multi award-winning [Transcribe Bentham](#), a project to crowdsource transcriptions of the manuscripts of Jeremy Bentham at University College London; a digital research environment dedicated to the work of James Macpherson, [Ossian Online](#); and [Personæ: A Character-Visualisation Tool for Dramatic Texts](#). He has completed periods of postdoctoral work at the University of Virginia and NUI Galway, and his research has been published in *Literary and Linguistic Computing*, *European Romantic Review*, and *Advances in Complex Systems*.

**All the named persons are employed by the beneficiary.**

**Relevant publications**

- McCrae, J. P., Arcan, M., Asooja, K., Gracia, J., Buitelaar, P. & Cimiano, P. (2016). Domain adaptation for ontology localization. *Web Semantics*, 36, 23-31.
- McCrae, J. P. & Cimiano, P. (2015). Linghub: a Linked Data based portal supporting the discovery of language resources. *Proceedings of the 11th International Conference on Semantic Systems, SEMANTICS 2015*, Vienna, Austria, September 15 - 17, 2015, pp. 88-91.
- McCrae, J. P., Cimiano, P., Rodriguez-Doncel, V., Vila-Suero, D., Gracia, J., Matteis, L., Navigli, R., Abele, A., Vulcu, G. & Buitelaar, P. (2015). Reconciling Heterogeneous Descriptions of Language Resources. *Proceedings of the 4th Workshop on Linked Data in Linguistics*, pp. 39-48.
- Arcan, M., McCrae, J. P. & Buitelaar, P. (2016). Expanding wordnets to new languages with multilingual sense disambiguation. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, Osaka, Japan, December 11-17 2016, pp. 97-108.
- McCrae, J. P., Aguado-de-Cea, G., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr D. & Wunner, T. (2012). Interchanging lexical resources on the Semantic Web. *Language Resources and Evaluation*, 46(6), 701-709.

**Previous projects/activities**

**LIDER** (<http://lider-project.eu/>), EU-FP7: Development of best practices and a reference architecture for the publication of linguistic linked data.

**MONNET** (Multilingual Ontologies for Networked Knowledge), EU-FP7: Development of a semantics-based solution for integrated information access across language barriers.

**MixedEmotions** (<http://mixedemotions-project.eu>) EU-FP7: This project is developing an innovative multilingual multi-modal Big Data analytics applications that will analyze a more complete emotional profile of user behavior using data from mixed input channels.

**Saffron** (<http://saffron.insight-centre.org/>): The Saffron project is a research framework for extracting knowledge from texts. It has been used to develop custom portals for Kenny's Bookstores and the Irish Times.

**YourDataStories** (<http://yourdatastories.eu/>) EU-FP7: Development of a software stack that will enable the Open Government Data (OGD) to reach citizens in their everyday online life.

**Description of infrastructure**

Pre-existing tools and services:

- NAISC (<https://github.com/jmccrae/naisc>)
- Yuzu (<https://github.com/jmccrae/yuzu>, <http://linghub.org>)

## 6. Austrian Academy of Sciences (OeAW)

### Role in the project (cf. Consortium as a whole)

- Work package leader (WG7)
- Partner with lexicographic data and expertise
- Partner with computational linguistics data and expertise
- Partner with expertise in standardisation
- Digital humanities partner (e-learning, digital editions)

### Description of the legal entity

**The Austrian Academy of Sciences (OeAW)** is Austria's largest non-university research facility with over 1300 employees looking back on a history of over 170 years. The 28 institutes perform applied and basic research in many different disciplines ranging from archaeology, cultural sciences, linguistics, and literary studies to quantum physics and molecular biology at the highest international level.

Several institutes of the Academy look back on a long tradition in the application of digital methods. Over the past few years, the Academy has developed a special focus on digital methods in the humanities. In 2015, these efforts culminated in the establishment of the Austrian Centre for Digital Humanities (ACDH-OeAW) which was jointly founded by the Academy and the Ministry of Science, Research and Economy with the declared intention of fostering the digital paradigm in the humanities.

In their research, the ACDH-OeAW team works on text- and language-related questions focusing on non-standard and historical linguistic varieties as well as eLexicography and scholarly digital editions. In numerous projects across a wide range of digitally mediated humanities disciplines, the ACDH-OeAW works on digital language resources and related standards, semantic technologies, domain-specific virtual research environments and repositories for preservation of valuable digital research data. As the Austrian coordinator for both ESFRI infrastructure consortia in the humanities (CLARIN and DARIAH), the Austrian Academy of Sciences has been playing a key role in the Austrian DH landscape, acting as a hub for many DH-related activities on the institutional, national and international. These activities also includes to raise awareness of the benefits of innovative approaches and to encourage researchers to make use of digital technologies with the organization of training events and conferences as well as a tailored outreach strategies.

### Short CVs of main team members

**Karlheinz Moerth** (PhD) (M) is the director of the ACDH-OeAW. Proceeding from a background in Near Eastern Studies (with a focus on modern languages and applied linguistics), he has been working throughout his academic life at the interface between modern IT and humanities disciplines. He conducted research on a wide range of text technological questions taking a special interest in digital editing, eLexicography, text lexicography, methodologies for the build-up and maintenance of digital corpora, annotation research and corpus-related encoding.

Beginning in lexicographic projects of the Commission for Functional Literary Text Types, he contributed substantially to the creation of the AAC - Austrian Academy Corpus which evolved over the



years as Austria's major digital language resource covering large parts of the 19th and 20th centuries. One of his areas of interest are standards in the digital humanities, in particular standards related to digital language resources, and the interface between digital dictionaries and digital corpora. Being a member of Austrian Standards, he worked among others on issues related to the de-facto standard TEI (Text Encoding Initiative), focusing on dictionary module of the TEI Guidelines and adapting the TEI for linguistic applications, and LMF (Lexicographical Markup Framework). In recent years, he was participating in the TUNICO project (Linguistic dynamics in the Greater Tunis Area: A corpus-based approach) which investigated a contemporary variety of spoken Arabic making use of innovative approaches combining both the analysis of corpus and lexicographic data. The project, which ended in 2016, produced two digital language resources: a corpus of spoken Tunisian youth language and the first freely available digital dictionary of a spoken Arabic variety.

Particularly worth mentioning is his engagement for the European infrastructure consortia CLARIN (Common Language Resources and Technology Infrastructure) and DARIAH (Digital Research Infrastructures for the Arts and Humanities) where he has been active in a number of bodies. On the Austrian level, he has been coordinating (on behalf of the Ministry of Science, Research and Economy) all CLARIN and DARIAH activities since January 2014. On the European level, he has been functioning as co-head of the DARIAH-EU Virtual Competency Centre I (e-Infrastructures) from 2011 until 2014 and he is part of the CLARIN standards committee.

**Tanja Wissik** (PhD) (F) is a senior researcher at the Austrian Academy of Sciences and a lecturer at the University of Graz. She holds a PhD from the University of Vienna in translation studies with a specialisation in the field of terminology and corpus linguistics. She has been working in numerous national and international research projects first as a junior researcher at the Institute for Specialised Communication and Multilingualism of the European Academy Bolzano and then as a researcher and lecturer at the University of Vienna. She started to work in dictionary and terminology projects (German – Italian) at the European Academy Bolzano, then she contributed at the University of Vienna to language resources projects like CLARIN preparatory phase, FLaReNet (Fostering Language Resources Network), CLARIN ERIC and LISE (Legal Language Interoperability Services) project, where she was deputy coordinator and WP leader of the dissemination workpackage. At the ACDH-OeAW she is involved in both research infrastructures for the humanities, CLARIN ERIC, especially as part of the user involvement working group, and in DARIAH-EU as part of the Digital Methods and Practices Working Group. She is involved in #dariahTeach, an Erasmus + project that aims at developing an open platform for open e-learning material in the Digital Humanities. At national level she is involved in CLARIAH-AT, the national CLARIN and DARIAH consortium in Austria, as well as in the Digital Humanities Austria initiative, a network to promote digital humanities research and foster exchange between researchers in Austria and abroad. Furthermore she is involved in the network and outreach activities of the ACDH with experience in scientific dissemination, communication strategies, social media as well as conference and workshop organization (e.g. FLaReNet Forum 2009, LSP2013, TEI2016, dha2015, dha2016, ACDH Tool Galleries).

**Thierry Declerck** (M) studied Philosophy at the University of Brussels and Computational Linguistics at the University of Tübingen. He has been working in numerous international and international research projects at the DFKI Language Technology Lab, the Computational Linguistic Department of the University of the Saarland, Institut für Maschinelle Sprachverarbeitung (IMS) of the University of Stuttgart and the Austrian Academy of Science. With his expertise in computational linguistics, natural language processing and linked open data he is ensuring the tight connection between language technology and semantic web technologies in the Digital Humanities, with a special focus on digital editions and lexicography projects. He was contributing to the LIDER project (<http://www.lider-project.eu/>), dealing with the establishment of the Linguistic Linked Open Data (LLOD) cloud, and to the FREME projects (<http://www.freme-project.eu/>), dealing with application scenarios for the LLOD. Before this he was the coordinator of the European FP7 project TrendMiner (<http://www.trendminer->

[project.eu/](http://project.eu/)) and he was involved in CLARIN, D-SPIN and the Monnet project for Multilingual Ontologies for Networked Knowledge. He is involved in the ENeL (European Network for e-Lexicography) COST action within Working Group 3 “Innovative e-dictionaries. As a consultant at the Austrian Academy of Sciences he has been instrumental in introducing and cooperating on topics related to the Linguistic Linked Open Data and other standards developed in the context of W3C and he has contributed to projects like Texttechnologische Methoden zur Analyse österreichischer Barockliteratur, funded by the Anniversary Fund of the Austrian National Bank or the Dictionary of Bavarian Dialects in Austria (WBÖ).

Furthermore he was an active member of ISO TC37/SC4 and DIN NAAT6. He was for example editor of the ISO standard for syntactic annotation (SynAF), which has been submitted as a final draft for international standard in June 2010. More recently he is involved in W3C community groups, like Ontolex and he is co-chair for the LREC conferences, since 2012.

**All the named persons are employed by the beneficiary or will be employed after project starts (Thierry Declerck) or have an equivalent relationship covering the eligibility of costs.**

#### Relevant publications

- Mörth, K. (forthcoming). Arabic lexicography in the Internet era. In: P. A. Fuertes-Olivera (Ed.), *The Routledge Handbook of Lexicography*.
- Declerck, T. & Mörth, K. (2016). Towards a Sense-based Access to Related Online Lexical Resources. In: T. Margalitadze & G. Meladze (Eds.): *Proceedings of the XVII EURALEX International Congress*. Tbilisi, Georgia, pp. 660-668.
- Declerck, T., Wandl-Vogt, E. & Mörth, K. (2015). Towards a Pan European Lexicography by Means of Linked (Open) Data. In Kosem, I., Jakubiček, M., Kallas, J., Krek, S. (eds.) *Electronic lexicography in the 21st century: linking lexical data in the digital age. Proceedings of the eLex 2015 conference, 11-13 August 2015, Herstmonceux Castle, United Kingdom*. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd., pp. 342-355.
- Wissik, T. (2014). Nutzung von Korpora bei Untersuchungen an der Schnittstelle zwischen Terminologiewissenschaft und Varietätenlinguistik: methodologische Überlegungen. In: Drewer, P./Mayer F./Schmitz, K. (Hrsg.): *Rechte, Rendite, Ressourcen. Wirtschaftliche Aspekte des Terminologiemanagements. Akten des DTT-Symposiums 27.-29. März 2013, Mannheim*. Köln: Deutscher Terminologie-Tag e.V.
- Wissik, T. & Resch, C. (2016). Digitale Tools und Methoden für die geisteswissenschaftliche Forschung praxisnah erklärt: Ein neues Format im Test. *Digital Humanities 2016: Conference Abstracts*. Jagiellonian University & Pedagogical University, Kraków, pp. 711-713.

#### Previous projects/activities

**Tunico** (Language Dynamics in the Greater Tunis Area: A corpus-based approach), FWF (Austrian Science Fund) project (<https://tunico.acdh.oeaw.ac.at>). The project investigated a contemporary variety of spoken Arabic making use of innovative approaches combining both the analysis of corpus and lexical data.

**LISE** (Legal Language Interoperability Services), CIP-ICT-PSP (<http://www.lise-termservices.eu/>). The project addressed the urgent need for consolidated administrative nomenclatures and legal terminologies as tools to enhance interoperability and cross border collaboration. The project elaborated best practices and guidelines for legal and administrative terminology work as well as tools to enhance interoperability.

**Texttechnologische Methoden zur Analyse österreichischer Barockliteratur**, funded by the Anniversary Fund of the Austrian National Bank. One of the outcomes of the project is the ABaC:us – Austrian Baroque Corpus (<https://acdh.oeaw.ac.at/abacus>).

**#dariahTeach** Erasmus+ Strategic Partnership for Higher Education (<http://dariah.eu/teach/>). The goal of the project is to develop an open source, extensible, online multilingual, community-driven platform for high quality and open teaching and training materials for the digital humanities specifically tailored for third-level education and the training materials themselves.

**Parthenos** (Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies), H2020 (<http://www.parthenos-project.eu/>). The ACDH is particularly involved in Work Packages 5 (Interoperability and semantics) & 6 (Services and tools), leading the task Resource Discovery Tools (WP6). In addition, the ACDH contributes to Work packages 2 (Community involvement and requirements), 3 (Common policies and implementation strategies), and 4 (Standardization) as well as to dissemination activities (WP8).

### Description of infrastructure

Lexicographic data:

- Dictionary of Bavarian Dialects of Austria
- Dagaare-Cantonese-English Dictionary
- Hausa-English Dictionary
- Database of Bavarian Dialects of Austria
- Russian Dialect Dictionary
- Tunico (Tunisian)
- Viennese Historical Dictionaries Online (ViDi)

## 7. Centar za digitalne humanističke nauke (Belgrade Center for Digital Humanities, BCDH)

### Role in the project (cf. Consortium as a whole)

- Work package leader (WG5)
- Partner with lexicographic data and expertise
- Digital humanities partner (e-learning, digital editions)

### Description of the legal entity

**The Belgrade Center for Digital Humanities** (BCDH) is a leading Serbian institution exploring the use of computational methods in the study of traditional humanities disciplines. With a strong focus on data modelling, digital editions, standards-compliant lexicographic resources and the development of training materials, BCDH is forging an infrastructural approach to humanities computing with the goal of making historical, literary and cultural heritage as well as contemporary artistic production accessible, reusable, and open to creative computational interventions.

BCDH is closely cooperating with the Institute for Serbian Language of the Serbian Academy of Arts and Sciences on a nationally-funded project digitizing Serbian lexicographic heritage and building a dictionary platform along with a set of tools and APIs for the scholarly exploration of lexical data.

As the National Coordinating Institution for the Digital Research Infrastructure in the Arts and Humanities in the Republic of Serbia (DARIAH-RS), the BCDH coordinates and directs a multi-institutional effort to build a digital research infrastructure for the arts, humanities and cultural heritage in Serbia. At the same time, as the institution chairing the National Coordinators Committee at DARIAH-EU, BCDH is also closely involved in the process of coordinating national roadmaps and infrastructural activities in 17 DARIAH member countries.

BCDH has played vital roles in various European initiatives and projects, including the European Network for e-Lexicography (ENeL), where it co-leads the Work Package "Retrodigitized Dictionaries". As work package leader in both #dariahTeach — an Erasmus+ funded Strategic Partnership for Higher Education, and DESIR — a project funded by a Horizont 2020 Infradev-3 grant, BCDH is already spearheading the effort in developing open-source, multilingual training materials and reference curricula in the area of Digital Humanities.

With experience in infrastructure building, scholarly digitization of lexicographic resources, as well as DH pedagogy, BCDH is uniquely positioned to contribute to ELEXIS, where it will lead WP5 "Training and Education" and actively participate in WP1 "Lexicographic Data and Workflow."

### Short CVs of main team members

**Toma Tasovac** (M) is director of the Belgrade Center for Digital Humanities. With degrees from Harvard and Princeton, Mr. Tasovac has more than 15 years of professional experience in the field of digital humanities. He is a recognized scholar in the field of electronic lexicography, with a particular focus on retrodigitization of legacy dictionaries, as well as a champion of humanistic infrastructure building on the European scale. He serves as the National Coordinator of DARIAH-RS; chair of the DARIAH-EU National Coordinators Committee, member of the Europeana Research Advisory Board; member of the CLARIN-DE/DARIAH-DE Technical Board, co-Leader of the DARIAH-EU Working Groups "Training and Education" and "Lexical Resources" etc. He also serves on the European Research Council Consolidator Grant Panel "Culture and Cultural Production." He is the leader of Working Group "Technology-oriented specifications for a flexible and robust solution" in COST Action CA16105: European Network for Combining Language Learning with Crowdsourcing Techniques (enetCollect) (2017-2020) and co-leader of the Working Group "Retro-digitizing dictionaries" in COST Action IS1305: European Network for eLexicography (ENeL) (2013-2017).

**All the named persons are employed by the beneficiary.**

### Relevant publications

- Tasovac, T., Rudan, S. & Rudan, S. (2015). Developing Morpho-SLaWS: An API for the Morphosyntactic Annotation of the Serbian Language. *Systems and Frameworks for Computational Morphology*. Proceedings of the 4th International Workshop, SFCM 2015, Stuttgart, Germany, September 17-18, 2015, pp. 137-47.
- Tasovac, T. & Petrović, S. (2015). Multiple Access Paths for Digital Collections of Lexicographic Paper Slips. In I. Kosem, M. Jakubiček, J. Kallas & S. Krek (Eds.), *Electronic lexicography in the 21st century: linking lexical data in the digital age. Proceedings of the eLex 2015 conference, 11-13 August 2015, Herstmonceux Castle, United Kingdom*. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd., pp. 384-396.
- Ermolaev, N. & Tasovac, T. (2012). Building a Lexicographic Infrastructure for Serbian Digital Libraries. *Libraries in the Digital Age (LIDA) Proceedings, 12*: Available online under <http://ozk.unizd.hr/proceedings/index.php/lida2012/article/view/55/27>.
- Tasovac, T. & Ermolaev, N. (2011). Encoding Diachrony: Digital Editions of Serbian 18th-Century

<p>Texts. <i>Springer Lecture Notes in Computer Science</i>, vol. 6966, 497–500.</p> <ul style="list-style-type: none"> <li>• Tasovac, T. (2012). Potentials and challenges of WordNet-based pedagogical lexicography: The Transpoetika Dictionary. In S. Granger &amp; M. Paquot (Eds.), <i>Electronic Lexicography</i>. Oxford: Oxford University Press, pp. 237-258.</li> </ul>
<p><b>Previous projects/activities</b></p>
<p><b>Horizont 2020 INFRADEV-3</b> grant: “DARIAH ERIC Sustainability Refined” (2017-2019). Leader of Work Package “Training and Education”.</p> <p><b>Erasmus+ Strategic Partnership or Higher Education (2015-2017)</b>: #dariahTeach. Leader of Work Package: “Curriculum Development”.</p> <p><b>Raskovnik</b>: Serbian Dictionary Portal and Platform for Comparative Lexical Studies. In cooperation with the Institute of Serbian Language of the Serbian Academy of Arts and Sciences. Funded by the Ministry of Culture and Information of the Republic of Serbia.</p>
<p><b>Description of infrastructure</b></p>
<p>Lexicographic data:</p> <ul style="list-style-type: none"> <li>• Karadžić, Serbian Dictionary (1818, 1852)</li> <li>• Miklošič, Lexicon Palaeoslovenico-Graeco-Latinum (1862—1865)</li> <li>• Daničić, Dictionary of Serbian Literary Antiquity (1863-4)</li> <li>• Bojanić &amp; Trivunac, Dictionary of Dubrovnik Dialect</li> <li>• Elezović, Dictionary of Kosovo-Metohija Dialect</li> <li>• Zlatanović, Dictionary of Southern Serbian Dialects</li> <li>• Žugić, Dictionary of Jablanica Region</li> </ul>

## 8. RILMTA Research Institute for Linguistics Hungarian Academy of Sciences

<p><b>Role in the project (cf. Consortium as a whole)</b></p>
<ul style="list-style-type: none"> <li>• Partner with lexicographic data and expertise</li> <li>• Partner with computational linguistics data and expertise</li> </ul>
<p><b>Description of the legal entity</b></p>
<p>Founded in 1949, Research Institute for Linguistics at the Hungarian Academy of Sciences is internationally recognized as the leading research centre for the Hungarian Language pursuing cutting edge research in a number of areas such as theoretical linguistics, synchronic and diachronic study of the structure and use of Hungarian. Its Finno-Ugric Department is actively engaged in projects concerned with the documentation and digital vitality of endangered languages. Its Research Centre for Multilingualism is part of the Mercator Network of Language Diversity Centers and has recently completed a major project on the standardization of Hungarian sign language. Apart from its theoretical research, the Institute is actively engaged in experimental research in neurolinguistics, empirical study of language production and language processing and aphasiac studies. The Lexicographic Department is currently working on the Unabridged Academic Dictionary of Hungarian, based on a historic corpus of Hungarian as well as a huge collection of manual records. With a staff of</p>

over a hundred well-qualified experts, the Institute not only has recognized expertise in the above areas but has developed a large set of textual and lexical resources, which are made available as webservice through the website of the Institute (<http://www.nytud.hu/adatb/index.html>).

The Institute is an active member of EFNIL, the European Federation of National Institutions for Language ([www.efnil.org](http://www.efnil.org)), it runs its Secretariat, with Tamás Váradi acting as its Secretary General.

The Department of Language Technology and Applied Linguistics is one of the leading Hungarian Language Technology centres. The Language Technology Research Group has coordinated several European projects such as iTranslate4.eu and CESAR projects and has taken part in a number of EU or national projects. The Department acts as the Coordinator of the Hungarian Language and Speech Technology Platform (<http://hlt-platform.hu>), it was one of the founding partner of the CLARIN infrastructure ([www.clarin.eu](http://www.clarin.eu)) and acts as the coordinator of the Hungarian CLARIN consortium. The Department was the lead of a national consortium to design an open Hungarian national language processing infrastructure.

The Department is well embedded in several collaborative network efforts currently underway to develop language technology infrastructure. It acts as the national coordinator for Hungarian in the ELRC project and COST projects including, notably, the European Network of e-Lexicography (COST Action IS1305).

Department of Lexicology and Lexicography's main research area is the compilation of monolingual Hungarian dictionaries. This working team created A magyar nyelv értelmező szótára 1–7. ('Explanatory Dictionary of Hungarian', abbr. ÉrtSz., chief editors: G. Bárczi and L. Országh; Akadémiai Kiadó, Budapest, 1959–1962), and its shortened version, Magyar értelmező kéziszótár ('Concise Hungarian Explanatory Dictionary', abbr. ÉKsz., edited by J. Juhász, I. Szőke, G. O. Nagy, and M. Kovalovszky; Akadémiai Kiadó, Budapest, 1972). The department also works on a new etymological dictionary called Új etimológiai szótár ('New Etymological Dictionary of Hungarian', abbr. ÚESz.).

#### Short CVs of main team members

**Dr. Tamás Váradi** (PhD) (M) is manager for projects and innovation and head of the Department of Language Technology and Applied Linguistics. He graduated in English, Spanish and General Linguistics at Eötvös Loránd University and holds a PhD in English Linguistics from the same university. Early in his career he acquired international recognition through his pioneering research in second language acquisition. His empirical research into the spontaneous speech of second language learners led him into the field of corpus linguistics. He was visiting research fellow at Lancaster University in 1990 and worked with Geoffrey Leech on the parsed LOB corpus. Between 1991 and 1995 we taught at the School of Slavonic and East European Studies at London University. Following his return to Budapest he founded the Department of Corpus Linguistics and created the Hungarian National Corpus. Under his leadership the Department has become a key player in Hungarian language technology, well known across Europe. Dr. Váradi was one of the initiators of the CLARIN research infrastructure initiative and served on its Executive Board in the preparatory phase of the project. He has been leading several European and national funded projects as coordinator. He is General Secretary of the European Federation of National Institutions of Language. Dr. Váradi is a member of the META-NET Executive Board. He serves as chairman of the Hungarian Language and Speech Technology Platform and is the coordinator of the Hungarian CLARIN network. He was the lead of a major national collaborative project to build an open integrated Hungarian Language Technology Infrastructure as well as a research project to support the revitalization of endangered Finno-Ugric languages through language technology.



**Dr. Bálint Sass** (PhD) (M) Research fellow ([http://www.nytud.hu/oszt/korpusz/Sass\\_Balint.html](http://www.nytud.hu/oszt/korpusz/Sass_Balint.html)). Bálint Sass has worked in the field of computational linguistics since 2001. He majored in computer science and then defended his thesis on an automatic lexical acquisition method to create a corpus driven dictionary of verb phrase constructions in 2011. He worked with several corpora created at RILMTA, e.g. the Hungarian National Corpus or the Old Hungarian corpus. His field of expertise is corpus querying and corpus query interfaces. He has been involved in several national and international projects. He is also the author of the new Hungarian contracted Braille script.

**Dr. Veronika Lipp** (PhD) (F) Research fellow. After graduating (2001) she was offered a young research fellow position in the Department of Lexicology and Lexicography at the Research Institute for Linguistics of the Hungarian Academy of Sciences. This department has one huge project, compiling the Comprehensive Dictionary of Hungarian. In the first few years she was responsible for developing and maintaining the electronic historical corpus, which is one of its main sources. Recently she writes entries and at the same time she has editorial tasks also. As one of the two national representatives she has been an active member of the European Network of e-Lexicography (COST Action IS1305) for four years. The latest international project she was involved in is a bibliography of diachronic orthography in Early Modern Europe (project manager: Prof. Marco Condorelli, University of Central Lancashire.) This bibliographic book is likely to be going to be part of the Handbook series attached to the Empirical Approaches to Linguistic Theory series (EALT) (Brill). Beside the lexicographic works she participated in some of the projects of the Institute for Literary Studies of the Hungarian Academy of Sciences, for example to publish and to interpret partly unknown philosophical, historical, linguistical texts from the 18th century (Magyarországi gondolkodók 18. század I-II.) and I took part in the works of Dictionary of Mikes Kelemen also.

**All the named persons are employed by the beneficiary.**

#### Relevant publications

- Váradi, T. & Lipp, V. (Fortcoming). Hungarian Lexicography. In P. Hanks & G. M. de Schryver (Eds.). *International Handbook of Modern Lexis and Lexicography*. Springer.
- Tadić, M. & Váradi, T. (2012). Central and South-East European Resources in META-SHARE. In: *Proceedings of COLING 2012: Demonstration Papers*. The COLING 2012 Organizing Committee, December, 2012, Mumbai, India, pp. 431-438.
- Eszter S., Benyeda, I. Z., Koczka, P. & Ludányi, Z. (2015). [Automatic creation of bilingual dictionaries for Finno-Ugric languages](#). In: *Proceedings of the First International Workshop on Computational Linguistics for Uralic Languages*, Tromsø, pp. 119-131.
- Oravecz C., Tamás, V., Bálint, S. (2014). [The Hungarian Gigaword Corpus](#). In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC2014*, Reykjavik, Iceland. Paris: ELRA, pp. 1719-1723.
- Bálint, S. & Pajzs, J. (2010). FDVC - Creating a Corpus-driven Frequency Dictionary of Verb Phrase Constructions for Hungarian. In: S. Granger & M. Paquot (Eds.), *eLexicography in the 21st century: New challenges, new applications. Proceedings of eLex 2009*, Louvain-la-Neuve, 22-24 October 2009. Cahiers du CENTAL 7. Presses universitaires de Louvain, pp. 263-272.

#### Previous projects/activities

**ELRC** (European Language Resource Coordination; 2015–)

**Finno-Ugric Digital Natives** (2013–): Linguistic support for Finno-Ugric digital communities in generating online content

**EFNILEX** (2008–2015): Automatically generated online prototype dictionaries

**CESAR** (2011–2013) Central and South-East European Resources

### Description of infrastructure

Lexicographic data:

- Hungarian Concise Dictionary

## 9. Institute for Bulgarian Language “Prof. Lyubomir Andreychin” (IBL)

### Role in the project (cf. Consortium as a whole)

- Partner with lexicographic data and expertise
- Partner with computational linguistics data and expertise
- Technology partner (machine learning, data mining, artificial intelligence)

### Description of the legal entity

**The Institute for Bulgarian Language** (IBL) at the Bulgarian Academy of Sciences is a leading research organisation focused on theoretical and applied research of Bulgarian: theoretical, applied and computational linguistics; general and contrastive linguistics; lexicography and terminology; etymology, history of language and dialects, etc.

The Institute for Bulgarian language develops and maintains various linguistic resources – dictionaries (explanatory, bilingual, syntactical, etymological, historical, spelling, phraseological, dialectal, etc.), lexical databases, Bulgarian dialect atlases, lexical semantic networks (Bulgarian wordnet), corpora (monolingual and multilingual, annotated, parallel, incl. the Bulgarian National Corpus), etc.

The Department of Computational Linguistics, founded in 1994 within the Institute, is a leading centre in the field of computational linguistics and natural language processing in Bulgaria. The Department deals with theoretical and applied research in computational linguistics, development of theoretical models for formal description of language, development of language technologies to be implemented in various computer applications and systems, development of language resources.

The Institute for Bulgarian Language is a leading actor in the field and is involved in numerous national and international research projects focused on theoretical issues in the formal description of language, morphological, syntactic and semantic analysis, ontologies and semantic relations, document categorization and annotation, information extraction and retrieval, word sense disambiguation, machine translation, etc.

The Institute for Bulgarian Language has developed a number of widely used software products and language resources, namely: Bulgarian language processing chain and language processing technologies adaptable for different languages; tools for extraction and alignment from multilingual content, for word sense disambiguation and semantic content enhancement; language and translation models, large amount of monolingual and multilingual resources, conceptual resources and language databases.



These products are targeted at various users such as Bulgarian governmental and non-governmental institutions; European institutions; educational bodies – schools, universities, colleges; language professionals; companies working in knowledge management, multilingual data retrieval, multilingual data processing, language data providers, and others.

### Short CVs of main team members

**Prof. Svetla Koeva** (PhD) (F) is a professor in computational linguistics. She is an expert in language processing, formal description of natural language, ontologies, and others. She has extensive management experience as team leader and coordinator of projects at national and international level, such as ATLAS: Applied Technology for Language-Aided CMS; CESAR: Central and South-East European Resources; Internet Monitoring and Multi-language Content Analysis; Bulgarian wordnet, and others.

**Prof. Diana Blagoeva** (PhD) (F) holds a PhD degree in Bulgarian language. She is the editor in chief of the Dictionary of Bulgarian language project. She has been involved in several lexicographical projects including Dictionary of new words in Bulgarian, databases of Bulgarian synonyms, antonyms, new words and phraseology. Prof. Blagoeva is the (co)author of many publications on language resources, Bulgarian lexicology and lexicography, and corpus linguistics.

**Martin Yalamov** (M) is software engineer. For the past four years he has been involved in the development of the several lexicographic databases: of Bulgarian synonyms, antonyms, new words and phraseology, and of the Dictionary of Bulgarian language. He has also an experience in automatic compilation of large scale corpora.

**Dimitar Hristov** (M) is software engineer. His research interests are in the field of NLP. For the past two years he has been involved in the development of components of language processing chain for Bulgarian. He has also an experience in semantic processing, text clustering and automatic domain labelling.

**All the named persons are employed by the beneficiary.**

### Relevant publications

- Koeva, S. (2015). Paraphrasing of Synonyms for a Fine-grained Data Representation. In: *Joint Proceedings of the Posters and Demos Track of 11th International Conference on Semantic Systems - SEMANTICS2015 and 1st Workshop on Data Science: Methods, Technology and Applications (DSci15)*, 1481, CEUR Workshop Proceedings, pp. 79-83.
- Koeva, S. & Blagoeva, D. (2013). The Dictionary Writing System LexIt and its Application in the Bilingual Lexicography. *Cognitive Studies | Études cognitives*, 13, 57-65.
- Koeva, S., Stoyanova, I., Todorova, M., Leseva, S. & Dimitrova, T. (2016). Metadata Extraction, Representation and Management within the Bulgarian National Corpus. In *Proceedings of the 4th Workshop on Challenges in the Management of Large Corpora 2016*, Portorož (LREC-2016 workshop), ELDA, pp. 33-39.
- Stoyanova I., Koeva S., Todorova M. & Leseva S. (2016). Semi-automatic Compilation of a Very Large Multiword Expression Dictionary for Bulgarian. In: *Proceedings of GLOBALEX 2016: Lexicographic Resources for Human Language Technology, Workshop at LREC2016*, Portorož, Slovenia, pp. 86–95.
- Stoyanova I., Yalamov, M., & Koeva, S. (2016). Quotation Retrieval System for Bulgarian Media Content. In: *Proceedings of the Second International Conference Computational Linguistics in Bulgaria 2016*, Sofia: Institute for Bulgarian Language, pp. 64–73.

<b>Previous projects/activities</b>
<p><b>ELRC</b> (European Language Resource Coordination; 2015–)</p> <p><b>CESAR</b> (2011–2013; CIP-ICT-PSP.2010.6.1) Central and South-East European Resources</p> <p><b>ATLAS: Applied Technology for Language-Aided CMS</b> (2010 – 2013 CIP-ICT-PSP-2009-3 71F – 0243/31.7.15)</p> <p><b>LASI: Internet Monitoring and Multi-language Content Analysis</b> (2014-2015)</p>
<b>Description of infrastructure</b>
<p>The institute for Bulgarian Language is a member of the Network of Excellence META-NET: Multilingual Europe Technology Alliance. META-SHARE within the Meta-Net is a sustainable network of repositories of language data, tools and related web services documented with high-quality metadata, aggregated in central inventories allowing for uniform search and access to resources. One of the META-SHARE repositories is maintained by the Institute for Bulgarian Language (<a href="http://metashare.ibl.bas.bg">http://metashare.ibl.bas.bg</a>).</p> <p>Lexicographic data:</p> <ul style="list-style-type: none"> <li>• Dictionary of synonyms</li> <li>• Dictionary of antonyms</li> <li>• Dictionary of new words</li> <li>• Dictionary of Bulgarian</li> </ul>

#### 10. Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa - Portugal

<b>Role in the project (cf. Consortium as a whole)</b>
<ul style="list-style-type: none"> <li>• Partner with expertise in standardisation</li> </ul>
<b>Description of the legal entity</b>
<p><b>Faculdade de Ciências Sociais e Humanas (FCSH)</b> of Universidade Nova de Lisboa (UNL) is the largest Portuguese institution of higher education in the fields of social sciences and the humanities. Its mission is to provide a high quality public service in tertiary education and research in the fields of social and human sciences, thereby raising national and foreign citizens' qualifications and producing scientifically and socially relevant knowledge. To fulfil this mission, the FCSH pursues: a) teaching and research excellence in the fields of the social and human sciences, at both national and international levels; b) a clear commitment to innovation and interdisciplinarity; c) the creation, dissemination and support for a humanistic spirit, and; d) the provision of services to the community in its areas of expertise. The global strategy of the FCSH is in clear connection with its social and economic environment, which enables it to respond to external stimuli in a sound and comprehensive way. FCSH's research policy stresses interdisciplinarity – both among the social sciences and between these and other scientific fields – and cooperation – both within the institution and with external institutions.</p>

**Centro de Linguística da Universidade Nova de Lisboa (CLUNL)** is a research unit that has as its main objectives the advancement of research in Theoretical and Applied Linguistics, the advanced training of researchers and the promotion of the publication of scientific data in those domains concerned with the nature and structure of languages and texts.

### Short CVs of main team members

**Rute Costa** (PhD) (F) is an Associate Professor (tenured) at the Universidade NOVA de Lisboa (UNL). She is an Invited Professor at the University of Luxemburg and at the Agostinho Neto University (Angola). In 2001, she was awarded a PhD in Terminology from Universidade NOVA de Lisboa. In 1989, she studied at the University of Paris III: Sorbonne Nouvelle for one semester and in 1993 she had a training period at the Centre ATO – CI (Centre for Research in Cognition and Information) at the University of Quebec, Montreal. At the UNL, she is scientific coordinator of the Master’s Programme in Terminology and Management of Information for Special Purposes and of the PhD Programme in Translation and Terminology. She is responsible for the Terminology seminars in the PhD Programme in Linguistics (Knowledge, Representation and Use). She is a researcher at the Linguistics Centre of the Universidade NOVA de Lisboa (CLUNL), and she is associate researcher at the Condillac Team of the Computer Science, Systems, Information and Knowledge Processing Laboratory (LISTIC) of the University of Savoy.

She was President of the European Association for Terminology (EAFT) between 2000 and 2006. She has been a liaison member of ISO TC 37 “Terminology and other language and content resources” since 2000. She has been the chairwoman of the ISO/TC 37/SC 2 “Terminographical and lexicographical working methods” since 2015. She is currently the convener of the ISO/TC 37/SC1 WG3 for revision of the standards 704 and 1087. In 2011, she was honoured with the title of Chevalier de l’Ordre des Arts et des Lettres by the Ministry of Culture and Communication of the French Republic.

**Christophe Roche** (PhD) (M). Since 1988, he is Full Professor in Artificial Intelligence at the University of Savoie, France. He lectures in Computer Science, Artificial Intelligence, Knowledge Representation, Logic, Terminology, and Ontology. He is currently in charge of the Condillac Research Group on “Terminology and Ontology” of the LISTIC Lab (Laboratoire d’Informatique, Systèmes, Traitement de l’Information et de la Connaissance) of Polytech’Annecy Chambéry (University Savoie Mont-Blanc).

He received his PhD in Artificial Intelligence (Expert System and Knowledge Representation) at the University of Grenoble (National Polytechnic Institute of Grenoble (INPG), France) in 1984. He worked in private research institutes in artificial intelligence in Grenoble and Paris, where he also taught Artificial Intelligence at the University of Paris VI. In 1986 he was Professor at the University of Neuchâtel (Switzerland) in charge of the post-graduate Artificial Intelligence course for the Universities of Neuchâtel, Lausanne and Fribourg. Since 2009, he is also Associate Researcher at the Linguistic Research Center (Centro de Linguística, Universidade NOVA de Lisboa) in Portugal, and lecturer in Ontology, Terminology and Knowledge Representation at the University NOVA of Lisbon.

He is an international Expert for the ISO (International Organization for Standardization) Technical Committee TC 37. He is the Project Leader of the ISO 1087-1 Standard on Terminology Vocabulary and of the ISO 704 Standards on Methods and Principles of Terminology. He is also the Chairman of the AFNOR (The French Organization for Standardization) Commission on Terminology, Principles and Coordination (X03A Commission).

He participated and is currently involved in several industrial projects on Terminology and Ontology: French Ministry of Education, Electricity of France (Nuclear plants, Water dams), INE (Instituto Nacional

de Estatística, Lisbon), GRETh (renewable energy), SAGEM (knowledge management), Wallonie-Bruxelles Federation (Terminology and Culture), etc.

He gives lectures in Terminology and Ontology at Master and PhD levels both for University Savoie Mont Blanc and University NOVA de Lisboa. He is responsible with Rute Costa (Lisbon) for the TOTH Training sessions in Terminology and Ontology.

He has published and given more than one hundred publications and invited talks. He participates in several international program committees of conferences in Ontology, Knowledge Engineering and Terminology: TOTH, NLPCS, EGC, OntoContent, VORTE, ICWIT, FLAIRS, ExaCt, FUBATEC, IEA/AIE, JFO, Web, TKE, etc.

**All the named persons are employed by the beneficiary or have an equivalent relationship covering the eligibility of costs (Cristophe Roche).**

### Relevant publications

- Almeida, B., Roche, C. & Costa, R. (2016). "[Terminology and ontology development in the domain of Islamic archaeology](#)" *TKE 2016, 12th International conference on Terminology and Knowledge Engineering*, Copenhagen, 22– 24 June 2016, pp. 147-156.
- Carvalho, S., Costa, R. & Roche, C. (2016). "LESS Can Indeed Be More. Linguistic and Conceptual Challenges in the Age of Interoperability". *TKE 2016, 12th International conference on Terminology and Knowledge Engineering*, Copenhagen, 22–24 June 2016.
- Costa, R. (2013). Terminology and Specialised Lexicography: two complementary domains. In R. Gouws, R. Hjalmar, U. Heid, S. J. Schierholz, W. Schweickard & H. E. Wiegand (Eds.), *Lexicographica. International Annual of Lexicography*. Volume 29, Issue 1. Berlin, New York: De Gruyter, pp. 29–42. DOI: 10.1515/lexi-2013-0004.
- Roche, C. (2015). Ontological definition. In H. Kockaert & F. Steurs (Eds.), *Handbook of Terminology*. Volume 1. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/hot.1, pp. 128-152.
- Santos, C. & Costa, R. (2015) Domain specificity: semasiological and onomasiological knowledge representation". In H. Kockaert & F. Steurs (Eds.), *Handbook of Terminology*. Volume 1. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/hot.1, pp. 153-179.

### Previous projects/activities

**UTOPIA**, Food and the Future: Utopian thinking and the construction of inclusive societies - a contribution from the Humanities [Ministry of Research and Science - PT] [PIDC/CPC-ELT/5676/2014] [2016-2019] The ultimate goal of the project is twofold: first, to contribute to the implementation of FS in Portugal, evincing the relevance of its humanistic perspective both at the university and at high schools; secondly, to aid the development of FS by offering a research methodology that enables scholars from any field to critically think about the future role of food.

**European Projet Européen FP 7 "AthenaPlus"** (2013 – 2015) - <http://www.athenaplus.eu/> Environment TMP for Thesaurus Management Platform version 2) for the construction of multilingual thesaurus based on Ontoterminology (OTe engine) in the respect of the standards ISO 25964-1 and ISO 25964-2. The principal objectives of the AthenaPlus project were to: I) Contribute more than 3.6 millions metadata records to Europeana; II) Improve search, retrieval and re-use of Europeana's content, bettering multilingual terminology management; III) Experiment with enriched metadata their re-use adapted for users with different needs .

**FP7 Siera** – Integrating Sina Institute into the European Research Area [FP7-INCO-2011-69] (2011-2014). The general objective of the project was to reinforce closer scientific cooperation between EU and Palestinian scientists in the field of multilingual and multicultural knowledge sharing technologies (The diversity of languages, cultures and standards are the main barriers to sharing and consuming knowledge).

**European Project - FP7 “Linked Heritage”** (2011-2013) - <http://www.linkedheritage.org/> "Linked Heritage - Coordination of Standards and Technologies for the enrichment of Europeana. The Linked Heritage Project had 3 main objectives: I) to contribute large quantities of new content to Europeana, from both the public and private sectors; II) to demonstrate enhancement of quality of content, in terms of metadata richness, re-use potential and uniqueness; III) to demonstrate enable improved search, retrieval and use of Europeana content.

**TKB:** a Transmedia Knowledge Base for contemporary dance. [PTDC/EAT-AVP/098220/2008 [Ministry of Research and Science - PT] (2010-2013) <http://tkb.fcsh.unl.pt/tags/dance>. In the interstices between linguistics and performing arts studies, TKB is an extensive and transdisciplinary project aiming at the design and construction of an open-ended multimodal knowledge-base to document, annotate and support the creation of contemporary dance pieces.

#### Description of infrastructure

N/A

### 11. K Dictionaries (KD)

#### Role in the project (cf. Consortium as a whole)

- Partner with lexicographic data and expertise
- Partner with computational linguistics data and expertise

#### Description of the legal entity

**K Dictionaries Ltd (KD)** is a technology-oriented-content creator that develops lexicographic data for 50 languages and cooperates with publishing, technology, academic and other professional partners worldwide on enhancing the interoperability of lexicography with language technologies. Its resources include cross-lingual varieties based on learners’ and multilingual dictionaries as well as tools for data compilation, auto-generation, processing, management and dissemination.

#### Short CVs of main team members

**Ilan Kernerman**, CEO (M). He has been involved in lexicographic research, development, project management, international marketing and cooperation with the industry and academia since 1991. He studied lexicography, publishing, business administration and mediation, and speaks at universities, language institutes and professional meetings across the world.

Ilan edits and publishes Kernerman Dictionary News (since 1994), has co-edited and published collections of lexicography papers (Lexicography in Asia, 1998, with Tom McArthur, and English Learners’ Dictionaries at the DSNA 2009, 2010, with Paul Bogaards), and was guest co-editor of the

International Journal of Lexicography special issue on bilingual learner's dictionaries (29:3, OUP, September 2016, with Arleta Adamska-Sal-aciak). He is the President of ASIALEX (The Asian Association for Lexicography), MC member of the COST-ENeL (European Network of e-Lexicography), initiator of GLOBALEX international constellation for lexicography, and on the Board of Trustees of Adam Kilgarriff Prize.

**Noam Ordan**, (PhD) Chief Research Officer (M). He graduated in translation studies from Bar Ilan University, did a post-doc at Haifa University, worked as teacher and researcher at universities in Germany and Israel, and took part in various academic and industry projects, including development of the Hebrew WordNet. He coordinates KD's research innovation and designs algorithms for using human-crafted lexicographic content for computational tasks such as cross-lingual information retrieval.

**Yifat Ben-Moshe**, Chief Content Officer (F). She graduated in Linguistics from Tel Aviv University, and joined KD in 2011. She is responsible for forming the company's editorial policy, developing cross-lingual lexicographic resources, coordinating projects with editors, translators, business and academic partners, and supervising student internship programs.

**Gidon Karni**, Chief Technology Officer (M). He graduated in software programming from Tel Aviv Engineering School, and is an IT system analyst in .NET environment. Before joining KD he was employed by Calanit software company (1987-2004) in technical support for products and automated systems, and developing various software programs.

He has extensive experience in creating and processing lexicographic data in XML format (and MS SQL systems), leads KD's data transformation to RDF, and is responsible for developing advanced tools for the data compilation, generation, manipulation and dissemination, statistics and quality assurance.

**All the named persons are employed by the beneficiary or will be employed after project starts or have an equivalent relationship covering the eligibility of costs.**

#### Relevant publications

- **K English Multilingual Dictionary.** English-English dictionary core with translations in 45 languages
- **The Global Series.** Monolingual, bilingual and multilingual lexicographic datasets for 24 languages
- **MultiGloss.** Semi-automatically generated glossaries for 22 languages, each including translation equivalents in 45 languages
- **Random House Webster's College Dictionary.** Extensive monolingual resource of American English
- **Aguado-de-Cea, G., Montiel-Ponsoda, E., Kernerman, I. & Ordan, N. 2016.** *From dictionaries to cross-lingual lexical resources.* META Forum, Lisbon. [http://www.meta-net.eu/events/meta-forum-2016/slides/29\\_cea.pdf](http://www.meta-net.eu/events/meta-forum-2016/slides/29_cea.pdf)

#### Previous projects/activities

**KIET** – K Index Editorial Tool – for developing smart word-to-sense L1>English indices.

**LDL4HELTA** – Linked Data Lexicography for High-End Language Technology Application – a EUREKA project as part of the Bilateral Austria-Israel R&D Framework, with Semantic Web Company (Vienna),



and the participation of the Ontology Engineering Group of Madrid Polytechnic University (UPM-OEG).  
<https://ldl4.com/>

**TIAD** – Translation Inference Across Dictionaries – shared task for automatic generation of bilingual dictionaries, thesauri and semantic fields based on existing lexicographic resources.  
<http://ldk2017.org/index.php/workshops/>

**Lexicographic data meet computational linguistics and knowledge based systems** – COST ENeL WG3 meeting, Brno (CZ) 16-17 September 2016. <http://www.elexicography.eu/working-groups/working-group-3/wg3-meetings/wg3-brno-2016/>

**GLOBALEX 2016** – Lexicographic Resources for Human Language Technology – workshop at LREC 2016.  
<http://ailab.ijs.si/globalex/>

### Description of infrastructure

KD has a solid infrastructure for lexicographic creation and data processing, manipulation, conversion, QA, dissemination and publication, including RDF modelling and automatic generation of cross-lingual lexical resources.

Lexicographic data:

- K English Multilingual Dictionary
- Global French Multilingual + L2-French
- Random House Webster's College Dictionary

## 12. CNR-ILC - Istituto di Linguistica Computazionale "A. Zampolli" - CNR - Italy

### Role in the project (cf. Consortium as a whole)

- Partner with computational linguistics data and expertise
- Partner with expertise in standardisation

### Description of the legal entity

**The Institute for Computational Linguistics “A. Zampolli”** (CNR-ILC) is a centre of reference in the field of Computational Linguistics at both national and international levels. The institute carries out research activities in strategic scientific areas of the discipline, as well as publishing activities, training and education activities and technology transfer. The research carried out within each area is highly interdisciplinary and involves different professional skills and expertise extending across several disciplines, including Linguistics, Computational Linguistics, and the Digital Humanities. The institute is able to boast a long-standing expertise in the creation, maintenance and promotion of language resources, with a special focus on lexical resources for natural language processing and for the digital humanities (and particularly in Latin and Ancient Greek). CNR-ILC has participated, and continues to participate, in important initiatives in the area of Language Resources and Technologies, such as, for instance, helping to organise, for instance, the biennial Language Resources and Evaluation Conference (LREC). CNR-ILC is also involved in the publishing of resources as Linked Open Data along with the development and evaluation of data models for this purpose. In addition, the institute has played an important role in the definition and promotion of standards for language resources. So that it has been involved in important EU projects on standardisation (e.g. LIRICS and INTERA) and is an active

participant within the International Standardisation Organization, ISO TC37 SC4. Its members collaborated in the creation of the Lexical Markup Framework, an ISO Standard for Machine Readable Dictionaries and NLP Dictionaries. CNR-ILC coordinated the EU FP7 thematic network on language resources, FLReNet, which produced important recommendations subsequently taken up by META-NET and CLARIN. The institute has also substantial experience in the development and adaptation of pipelines for linguistic annotation such as those developed for the project OPENER, and of orchestrated services for the automatic acquisition of language resources thanks to its active involvement in the PANACEA project. The promotion of digital content and resources for European minority languages is also one of the aims of the institute.

Finally, CNR-ILC has significant expertise in the development of platforms and infrastructures within the field of NLP and the digital humanities: it has been part of the CLARIN infrastructure since its preparatory phase as an FP7 project, and currently acts as the national coordinator of CLARIN-IT. CNR-ILC hosts a CLARIN-IT data center (C, aiming for B) and is active in the CLARIN standard committee, the standing committee for CLARIN technical centers, the user involvement WG, and in the National Coordinator Forum. CNR-ILC also acts as an executive institute within the DARIAH infrastructure.

### Short CVs of main team members

**Monica Monachini** (PhD) (F) is a senior researcher. She leads the Language Resources and Infrastructure group (LaRI) at CNR-ILC. Her main research interests include: Computational Lexicography and Lexicology; Lexical Semantics; Lexicons; Terminologies; Standards; Metadata; Infrastructural issues related to Language Resources. She has published over 150 articles in journals, books and conferences at an international level; organized several scientific events and participated in several international and national projects. She represents CNR in UNI Committees devoted to Terminology and is an expert delegate in ISO TC37 SC4 Committees for Language Resources Management. She has been nominated by MIUR as National Coordinator of the CLARIN-ERIC Infrastructure for Italy, CLARIN-IT.

**Valeria Quochi** (PhD) (F) is a researcher at CNR-ILC. Her main research interests include: Computational Lexicography and Lexicology, Lexical Semantics, Standards, Research Infrastructures. In particular she works on the representation, modelling and (semi-)automatic acquisition of lexical information, with a focus on multiword expressions. She has participated in the organisation of several conferences and workshops as a member of the scientific committee, including LREC from 2008 onwards. She has actively participated in several EU projects, including FLReNet, CLARIN preparatory phase, and PANACEA, for which she led the evaluation work package.

**Fahad Khan** (PhD) (M) is a post-doctoral researcher at CNR-ILC. Currently his two main research interests are in the modelling of language resources as Linked Open Data, standards for representing temporal lexical data, as well as in the application of techniques from Knowledge Representation and Engineering to the Digital Humanities. He is currently involved in the development of computational formats and standards for the modelling of etymological data and attestations in lexical datasets and is the project leader for the Diachrony-Etymology extension of LMF.

**Irene Russo** (PhD) (F) is a post-doctoral researcher at CNR-ILC with a research interest in computational semantics - more specifically corpus based methodologies for the representation of adjectival and verbal meanings - and sentiment analysis. She has experience in the definition of annotation guidelines and in crowdsourcing experiments. More recently she was involved in promoting digital production of content and resources for European minority languages through training programmes for adult learners with the Erasmus+ project DLDP.



**All the named persons are employed by the beneficiary or will be employed after project starts or have an equivalent relationship covering the eligibility of costs**

**Relevant publications**

- Francopoulo, G., George, M., Calzolari, N., Monachini, M., Bel, N., Pet, M. & Soria, C. (2006). Lexical markup framework (LMF). In Proceedings of LREC, Fifth International Conference on Language Resources and Evaluation (Vol. 6), pp. 233-236.
- Khan, F., Diaz-Vera, J. E. & Monachini, M. (2016). Representing Polysemy and Diachronic Lexico-Semantic Data on the Semantic Web. *Proceedings of the Second International Workshop on Semantic Web for Scientific Heritage Co-Located with 13th Extended Semantic Web Conference (ESWC 2016)*, Heraklion, Greece, May 30th, 2016, pp. 37-45.
- Monachini, M. & Frontini, F. (2016). CLARIN, l'infrastruttura europea delle risorse linguistiche per le scienze umane e sociali e il suo network italiano CLARIN-IT. *JCoL - Italian Journal of Computational Linguistics*, 2(2), Special Issue on NLP and Digital Humanities, 11-30.
- Quochi, V., Frontini, F. & Rubino, F. (2012). "[A MWE Acquisition and Lexicon Builder Web Service](#)". In *Proceedings of COLING 2012: Technical Papers*, pp. 2291–2306, COLING 2012, Mumbai, December 2012.
- Russo, I. & Caselli, T. (2013). Changeable Polarity of Verbs through Emotions' Attribution in Crowdsourcing Experiments. In *First International Workshop on Emotion and Sentiment in Social and Expressive Media: approaches and perspectives from AI (ESSEM 2013)*, A workshop of the XIII International Conference of the Italian Association for Artificial Intelligence (AI\*IA 2013), pp. 131-139.

**Previous projects/activities**

The project **KYOTO** (Knowledge Yielding Ontologies for Transition-based Organization) (FP7) had the goal of providing a system for deep semantic content search, allowing experienced users to model and improve their domain ontology using automatically extracted terms and concepts. The system enables expert users to assess and integrate candidate concepts and terms in their knowledge base, thus enhancing the analysis of new documents. It resulted in a format for lexicons and facilitated the linking of lexicons with ontologies for improving NLP systems.

**PANACEA** (Platform for Automatic, Normalized Annotation and Cost-Effective Acquisition of Language Resources for Human Language Technologies) was a STREP action that developed a factory of Language Resources (LRs) in the form of a production line that automated all the steps involved in the acquisition, production, maintenance and updating of the LRs required by Language Technologies

**FLaReNet** (Fostering Language Resources Network) was a European forum to facilitate interaction among Language Resources stakeholders. It aimed at developing a common vision of the area as well as fostering a European strategy for consolidating the sector, thus enhancing competitiveness at the EU level and worldwide. By creating a consensus among major players in the field, the mission of FLaReNet was to identify priorities as well as short, medium, and long-term strategic objectives and to provide consensual recommendations in the form of a plan of action for European, national organisations and industry. Through the exploitation of new collaborative modalities as well as workshops and meetings, the FLaReNet Network sustained international cooperation and (re)created a wide Language community.

**OpeNER** (Open Polarity Enhanced Named Entity Recognition) developed a suite of NLP modules in 6 languages to perform natural language processing tasks. These modules are available for free and have been designed so that they are easy for those working in academia and research or in Small and Medium Enterprises to adapt and integrate them into their workflows. The OpeNER components have

been proven and tested on Tourism domain; it is reasonable easy, however, to adapt them to another specific domain.

#### Description of infrastructure

CNR-ILC has implemented the first national data center (ILC4CLARIN) in the CLARIN-IT infrastructure. The center provides the SSH scientific community with easy and secure access to language resources in CLARIN repositories. CLARIN-IT is now offering data, lexicons, tools and applications to support language processing tasks, with a special emphasis on Italian and classical languages such as Greek and Latin.

META-SHARE was an open language resource exchange facility, devoted to the sustainable sharing and dissemination of language resources (LRs). It aimed at increasing access to such resources on a global scale. CNR was an official META-SHARE node.

### 13. DSL – The Society for Danish Language and Literature

#### Role in the project (cf. Consortium as a whole)

- Partner with lexicographic data and expertise (as well as computational linguistics expertise)

#### Description of the legal entity

**The Society for Danish Language and Literature** (Det Danske Sprog- og Litteraturselskab, DSL) is an independent institution supported by the Danish Ministry of Culture, the Carlsberg Foundation, and a series of other foundations. DSL is charged with publishing and documenting Danish language and literature from the earliest times to the present day. The aims of DSL are to edit and publish works from all periods of Danish literature, bibliographies as well as documents regarding Danish history written before the year 1450. Furthermore DSL edits and publishes a large number of scholarly dictionaries as well as lexicons for NLP, and builds text corpora and other electronic language resources, e.g. as resources for dictionary-making. The society consists of around 85 members and employs around 30 academic staff. Many of the works published by DSL are freely accessible on the internet, including the Archive of Danish Literature, [www.adl.dk](http://www.adl.dk), Diplomatarium Danicum, [www.diplomatarium.dk](http://www.diplomatarium.dk), a corpus of medieval Danish texts <http://middelaldertekster.dk>, and a series of Danish dictionaries, published at [www.ordnet.dk](http://www.ordnet.dk), including DDO and ODS. DSL has a high level of expertise in digital lexicography, retro-digitalization, corpus linguistics and lexical semantics. DSL is involved in several research projects with UCPH, and has developed a number of resources over the years in collaboration with CST: The Danish WordNet *DanNet* (<https://wordnet.dk>); The SemDax Corpus (semantically annotated corpus) (<https://github.com/coastalcph/semDax>), and currently FrameNet for Danish (with CST and Anders Søgaard UCPH).

#### Short CVs of main team members

**Sanni Nimb** (PhD) (F), PhD in Computational lexicography; Senior Editor and researcher at DSL. She edits modern Danish Dictionaries (the online dictionary DDO and the Danish Thesaurus), and compiles lexicons for NLP based on the two dictionaries. She is the main editor of the published 1. edition of the Danish Thesaurus (2014), and established the xml-structure and the set of formal relations used to

annotate semantic groups across the thesaurus vocabulary. Her research interests are devoted to computational lexicography and formal lexical semantics. Since 2013 she has been involved in the national research project “Semantic Processing across Domains” together with UCPH, mainly with the compilation of lexical semantic data of different granularity to be used in semantic annotation tasks as well as in experiments with sense tagging of Danish texts. She is the main linguist in the FrameNet project for Danish (2016-2017) where thesaurus groups are assigned frames from Berkeley FrameNet. She is partner in the Danish wordnet project, DanNet (grant project from the Danish Research Council 2005-2008) with CST and member of Board of Representatives of the Danish Language Council 2017-. For full cv, see: <https://dsl.dk/medarbejdere/sn>.

**Nicolai H. Sørensen** (M), Senior Editor, works as a computational linguist on dictionary projects carried out at the Society for Danish Language and Literature. He is technically in charge of the development and publishing of online dictionaries and apps as well as the optimisation and maintenance of DSL’s dictionary databases. Furthermore he works on retro-digitisation and publishing of historical dictionaries, including recognition and mark-up of information types. He also participates in research projects with UCPH, e.g. the Danish WordNet DanNet (grant project from the Danish Research Council 2005-2008) and Semantic Processing across Domains (grant project from the Danish Research Council 2013-2017) where he works with semantic parsing and machine learning. For full cv, see: <https://dsl.dk/medarbejdere/nhs>

**Thomas Troelsgård** (M), Senior Editor, works as a computational linguist on several of the dictionary projects carried out at the Society for Danish Language and Literature. His work includes development, optimisation and maintenance of dictionary databases and entry structures, error detection, error correction, and consistency checking in dictionary data, as well as lexicographic work as editor of dictionary entries. He also works with extraction and conversion of dictionary data, e.g. in connection with the compilation of The Danish Thesaurus. Furthermore he works on retro-digitisation of (mainly historical) dictionaries, e.g. on recognition and mark-up of information types. For full cv, see: <https://dsl.dk/medarbejdere/tt>.

**All the named persons are employed by the beneficiary.**

#### Relevant publications

- Martinez Alonso, H., Johannsen, A. T., Olsen, S., Nimb, S., Sørensen, N. H., Braasch, A., Søgaard, A., Pedersen, B. S. (2015). Supersense tagging for Danish. In: *Proceedings of the 20th Nordic Conference of Computational Linguistics NODALIDA 2015. Vol. 109* Linköping University Electronic Press. (NEALT (Northern European Association of Language Technology) Proceedings Series, Vol. 23).
- Nimb, S., Pedersen, B. S., Braasch, A., Sørensen, N. H. & Troelsgård, T. (2013). Enriching a wordnet from a thesaurus. In: *Workshop Proceedings on Lexical Semantic Resources for NLP from the 19th Nordic Conference on Computational Linguistics (NODALIDA)*. Linköping Electronic Conference Proceedings; Volume 85, pp. 36-50.
- Pedersen, B. S., Braasch, A., Johannsen, A., Martínez Alonso, H., Nimb, S., Olsen, S., Søgaard, A. & Sørensen, N. H. (2016). The SemDaX Corpus - sense annotations with scalable sense inventories. In *Proceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC)*, Portorož, Slovenia.
- Pedersen, B. S., Nimb, S., Asmussen, J., Sørensen, N. H., Trap-Jensen, L., Lorentzen, H. (2009). DanNet – the challenge of compiling a WordNet for Danish by reusing a monolingual dictionary. *Language Resources and Evaluation, Computational Linguistics Series*, pp. 269-299.
- Troelsgård, T., Akhøj Nielsen, M. (2016). Moth’s Danish Dictionary. Publishing a Dictionary with a 300-year Delay. In T. Margalitadze & G. Meladze (Eds.), *Proceedings of the XVII EURALEX International Congress*, 6-10 September, 2016. Ivane Javakhishvili Tbilisi University, Tbilisi, pp.

622-634.
<b>Previous projects/activities</b>
<p><b>Semantic Processing across Domains</b>, Collaborate project with UCPH, funded by the Danish Research Council (2013-2017) concerned with semantic processing and domain adaptation and includes annotation and processing of Danish texts (i. e. The SemDax Corpus, a semantically annotated Danish corpus). <a href="http://cst.ku.dk/projekter/semantikprojekt/">http://cst.ku.dk/projekter/semantikprojekt/</a></p> <p><b>From thesaurus to FrameNet</b> (2016-2017) Collaborate project funded project (Carlsberg Foundation) with the goal of developing a publicly accessible, manually constructed and validated Danish FrameNet on the basis of the Danish Concept Dictionary / Thesaurus (UCPH)</p> <p><b>DanNet</b>, funded by the Danish Research Council (ended 2008). On compiling a Danish wordnet on the basis of a monolingual dictionary (with UCPH).</p>
<b>Description of infrastructure</b>
<p>Lexicographic data:</p> <ul style="list-style-type: none"> <li>• The Danish Dictionary</li> <li>• Dictionary of the Danish Language</li> <li>• Moths Dictionary</li> <li>• Old Danish Dictionary</li> <li>• Danish Thesaurus</li> </ul>

#### 14. Centre for Language Technology at University of Copenhagen (UCPH)

<b>Role in the project (cf. Consortium as a whole)</b>
<ul style="list-style-type: none"> <li>• Work package leader (WG9)</li> <li>• Partner with computational linguistics data and expertise</li> </ul>
<b>Description of the legal entity</b>
<p><b>Centre for Language Technology at the Department of Nordic Research, University of Copenhagen</b> (UCPH) employs approximately 16 staff, including computational linguists, linguists, lexicographers, terminologists, computer scientists and engineers. UCPH has a high level of expertise in many areas of LT and particular expertise in computational lexical semantics.</p> <p>Teaching and research activities at the Centre are organized around a number of themes: research infrastructure, language technology applications, multimodal communication, language processing and resources, computational lexicography, and machine learning.</p> <p>The Centre has developed a large number of resources, tools and platforms over the years in collaboration with other partners of which five are selected below:</p> <ul style="list-style-type: none"> <li>• The Danish WordNet, DanNet (with DSL) (<a href="https://wordnet.dk">https://wordnet.dk</a>)</li> <li>• The SemDax Corpus (semantically annotated corpus) (with DSL) (<a href="https://github.com/coastalcph/sem dax">https://github.com/coastalcph/sem dax</a>)</li> </ul>

- A Danish sense tagger ([https://github.com/coastalcph/dsl\\_semtagger](https://github.com/coastalcph/dsl_semtagger))
- AndreOrd – a DanNet wordnet browser (<https://andreord.dk>)
- WordTies – a multilingual wordnet browser for Nordic, Baltic and Polish wordnets (<https://wordties.cst.dk>)

### Short CVs of main team members

**Bolette S. Pedersen** (F), PhD in Computational Linguistics. Project leader of a number of projects concerned with computational lexicography and NLP: Semantic Processing across Domains (collaborate grant project from the Danish Research Council 2013-2017 with DSL). Project leader of the Danish wordnet project, DanNet (grant project from the Danish Research Council 2005-2008) with DSL. Member of Board of Representatives of the Danish Language Council 2010-2016. President of NEALT (Northern European Association of Language Technology) 2014-2016. Leader of the Danish METANET/METANORD group (an EU project with aim of developing language resources for the less resourced languages in the Baltic and Nordic countries, 2011-2013). For full cv, see: <http://cst.dk/bolette/my.html>

**Lene Offersgaard** (F), Senior Development Engineer, is Master of Science in Electronic Engineering. Since 1999, she is a senior R&D Engineer at the Centre for Language Technology. Her areas of expertise are in particular language resources and tools, research infrastructure, and statistical machine translation. Technical responsible for STO (Danish NLP lexicon). Now participating in Eurostar ReProsis project (Real Time Big Data Product Analysis) with development and evaluation of linguistically-aware models that will be integrated in a multilingual ontology. Chair of CLARIN ERIC Assessment committee.

**Sussi Olsen** (F), Research Associate, coordinated the manual semantic annotation of the SemDaX-corpus using the WebAnno tool in the Danish research project Semantic Processing across Domains, and is still part of the FrameNet annotation of the same corpus. She participated in several large-scale European projects such as Language Technology Observatory, CLARIN-PLUS, META-NORD, EuroTermBank, Let'sMT!, as well as many national research projects. She has been involved in several lexicographical projects including STO - the Danish lexical database for language technology applications. She worked as an expert validator of written language resources for ELRA and Netherlands Organisation for Scientific Research and is the (co)author of many publications on language resources, encompassing lexicography, corpus compilation and annotation, metadata and machine translation.

**All the named persons are employed by the beneficiary.**

### Relevant publications

- Offersgaard, L., & Hansen, D. H. (2016). *Facilitating Metadata Interoperability in CLARIN-DK*. Abstract from 10th edition of the Language Resources and Evaluation Conference, Portorož, Slovenia.
- Olsen, S., Pedersen, B. S., Martínez Alonso, H., Johannsen, A. (2015). [Coarse-grained sense annotation of Danish across textual domains](#). In *Proceedings of the Workshop on Semantic resources and Semantic Annotation for Natural Language Processing and the Digital Humanities at NODALIDA 2015*, Linköping University Electronic Press, Sweden, pp. 36-43.
- Pedersen, B. S., Braasch, A., Johannsen, A., Martínez Alonso, H., Nimb, S., Olsen, S., Sjøgaard, A. & Sørensen, N. (2016). The SemDaX Corpus - sense annotations with scalable sense inventories. In *Proceedings of the 10th edition of the Language Resources and Evaluation Conference*, Portorož, Slovenia, pp. 842-847.

- Pedersen, B. S., Lindén, K., Vider, K., Forsberg, M., Kahusk, N., Niemi, J., Nygaard, L., Seaton, M., Orav, H., Borin, L., Voionmaa, K., Nisbeth, N. & Rognvaldsson, E. (2013). Nordic and Baltic wordnets aligned and compared through “WordTies”. *Proceedings from the 19th Nordic Conference on Computational Linguistics (NODALIDA)*. Linköping Electronic Conference Proceedings; Volume 85, pp. 147-162.
- Pedersen, B. S., Nimb, S., Asmussen, J., Sørensen, N., Trap-Jensen, L., Lorentzen, H. (2009). DanNet – the challenge of compiling a WordNet for Danish by reusing a monolingual dictionary. *Language Resources and Evaluation, Computational Linguistics Series*, pp. 269-299.

#### Previous projects/activities

**Semantic Processing across Domains**, Collaborate project funded by the Danish Research Council (ends 2017) concerned with semantic processing and domain adaptation and includes annotation and processing of Danish texts (with DSL)

**From thesaurus to FrameNet** (ends 2017) Collaborate project funded project (Carlsberg Foundation) with the goal of developing a publicly accessible, manually constructed and validated Danish FrameNet on the basis of the Danish Concept Dictionary (with DSL)

**ReProsis, EU Eurostars project** (ends 2018) about tracking products across different platforms, languages and countries. CSTs focus is on recognizing semantic relations across languages.

**METANET/METASHARE: META-NORD** (ended 2013) was an EU project which aimed at developing and documenting methodologies for building language resources for the under-resourced languages in the Baltic and Nordic countries, with focus on semi-automatic/machine assisted resource generation. Centre for Language Technology has set up a META-SHARE node as a repository for Danish language tools and resources.

**DanNet**, funded by the Danish Research Council (ended 2008). On compiling a Danish wordnet on the basis of a monolingual dictionary (with DSL).

#### Description of infrastructure

The Centre administrates the CLARIN-DK infrastructure platform for language resources and technology (<http://info.clarin.dk/>), as well as the national META-SHARE node <http://metashare.cst.dk/>.

Lexicographic data:

- Dictionary of Danish Insular Dialects
- Dictionary of Old Norse Prose

### 15. Trier Center for Digital Humanities (TCDH)/Trier University - Germany

#### Role in the project (cf. Consortium as a whole)

- Partner with lexicographic data and expertise
- Digital humanities partner (e-learning, digital editions)

#### Description of the legal entity



**The Trier Center for Digital Humanities (TCDH)** is an internationally renowned institution for Digital Humanities. Founded as one of the first of its kind in 1998, the centre has been conducting pioneering work in the growing field of research that is Digital Humanities for nearly 20 years. Since its foundation, more than 100 diverse projects in the field of Digital Humanities were implemented at the centre – in cooperation with more than 150 regional, national, and global partners at universities, research institutions, academies of sciences, publishers, and institutions of memory. The major focal points in research and development lie in the areas of digital publication and networking of dictionaries, sources, editions, and other fundamental works and artefacts in the Humanities, as well as on developing tools for the preparation, indexation, analysis, and visualisation of corpora. By developing virtual research environments and workflow management tools, the centre supports research processes in the Humanities and enables inter- as well as transdisciplinary cooperation. In the last about 20 years, the TCDH formed an interdisciplinary Digital Humanities team with competences in Lexicography, in Computer Science, and Computer Linguistics and collected a huge amount of dictionary data as well as outstanding expertise in encoding and publishing lexicographical resources.

Since 1998, the centre has digitised, published, and interlinked several Middle High German Dictionaries, the *Deutsche Wörterbuch* by Jacob and Wilhelm Grimm, the *Rheinische Wörterbuch*, the *Pfälzische Wörterbuch*, the *Wörterbuch der deutsch-lothringischen Mundarten*, the *Wörterbuch der elsässischen Mundarten*, the Goethe Dictionary. These dictionaries are part of the Trier Dictionary Network which cooperatively interlinks dictionaries such as for example the New Middle High German dictionary, the *Deutsche Rechtswörterbuch* or the *Schweizerisches Idiotikon*. Furthermore, as a service centre the Trier Center for Digital Humanities digitises and, in some cases, encodes dictionaries for external partners.

Additionally, its expertise and practical experience shape the curriculum of the bilingual MSc programme in Digital Humanities, a study programme established in 2014 at the University of Trier in cooperation with the Trier Center for Digital Humanities. Modules on Digital Lexicography and Intellectual Property Rights (the latter in cooperation with the Trier University professor for Civil Law who is an expert especially on Information Society Law and Intellectual Property Rights) are part of the curriculum enabling a great degree of training to obtain practical skills.

#### Short CVs of main team members

**Vera Hildenbrandt** (PhD) (F) is a member of academic staff at the University of Trier and Executive Director at the Trier Center for Digital Humanities in the field of Digital Philology. She supervises projects at the interface between traditional and digital philologies – amongst others numerous endeavours in digital lexicography (e.g. the digitisation and digital publication of the first as well as the revised edition of the German Dictionary by Jacob and Wilhelm Grimm and the Goethe Dictionary). Dr Vera Hildenbrandt is a member of various international research groups such as the DFG network “Internet Lexicography” and the Cost Action IS1305 “European Network of eLexicography” (ENeL; member of the steering group, chair of Working Group 2 “Retro-digitised Dictionaries”). Her main research interests lie in the areas of Digital Humanities and Digital Lexicography. Since October 2015, she has held the position of Deputy Professor for Digital Humanities at the University of Trier.

**Thomas Burch** (PhD) (M) is a member of academic staff at the University of Trier and Executive Director at the Trier Center for Digital Humanities in the field of Information Technology. He develops software solutions aimed at answering ambitious research questions in the Humanities with particular conceptual foci on intuitively ergonomic graphical user interfaces on the one hand and the implementation of newest technical standards in the areas of software design and data modelling on the other hand. The software systems developed in this manner (e.g. “Dictionary Network”, “Transcribo”, “Research Network and Database System – FuD”, and the “Trierer Artikel Redaktionssystem – TAREs”) are constantly used in manifold digital projects such as “Arthur Schnitzler:

Digital Historical-Critical Edition” and “New Middle High German Dictionary”. Dr Thomas Burch’s main research interests include the conceptualisation and adaptation of methods and processes in the field of Information Technology and their implementation for research questions in the Humanities.

**All the named persons are employed by the beneficiary.**

### Relevant publications

- Hildenbrandt, V. & Klosa, A. (Eds.). (2016). *Lexikographische Prozesse bei Internetwörterbüchern*. Mannheim: Institut für Deutsche Sprache. Doi: 10.14618/opal\_01-2016.
- Hildenbrandt, V. & Moulin, C. (2012). Das Trierer Wörterbuchnetz. Vom Einzelwörterbuch zum lexikographischen Informationssystem. *Korrespondenzblatt des Vereins für niederdeutsche Sprachforschung* 119, 73–81.
- Hildenbrandt, V. (2011). TEI-basierte Modellierung von Retrodigitalisaten (am Beispiel des Trierer Wörterbuchnetzes). In A. Klosa & C. Müller-Spitzer (Eds.), *Datenmodellierung für Internetwörterbücher. 1. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikographie“*. *Online publizierte Arbeiten zur Linguistik* 2011 (2), 21–35.
- Burch, T. & Rapp, A. (2007). Das Wörterbuch-Netz: Verfahren – Methoden – Perspektiven. In Daniel Burckhardt, Rüdiger Hohls und Claudia Prinz, unter Mitwirkung von S. Barteleit, G. Gersmann, P. Haber, M. Herren, P. Sahle, D. Schlögl, G. Vogeler, C. Wagner & I. Zündorf, *Geschichte im Netz: Praxis, Chancen, Visionen. Beiträge der Tagung .hist 2006*, Berlin 2007 (Historisches Forum 10, Teilband I), pp. 607–627.
- Bartz, H.-W., Burch, T., Christmann, R., Gärtner, K., Hildenbrandt, V., Schares, T. & Wegge, K. (Eds.). (2004) *Der Digitale Grimm. Deutsches Wörterbuch von Jacob und Wilhelm Grimm. Elektronische Ausgabe der Erstbearbeitung*. Hg. vom Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier in Verbindung mit der Berlin-Brandenburgischen Akademie der Wissenschaften. 2 CD-ROMs, Benutzerhandbuch, Begleitbuch. Frankfurt am Main: Zweitausendeins 2004.

### Previous projects/activities

**eHumanities Centre for Historical Lexicography (ZHistLex):** The TCDH is one of six partners in Germany in an initiative aiming to build an eHumanities Centre for Historical Lexicography (ZHistLex) which is funded by the German Federal Ministry of Education and Research since October 2016. Based on dictionaries and corpora already available in the collaborating institutions, the project aims at developing generic methods for the integration, presentation, and usage of data. These methods are used to establish a research portal that will also be available for future research. The project will develop new possibilities of data access and representation in digital lexicographical systems. The goals of this project are the creation of a portal, web-services, a metadata scheme for an overall bibliography of dictionary sources, an example illustrating the bidirectional linking of quotations in the dictionary with their digital full texts, practice cases illustrating word family structures across epochs, and many more. From a long-term perspective, the project aims at establishing a centre for Historical Lexicography that develops innovative methods and techniques in the area of Historical Lexicography and offers service, advice and support within this field.

**The Trier Dictionary Network:** The Dictionary Network presents a network of various reference works accessible via a single user interface. It enables its users to consult more than 20 reference works at the same time, answering modern research questions within seconds. Apart from the linking of its own dictionary resources provided by the TCDH, the Dictionary Network also offers the possibility of integrating external electronic documents. ([www.woerterbuchnetz.de](http://www.woerterbuchnetz.de))



**Digital Network of Dialect Dictionaries:** Within the frame of the project, the TCDH digitised four dialect dictionaries, published them online, and interlinked them with each other. The *Pfälzische Wörterbuch* turned out to be the ideal starting point as it was the most recently edited dictionary of the corpus and parts of it had already been digitally captured. Moreover, it also refers to other dictionaries of bordering dialects wherever these have a corresponding entry. Afterwards, the Network of Dialect Dictionaries was linked to the digitised version of Matthias Lexer's *Mittelhochdeutsches Handwörterbuch*. Enhancing the scope, the first edition of the *Deutsche Wörterbuch* by the Brothers Grimm was integrated into the network with hyperlinks, enabling users to draw upon a standard language approach to dialect lexis. Furthermore, since the TCDH also developed a digital map of the dialects in question, the Network exhibits profound points of contact with dialect geography. Electronic mappings which rely on the places referred to in the dictionaries allow for documenting a lexicon's regionality much more clearly than before. Thus, the Network of Dialect Dictionaries now represents an innovative resource for research into German dialects which can also be extended by further dictionaries.

**TAReS: A Webbased System for Editing, Producing, and Publishing Dictionaries in Distributed Offices:** Between 2002 and 2005, the Trier Center for Digital Humanities collaborated with the editorial division of the Middle High German Dictionary at the University of Trier ([Arbeitsstelle des Mittelhochdeutschen Wörterbuches an der Universität Trier](#)) to create a database-driven article editing system for decentralised editing of dictionaries which can be used via the internet. The programme supports all phases of the division's workflow: the ascertainment and lemmatisation of references in an entry, the gradual structuring, analysis, and commenting of references to establish a detailed article, and, finally, the publication of the dictionary and its content in different media. The system – which was originally developed for a historical reference dictionary – is ready to be expanded to and adapted for other dictionary projects.

**Middle High German Dictionary Network:** From 1997 to 2002, the TCDH digitised the *Mittelhochdeutsche Wörterbuch* by Georg Friedrich Benecke, Wilhelm Müller, and Friedrich Zarncke (BMZ, 1854–1866), the *Mittelhochdeutsche Handwörterbuch* by Matthias Lexer (1872–1878), including its supplements, and the *Findebuch zum mittelhochdeutschen Wortschatz* by Kurt Gärtner and Gerhard Hanrieder (1986–1992). The textual data were encoded in XML and published online as well as on CD-ROM. The digitised network of dictionaries represents, in an idealised manner, the structure of references that has been established in the dictionaries. It also offers an effective search engine which interlinks all four dictionaries.

#### Description of infrastructure

Lexicographic data:

- The German Dictionary by Jacob and Wilhelm Grimm (first edition)
- Rhenish Dictionary
- Palatinate Dictionary
- Dictionary of the German-Lorraine Dialects
- Dictionary of the Alsatian Dialects
- Grammatical-Critical Dictionary of the High-German Idiom (Adelung, second edition)
- Middle High German Dictionary (Benecke, Müller, Zarncke)
- Middle High German Dictionary (Lexer)

## 16. Institute of the Estonian Language

### Role in the project (cf. Consortium as a whole)

- Partner with lexicographic data and expertise

### Description of the legal entity

**The Institute of the Estonian Language** is a research and development institution under the Ministry of Education and Research. The Institute is one of the three centres where Estonian is systematically studied, and the only institution dealing with the development of the language. Its main aim is to contribute to the long-term preservation of the Estonian language and to guarantee the functioning of Estonian as the official national language. The Institute performs various public functions and fulfils major national tasks: compilation and upgrading dictionaries and databases essential for the country and national culture (e.g. Dictionary of Standard Estonian, Explanatory Dictionary, bilingual dictionaries), free public linguistic advice, language care and language planning (the Institute has the status of the Office of Onomastic Expertise), coordination of all-Estonian terminology work, incl. compilation of multilingual termbases, services for people with special needs (speech technology, sign language dictionary), online services for language learners (learners' dictionaries for different language proficiency levels), development of speech synthesis for Estonian, and the collection and development of language archives. The electronic resources and applications of the Institute are available to the public for free.

The Institute investigates contemporary Estonian, the history of the Estonian language, Estonian dialects and cognate Finno-Ugric languages. The whole lexicographic work is based on corpus linguistics, lexical semantics and language technology. The Institute aims to create a general framework for studying and developing the national language, for its cultivation, codification, and corpus-based lexical/linguistic monitoring and analysis. The Institute participates in the development of language policies in Estonia and the EU.

### Short CVs of main team members

**Jelena Kallas** (PhD) (F) is a computational lexicographer at the Institute of the Estonian Language, Tallinn, Estonia. She has been involved in different lexicographical projects, including monolingual and bilingual print and online dictionaries. Her research interests include corpus lexicography, automated lexicography, scholarly lexicography, dictionary use and innovative ways for presenting lexicographic data. Jelena has published on various aspects of presenting morphological and syntactic data in learners' dictionaries, on Estonian module of Sketch Engine and on lexicographic tools produced at the Institute of the Estonian Language. Jelena has been a member of EURALEX Executive Board since 2014, she is a member of organizing and scientific committee of eLex (Electronic lexicography in the 21st century) conferences since 2013 and a member of a management committee of ISCH COST Action IS1305 European Network of e-Lexicography (ENeL).

**Margit Langemets** (PhD) (F) is a senior lexicographer and the chief editor of dictionaries at the Institute of the Estonian Language. Her research interests include e-lexicography, corpus linguistics and lexical semantics. She has been involved in several bilingual and monolingual dictionary projects, as well as in the development of in-house DWS system EELEX, and as an expert in developing the new DWS EKILEX (since 2017).

**Kristina Koppel** (MA) (F) is a lexicographer and junior researcher at the Institute of the Estonian Language and a PhD student at the University of Tartu. Her research interests are mainly devoted to e-

lexicography, namely automatic extraction of lexicographic data from large text corpora. Her thesis is about parameters of example sentences for Estonian learners' dictionaries. For the past eight years she has been involved in three dictionary projects: Estonian Word Families (2012), Basic Estonian Dictionary (2014), Estonian Collocations Dictionary (2018) and their web interfaces and also in development of in-house DWS EELEX.

**All the named persons are employed by the beneficiary.**

#### Relevant publications

- Koppel, K. & Kallas, J. (2016). Õppijasõbralik korpuslause: automaatse valiku võimalusi [User-friendly sentence: possibilities for automatic acquisition]. In *Lähivõrdlusi*. Lähivertailuja, 26, pp. 222–250.
- Kallas, J., Kilgarriff, A., Koppel, K., Kudritski, E., Langemets, M., Michelfeit, J., Tuulik, M. & Viks, Ü. (2015). Automatic generation of the Estonian Collocations Dictionary database. Kosem, I., M. Jakubíček, J. Kallas & S. Krek (Eds.). *Electronic lexicography in the 21st century: linking lexical data in the digital age. Proceedings of the eLex 2015 conference, 11-13 August 2015, Herstmonceux Castle, United Kingdom*. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd., 1–20.
- Kallas, J., Tuulik, M. & Langemets, M. (2014). The Basic Estonian Dictionary: the first Monolingual L2 learner's Dictionary of Estonian. In A. Abel, C. Vettori, N. Ralli (Eds.), *Proceedings of the XVI EURALEX International Congress: The User in Focus, 15-19 July 2014, Bolzano/Bozen*. Bolzano/Bozen: European Academy, pp. 1109–1119.
- Kallas, J. & Langemets, M. (2012). Automatic Generation of Specialized Dictionaries Using the Dictionary Writing System EELEX. In A. Tavast, K. Muischnek & M. Koit, *Human Language Technologies – The Baltic Perspective, 247: Proceedings of the Fifth International Conference Baltic HLT 2012*. IOS Press, pp. 103–110. (Frontiers in Artificial Intelligence and Applications).
- Langemets, M., Loopmann, A. & Viks, Ü. (2010). Dictionary management system for bilingual dictionaries. In S. Granger, M. Paquot (Eds.). *eLexicography in the 21st century: New challenges, new applications. Proceedings of eLex 2009, Louvain-la-Neuve, 22-24 October 2009*. Louvain-la-Neuve: Presses universitaires de Louvain, Cahiers du CENTAL, pp. 425–430.

#### Previous projects/activities

**"The Creation of the Dictionary and Terminology Management System EELEX2 and Supporting the Mobility of the Institute's Scholars, Involvement of Foreign Researchers and Continuation of the Graduate School (2016–2022)"**, a EU project funded within [the Framework of the Institutional Development Programme for R&D and Higher Education Institutions \(ASTRA\)](#).

**"Modelling intermodular phenomena in Estonian (2009-2014)"**, Estonian Institutional Research Funding SF0050023s09. The outcome of the project is a fundamental study into several aspects of Estonian and of language structure in general, and the basic data models for complex Estonian language technological applications.

**"Estonian Hotline e-keelenõu (2013)"**, EKT39, a project funded by the Ministry of Education and Research within "National Programme for Estonian Language Technology" "Tools for lexical resources (2014–2017)", EKT61, a project funded by the Ministry of Education and Research within "National Programme for Estonian Language Technology"

**"Estonian Collocations Dictionary (2014-2018)"**, EKKM14-315, a project funded by the Ministry of Education and Research within national programme "Estonian language and cultural memory"

**"Estonian Basic Dictionary (2010-2013)"**, EKKM10-2014, a project funded by the Ministry of Education and Research within national programme "Estonian language and cultural memory"

### Description of infrastructure

The Institute possesses the world's biggest Archive of Estonian dialects and Finno-Ugric languages (EMSUKA). The main sub-collections are the archive of Estonian dialects and the archive of recordings of Estonian dialects and Finno-Ugric languages.

The Institute is a member of the consortium that forms the Centre of Estonian Language Resources (CELR). CELR performs as an organisational framework for coordinating and implementing the policies of Estonia as a member of CLARIN ERIC (Common Language Resources and Technology Infrastructure of the European Research Infrastructure Consortium). CELR is a member of META-NET, a Network of Excellence dedicated to fostering the technological foundations of a multilingual European information society. The Institute's contribution to the core infrastructure contains 58 objects, including various lexical resources, speech databases, speech processing tools, text corpora and text processing tools.

Lexicographic data:

- The Dictionary of Standard Estonian 2013
- The Explanatory Dictionary of the Estonian Language
- The Dictionary of Foreign Words
- The Estonian Etymological Dictionary
- The Basic Estonian Dictionary
- The Estonian-Russian Dictionary
- The Russian-Estonian Dictionary

## 17. Real Academia Española (RAE)

### Role in the project (cf. Consortium as a whole)

- Partner with lexicographic data and expertise

### Description of the legal entity

**The Royal Spanish Academy** (*Real Academia Española*, RAE) is the Spanish official institution responsible for overseeing the Spanish language since 1713. The RAE is affiliated with national language academies in twenty-two other Spanish-speaking nations through the Association of Spanish Language Academies (*Asociación de Academias de la Lengua Española*, ASALE). The RAE and ASALE promote a pan-hispanic linguistic unity and integrity within and among their respective territories to ensure a common language standard while acknowledging polycentric koinés. To achieve this goal, RAE regularly publishes dictionaries, grammars, orthographies and classical literary works. Free linguistic

advice is also provided. The RAE has been recently advising in the Spanish National Plan for the Advancement of Language Technology.<sup>3</sup>

The main priority is the [Diccionario de la lengua española](#) (Dictionary of Spanish Language, Spanish acronym DLE, formerly known as DRAE), edited periodically twenty-three times since 1780. Student and essential dictionaries have also been published by RAE. Other recurrent major works are the orthography series (last published in 2010-2013) and the grammar series (last published in 2009-2011). All these works are co-authored with the ASALE. Also, a long-time effort is made to publish a Historical Dictionary. Electronic resources at the RAE are freely accessible on the Internet (<http://www.rae.es>), where the [DLE](#) is searched more than 60 million times a month on average, peaking at 80 million in October 2016.

RAE works are based on empirical data, both as a huge collection of manual records started on XVIII<sup>th</sup> century and as language corpora with written and spoken (when available) samples from XI<sup>th</sup> century to-date.

### Short CVs of main team members

**Paz Battaner-Arias** (PhD) (F). Full member of Real Academia Española. She is a retired Professor of Spanish language at Universidad Pompeu Fabra (UPF), where she was chief researcher in several projects about language teaching, lexicography or corpus creation and dean of Translating and Interpreting School. She is a member of Euralex (European Association for Lexicography) and serves on the editorial board of scholarly publications. She is chief editor of a variety of dictionaries.

**Jordi Porta-Zamorano** (PhD) (M). Computer scientist working in the area of computational linguistics. His areas of expertise ranges from NLP applied to standard and nonstandard varieties, grammar checking, or search engines for corpora and dictionaries.

**José-Luis Sancho-Sánchez** (M). Computational linguist. With a lifetime career at the RAE, his experience ranges from lexical database population to Dictionary Writing Software development, with deeper hands on experience on reference corpus creation, annotation and exploitation and electronic lexicography.

**Rafael-J. Ureña-Ruiz** (M). Computational linguist. Also with a lifetime career at the RAE (except a year in RACEFyN developing a sci-tech corpus), his experience ranges from corpora annotation resources to multi-format digital publishing, both in backend and frontend tasks.

### Relevant publications

- Battaner P. (2012) «El léxico, como pilar inicial de la reflexión lingüística, y el diccionario». Ponencia plenaria en el V Congreso Internacional de Lexicografía Hispánica, Universidad Carlos III de Madrid, 26 de junio, 2012.
- Battaner P. (2014) «El diccionario como punto de encuentro en la descripción de la lengua española». Ponencia invitada en el Simposio Internacional El futuro de los diccionarios en la era digital, Real Academia Española, Madrid, 5, 6 y 7 de noviembre de 2014.

<sup>3</sup> <http://www.agendadigital.gob.es/tecnologias-lenguaje/Bibliotecaimpulsotecnologiaslenguaje/Detalle%20del%20Plan/Plan-Advancement-Language-Technology.pdf>

- Battaner P. & I. Renau (2011). «The Spanish Learner’s Dictionary DAELE on the Panorama of the Spanish E-Lexicography». En Kosem, I.; Kosem, K. (eds.). *Electronic Lexicography in the 21st Century, New Applications for New Users*. Proceedings of eLex 2011. Trojina: Institute for Applied Slovene Studies. 221-227.
- Battaner P. (2008) «El fenómeno de la polisemia en la lexicografía actual: otra perspectiva». *Revista de Lexicografía*, XIV.7-25.

**Previous projects/activities**

*Diccionario de la lengua española*: Rooted in the first dictionary ever published by RAE, DLe describes the Spanish general vocabulary while also signaling local or terminological specificities.

*Nuevo diccionario histórico del español*:\_NDHE aims at unveiling Spanish lexical evolution along time.

*Diccionario del español jurídico*:\_This Law-related dictionary is a joint product between RAE and the CGPJ (Spain's judiciary power governing body). Their definitions are concise, cross-referenced and lexicographically-constructed, but also include samples of real use in legal texts and law excerpts when necessary.

*Corpus del nuevo diccionario histórico del español*:\_CDH contains selected texts from all times, areas, genres and topics and serves as empirical evidence for the Historical Dictionary.

*Corpes XXI*:\_The 21st century Spanish Corpus is a collection of selected written and spoken texts from all genres, varieties and Spanish-speaking countries. It is designed to be fully balanced in order to serve as a reference corpus.

**Description of infrastructure**

Both development (corpus processing, indexing, dictionary storing,...) and operation environments (dictionary and corpus querying) are hosted in-house.

Lexicographic data:

- *Diccionario de la lengua española*, 22nd Ed. (2001)

**4.2 4.2. THIRD PARTIES INVOLVED IN THE PROJECT (INCLUDING USE OF THIRD PARTY RESOURCES)**

**Participant 17: Real Academia Española**

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	N
Does the participant envisage that part of its work is performed by <b>linked third parties</b> <i>RAE GESTIÓN, S.L. is a limited liability company with a sole proprietor, the Real Academia Española. RAE GESTIÓN, S.L. has a service contract with the Real Academia Española for providing technical support to all its activities and projects in areas of lexicography, corpus and technology through its multidisciplinary and qualified staff.</i>	Y
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	N

## 5 SECTION 5: ETHICS AND SECURITY

---

### 5.1 5.1 ETHICS

ELEXIS will respect the integrity of the involved individuals, making sure that no personal information will be exploited for making profit (although it will allow the development of commercial services driven by the end products of the project). In the case of test data, no personal data, related to user identity, will be stored locally or remotely, thus ensuring confidentiality.

The members of the ELEXIS consortium declare that the proposal conforms to current legislation and regulations in the countries where the research will be carried out, and will comply with obligations from ARTICLE 34 on ETHICS in Grant Agreement. Moreover, the proposal conforms to relevant EU legislation, such as:

- The latest version of the Helsinki Declaration
- The Charter of Fundamental Rights of the EU
- Directive 95/46/EC on protection of personal data, which describes a principle that personal data should not be processed at all, except when certain conditions are met. These conditions fall into three categories: transparency, legitimate purpose and proportionality.
- Directive 97/66/EC concerning the processing of personal data and the protection of privacy in the telecommunications sector.
- Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector.
- Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC.

The consortium commits that the approval by relevant authority/ies at national level (local ethics committee(s), data protection authorities) is ensured and reported on to the EC before the processing of data. Moreover, the consortium will comply with any future legislation and possible changes, especially under the light of the Commission's current reviewing of the general EU legal framework on the protection of personal data.

Any personal data collected will be processed in accordance with the GDPR (Regulation EU 2016/679) and the relevant national laws derived from its implementation. Participants in workshops, surveys, questionnaires, interviews and panels will be provided in writing with details of what personal information will be processed and whether or not their comments/contributions will be personally identifiable or only their institution will be identified.

The consortium partner K Dictionaries Ltd from Israel confirms that the ethical standards and guidelines of Horizon 2020 will be rigorously applied and will perform all research activities within the project with the highest research integrity and in accordance with the data privacy and other ethical standards and guidelines according to national legislation as well as EC Ethical Legislation.

### 5.2 5.2 SECURITY

**Please indicate if your project will involve:**

- activities or results raising security issues: NO
- 'EU-classified information' as background or results: NO



ESTIMATED BUDGET FOR THE ACTION (page 1 of 2)

Estimated eligible <sup>1</sup> costs (per budget category)										EU contribution			Additional information		
A. Direct personnel costs				B. Direct costs of subcontracting	C. Direct costs of fin. support	D. Other direct costs	E. Indirect costs <sup>2</sup>	F. Special unit costs	Total costs	Reimbursement rate %	Maximum EU contribution <sup>3</sup>	Maximum grant amount <sup>4</sup>	Information for indirect costs	Information for auditors	Other information:
A.1 Employees (or equivalent)		A.4 SME owners without salary				D.1 Travel	E. Indirect costs <sup>2</sup>	F. 1 "Costs for providing trans-national access to research infrastructure" **					Estimated costs of in-kind contributions not used on premises	Declaration of costs under Point D.4	Estimated costs of beneficiaries/linked third parties not receiving EU funding
A.2 Natural persons under direct contract		A.5 Beneficiaries that are natural persons without salary													
A.3 Seconded persons						D.2 Equipment									
[A.6 Personnel for providing access to research infrastructure]						D.3 Other goods and services									
						D.4 Costs of large research infrastructure									
Form of costs <sup>6</sup>	Actual	Unit <sup>7</sup>	Unit <sup>8</sup>		Actual	Actual	Actual	Flat-rate <sup>9</sup>							
	(a)	Total (b)	No hours	Total (c)	(d)	(e)	(f)	(g)=0,25x ((a)+(b)+(c)+(f)+[(h1)+(h2)]-(m))	Total (h1)	(i)=(a)+(b)+(c)+(d)+(e)+(f)+(g)+(h1)+(h2)+(h3)	(j)	(k)	(l)	(m)	Yes/No
1. JSI	503050.00	0.00	0	0.00	0.00	0.00	121395.00	156111.25		780556.25	100.00	780556.25	780556.25	0.00	No
2. LC	347100.00	0.00	0	0.00	0.00	0.00	93065.00	110041.25		550206.25	100.00	550206.25	550206.25	0.00	No
3. IVDNT	336000.00	0.00	0	0.00	0.00	0.00	40176.00	94044.00		470220.00	100.00	470220.00	470220.00	0.00	No
4. UNIROMA1	330000.00	0.00	0	0.00	0.00	0.00	33000.00	90750.00		453750.00	100.00	453750.00	453750.00	0.00	No
5. NUI GALWAY	344000.00	0.00	0	0.00	0.00	0.00	34400.00	94600.00		473000.00	100.00	473000.00	473000.00	0.00	No
6. OEAW	324500.00	0.00	0	0.00	0.00	0.00	38354.00	90713.50		453567.50	100.00	453567.50	453567.50	0.00	No
7. BCDH	211680.00	0.00	0	0.00	0.00	0.00	36648.00	62082.00		310410.00	100.00	310410.00	310410.00	0.00	No
8. MTANYTI	86800.00	0.00	0	0.00	0.00	0.00	17916.00	26179.00		130895.00	100.00	130895.00	130895.00	0.00	No
9. IBL	123444.00	0.00	0	0.00	0.00	0.00	23412.00	36714.00		183570.00	100.00	183570.00	183570.00	0.00	No
10. FCSH-UNL	63420.00	0.00	0	0.00	0.00	0.00	9513.00	18233.25		91166.25	100.00	91166.25	91166.25	0.00	No
11. K Dictionaries	145000.00	0.00	0	0.00	0.00	0.00	28374.00	43343.50		216717.50	100.00	216717.50	216717.50	0.00	No
12. CNR	105678.00	0.00	0	0.00	0.00	0.00	15851.00	30382.25		151911.25	100.00	151911.25	151911.25	0.00	No
13. DSL	161800.00	0.00	0	0.00	0.00	0.00	23856.00	46414.00		232070.00	100.00	232070.00	232070.00	0.00	No
14. UCPH	157500.00	0.00	0	0.00	0.00	0.00	23625.00	45281.25		226406.25	100.00	226406.25	226406.25	0.00	No
15. UT	92200.00	0.00	0	0.00	0.00	0.00	25539.00	29434.75		147173.75	100.00	147173.75	147173.75	0.00	No
16. EKI	44500.00	0.00	0	0.00	0.00	0.00	13058.00	14389.50		71947.50	100.00	71947.50	71947.50	0.00	No
17. RAE	5600.00	0.00	0	0.00	0.00	0.00	0.00	1400.00		7000.00	100.00	7000.00	7000.00	0.00	No
- RAEG <sup>14</sup>	27400.00	0.00	0	0.00	0.00	0.00	12120.00	9880.00		49400.00	100.00	49400.00	49400.00	0.00	No
Total beneficiary 17	33000.00	0.00			0.00	0.00	12120.00	11280.00		56400.00		56400.00	56400.00	0.00	
Total consortium	3409672.00	0.00		0.00	0.00	0.00	590302.00	999993.50		4999967.50		4999967.50	4999967.50	0.00	0.00



ESTIMATED BUDGET FOR THE ACTION (page 2 of 2)

- (1) See Article 6 for the eligibility conditions
- (2) The indirect costs covered by the operating grant (received under any EU or Euratom funding programme; see Article 6.5.(b)) are ineligible under the GA. Therefore, a beneficiary that receives an operating grant during the action's duration cannot declare indirect costs for the year(s)/reporting period(s) covered by the operating grant (see Article 6.2.E).
- (3) This is the theoretical amount of EU contribution that the system calculates automatically (by multiplying all the budgeted costs by the reimbursement rate). This theoretical amount is capped by the 'maximum grant amount' (that the Commission/Agency decided to grant for the action) (see Article 5.1).
- (4) The 'maximum grant amount' is the maximum grant amount decided by the Commission/Agency. It normally corresponds to the requested grant, but may be lower.
- (5) Depending on its type, this specific cost category will or will not cover indirect costs. Specific unit costs that include indirect costs are: costs for energy efficiency measures in buildings, access costs for providing trans-national access to research infrastructure and costs for clinical studies.
- (6) See Article 5 for the forms of costs
- (7) Unit : hours worked on the action; costs per unit (hourly rate) : calculated according to beneficiary's usual accounting practice
- (8) See Annex 2a 'Additional information on the estimated budget' for the details (costs per hour (hourly rate)).
- (9) Flat rate : 25% of eligible direct costs, from which are excluded: direct costs of subcontracting, costs of in-kind contributions not used on premises, direct costs of financial support, and unit costs declared under budget category F if they include indirect costs
- (10) See Annex 2a 'Additional information on the estimated budget' for the details (units, costs per unit).
- (11) See Annex 2a 'Additional information on the estimated budget' for the details (units, costs per unit, estimated number of units, etc)
- (12) Only specific unit costs that do not include indirect costs
- (13) See Article 9 for beneficiaries not receiving EU funding
- (14) Only for linked third parties that receive EU funding

## ANNEX 2a

### ADDITIONAL INFORMATION ON THE ESTIMATED BUDGET

#### Unit cost for SME owners/natural beneficiaries without salary

#### 1. Costs for a /SME owner//beneficiary that is a natural person/ not receiving a salary

Units: hours worked on the action

Amount per unit ('hourly rate'): calculated according to the following formula:

{ { EUR 4,650 / 143 hours }  
multiplied by  
{country-specific correction coefficient of the country where the beneficiary is established }

Country-specific correction coefficient (in force at the time of the call):

#### EU Member States

country	coefficient	country	coefficient	country	coefficient	country	coefficient	country	coefficient
AT	104.8%	DK	135.3%	HR	97.5%	LV	75.9%	SE	111.7%
BE	100.0%	EE	78.3%	HU	76.2%	MT	89.6%	SI	86.1%
BG	71.5%	EL	92.7%	IE	113.5%	NL	104.3%	SK	82.6%
CY	91.8%	ES	97.6%	IT	106.7%	PL	76.4%	UK	120.3%
CZ	83.8%	FI	116.6%	LT	73.1%	PT	89.1%		
DE	98.8%	FR	111.0%	LU	100.0%	RO	68.3%		

#### H2020 associated countries

country	coefficient	country	coefficient	country	coefficient	country	coefficient	country	coefficient
AL	76.1%	FO	134.1%	LI	110.0%	MK	68.4%	TR	86.6%
BA	73.6%	IL	108.7%	MD	61.1%	NO	131.9%		
CH	113.1%	IS	116.6%	ME	66.9%	RS	67.1%		

#### Other countries

country	coefficient	country	coefficient	country	coefficient	country	coefficient	country	coefficient
AM	89.9%	CU	83.8%	JP	115.9%	NI	57.3%	TJ	64.9%
AO	114.6%	CV	76.4%	KE	78.1%	NP	73.5%	TL	78.3%
AR	58.5%	DJ	93.4%	KG	83.1%	NZ	94.1%	TN	70.5%
AU	105.0%	DO	66.9%	KH	70.5%	PA	57.0%	TO	85.0%
AZ	93.0%	DZ	81.7%	KR	105.2%	PE	75.5%	TT	74.1%
BB	116.6%	EC	68.8%	KZ	100.2%	PG	83.0%	TW	83.6%
BD	47.2%	EG	48.6%	LA	77.7%	PH	65.8%	TZ	65.2%
BF	93.8%	ER	61.2%	LB	86.4%	PK	49.4%	UA	92.3%
BJ	92.6%	ET	85.2%	LK	61.6%	PS	100.4%	UG	65.7%
BM	151.5%	FJ	68.1%	LR	100.1%	PY	71.9%	US	99.4%
BO	51.3%	GA	113.1%	LS	56.7%	RU	115.5%	UY	75.3%
BR	92.0%	GE	89.5%	LY	60.0%	RW	87.3%	UZ	51.4%
BW	55.3%	GH	68.2%	MA	83.5%	SA	84.8%	VE	70.0%
BY	65.0%	GM	67.7%	MG	80.0%	SB	93.3%	VN	51.1%
BZ	75.3%	GN	60.4%	ML	90.4%	SD	65.1%	VU	112.6%
CA	86.4%	GT	78.8%	MR	64.5%	SG	102.5%	WS	75.8%
CD	127.6%	GW	102.7%	MU	72.7%	SL	85.2%	XK	58.6%
CF	114.3%	GY	58.9%	MW	76.0%	SN	86.2%	YE	68.1%
CG	124.9%	HK	93.8%	MX	70.4%	SR	50.6%	ZA	55.8%
CI	102.0%	HN	69.0%	MY	71.6%	SV	74.3%	ZM	66.4%

H2020 Model Grant Agreements: H2020 General MGA — Multi: v3.0 – dd.mm.2016

CL	67.1%	HT	108.7%	MZ	71.6%	SY	74.8%	ZW	47.2%
CM	103.3%	ID	75.3%	NA	68.3%	SZ	56.8%		
CN	85.0%	IN	52.8%	NC	128.9%	TD	125.3%		
CO	76.6%	JM	94.9%	NE	87.9%	TG	88.7%		
CR	76.7%	JO	75.5%	NG	92.4%	TH	65.0%		

*[additional OPTION for beneficiaries/linked third parties that have opted to use the unit cost (in the proposal/with an amendment):* For the following beneficiaries/linked third parties, the amounts per unit (hourly rate) are fixed as follows:

- Beneficiary/linked third party [short name]: EUR [insert amount]
  - Beneficiary/linked third party [short name]: EUR [insert amount]
- [same for other beneficiaries/linked third parties, if necessary] ]

Estimated number of units: see Annex 2

### **Energy efficiency measures unit cost**

*[OPTION if specific unit cost applicable to the grant: 2. Costs for energy efficiency measures in buildings*

Unit: m<sup>2</sup> of eligible ‘conditioned’ (i.e. built or refurbished) floor area

Amount per unit\*: see (for each beneficiary/linked third party and BEST table) the ‘unit cost table’ attached

\* Amount calculated as follows:  
{EUR 0.1 x estimated total kWh saved per m<sup>2</sup> per year x 10}

Estimated number of units: see (for each beneficiary/linked third party and BEST table) the ‘unit cost table’ attached

Unit cost table (energy efficiency measures unit cost)<sup>1</sup>

Short name beneficiary/linked third party	BEST No	Cost Amount per unit	Estimated No of units	Total unit cost (cost per unit x estimated no of units)

]

### **Research infrastructure unit cost**

*[OPTION if specific unit cost applicable to the grant: 3. Access costs for providing trans-national access to research infrastructure*

Units<sup>2</sup>: see (for each access provider and installation) the ‘unit cost table’ attached

<sup>1</sup> Data from the ‘building energy specification table (BEST)’ that is part of the proposal and Annex 1.

H2020 Model Grant Agreements: H2020 General MGA — Multi: v3.0 – dd.mm.2016

Amount per unit\*: see (for each access provider and installation) the ‘unit cost table’ attached

\* Amount calculated as follows:

$$\frac{\text{average annual total access cost to the installation (over past two years}^3)}{\text{average annual total quantity of access to the installation (over past two years}^4)}$$

Estimated number of units: see (for each access provider and installation) the ‘unit cost table’ attached

Unit cost table (access to research infrastructure unit cost)<sup>5</sup>

Short name access provider	Short name infrastructure	Installation		Unit of access	Amount per unit	Estimated No of units	Total unit cost (cost per unit x estimated no of units)
		No	Short name				

]

### Clinical studies unit cost

#### *[OPTION if specific unit cost is applicable to the grant: 4. Costs for clinical studies*

Units: patients/subjects that participate in the clinical study

Amount per unit\*: see (for each clinical study and beneficiary/linked third party) the ‘unit cost table’ attached

Estimated number of units: see (for each clinical study and beneficiary/linked third party) the ‘unit cost table’ attached

\* Amount calculated, for each task described in the protocol, as follows:

{Task 1

- {unit cost component ‘personnel costs’
- + unit cost component ‘costs of consumables’
- + unit cost component ‘costs of medical equipment’
- + unit cost component ‘costs of other specific services’
- + unit cost component ‘indirect costs’}

+ Task 2

- {unit cost component ‘personnel costs’
- + unit cost component ‘costs of consumables’
- + unit cost component ‘costs of medical equipment’
- + unit cost component ‘costs of other specific services’

<sup>2</sup> Unit of access (e.g. beam hours, weeks of access, sample analysis) fixed by the access provider in proposal.

<sup>3</sup> In exceptional and duly justified cases, the Commission/Agency may agree to a different reference period.

<sup>4</sup> In exceptional and duly justified cases, the Commission/Agency may agree to a different reference period.

<sup>5</sup> Data from the ‘table on estimated costs/quantity of access to be provided’ that is part of the proposal and Annex 1.

H2020 Model Grant Agreements: H2020 General MGA — Multi: v3.0 – dd.mm.2016

+ unit cost component ‘indirect costs’}  
[same for all other tasks]}

Unit cost components calculated as follows:

Unit cost component ‘**personnel costs**’ (i.e. ‘personnel costs of doctors’ + ‘personnel costs of other medical personnel’ + ‘personnel costs of technical personnel’)

For unit cost component ‘personnel costs of doctors’:

{ ‘average hourly cost for doctors’, i.e.:  
$$\frac{\text{certified or auditable total personnel costs for doctors for year N-1}}{\{1720 * \text{number of full-time equivalent for the personnel category doctors for year N-1}\}}$$
  
multiplied by  
estimated number of hours worked by doctors for the task (per patient/subject) }

For unit cost component ‘personnel costs of other medical personnel’:

{ ‘average hourly cost for other medical personnel’, i.e.:  
$$\frac{\text{certified or auditable total personnel costs for other medical personnel for year N-1}}{\{1720 * \text{number of full-time equivalent for the personnel category other medical personnel for year N-1}\}}$$
  
multiplied by  
estimated number of hours worked by other medical personnel for the task (per patient/subject) }

For unit cost component ‘personnel costs of technical personnel’:

{ ‘average hourly cost for technical personnel, i.e.:  
$$\frac{\text{certified or auditable total personnel costs for technical personnel for year N-1}}{\{1720 * \text{number of full-time equivalent for the personnel category technical personnel for year N-1}\}}$$
  
multiplied by  
estimated number of hours worked by technical personnel for the task (per patient/subject) }

‘total personnel costs’ means actual salaries + actual social security contributions + actual taxes and other costs included in the remuneration, provided they arise from national law or the employment contract or equivalent appointing act

Unit cost component ‘**costs of consumables**’ (i.e. ‘costs of consumables category 1’ + ‘costs of consumables category 2’ + ‘costs of consumables category 3’, etc)

For each category of consumables:

{ ‘average price per item’, i.e.:  
$$\frac{\text{certified or auditable total costs of purchase of the consumables in year N-1 for the category of consumables concerned}}{\text{total number of items purchased in year N-1 for the category of consumables concerned}}$$
  
multiplied by  
estimated number of items used for the task (per patient/subject) }

‘total costs of purchase of the consumables’ means total value of the supply contracts (including related duties, taxes and charges such as non-deductible VAT) concluded by the beneficiary for consumables delivered in year N-1, provided the contracts were awarded according to the principle of best value-for-money and without any conflict of interests

Unit cost component ‘**costs of medical equipment**’ (i.e. ‘costs of medical equipment category 1’ + ‘costs of medical equipment category 2’ + ‘costs of medical equipment category 3’, etc.)

For each category of medical equipment:

{ ‘average cost of depreciation and directly related services per unit of use’, i.e.:

H2020 Model Grant Agreements: H2020 General MGA — Multi: v3.0 – dd.mm.2016

{certified or auditable total depreciation costs in year N-1 for the category of equipment concerned + certified or auditable total costs of purchase of services in year N-1 for the category of equipment concerned}  
 -----  
 total capacity in year N-1  
 multiplied by  
 estimated number of units of use of the equipment for the task (per patient/subject)}

‘total depreciation costs’ means total depreciation allowances as recorded in the beneficiary’s accounts of year N-1 for the category of equipment concerned, provided the equipment was purchased according to the principle of best value-for-money and without any conflict of interests + total costs of renting or leasing contracts (including related duties, taxes and charges such as non-deductible VAT) in year N-1 for the category of equipment concerned, provided they do not exceed the depreciation costs of similar equipment and do not include finance fees

Unit cost component ‘costs of other specific services’ (i.e. ‘costs of contracts for specific service 1’ + ‘costs of contracts for specific service 2’ + ‘costs of contracts for specific service 3’, etc.)

For each category of specific service:

‘average cost of a specific service per patient or subject’, i.e.:  
 -----  
 certified or auditable total costs of purchase of a service in year N-1 for the category of specific services necessary for the conduct of clinical studies  
 -----  
 total number of patients or subjects included in the clinical studies for which the specific service was delivered in year N-1

‘total costs of purchase of a service’ means total value of the contracts concluded by the beneficiary (including related duties, taxes and charges such as non-deductible VAT) for the specific service delivered in year N-1 for the conduct of clinical studies, provided the contracts were awarded according to the principle of best value-for-money and without any conflict of interests

Unit cost component ‘indirect costs’

{ 25%  
 multiplied by  
 {unit cost component ‘personnel costs’ + unit cost component ‘costs of consumables’ + unit cost component ‘costs of medical equipment’} }

The following must be excluded:

- costs of in-kind contributions provided by third parties which are not used on the beneficiary’s premises and
- costs of providing financial support to third parties (if any).

Unit cost table: clinical studies unit cost<sup>6</sup>

[Insert name of clinical study]							
Tasks and unit cost components	Resources per patient	Amount per unit for beneficiary /linked third party	Amount per unit for beneficiary /linked third party	Amount per unit for beneficiary/linked third party 3 [insert short name]	...		

<sup>6</sup> Same table as in proposal and Annex 1.

H2020 Model Grant Agreements: H2020 General MGA — Multi: v3.0 – dd.mm.2016

			1 [insert short name]	2 [insert short name]		in-kind contributions by third party*	
<b>Task No. 1 Blood sample</b>							
<b>Personnel costs</b>	doctors	----	0	0	0	0	
	other medical personnel	Phlebotomy (nurse), 10 minutes	8,33 EUR	11,59 EUR	10,55 EUR	9,76 EUR	
	technical personnel	Sample Processing (lab technician), 15 minutes	9,51 EUR	15,68 EUR	13,77 EUR	12,35 EUR	
<b>Costs of consumables</b>	Category 1	Syringe, 1	XX EUR	XX EUR	XX EUR	XX EUR	
	Category 2	Cannula, 1	XX EUR	XX EUR	XX EUR	XX EUR	
	Category 2	Blood container, 1	XX EUR	XX EUR	XX EUR	XX EUR	
	...						
<b>Costs of medical equipment</b>	Category 1	Use of -80° deep freezer, 60 days	XX EUR	XX EUR	XX EUR	XX EUR	
	Category 2	Use of centrifuge, 15 minutes	XX EUR	XX EUR	XX EUR	XX EUR	
	....						
<b>Costs of other specific services</b>	Category 1						
	Category 2						
	...						
<b>Indirect costs</b>							
<b>Task No. 2</b>							
...							
<b>Total amount per unit</b>			XX EUR	XX EUR	XX EUR	XX EUR**	
<b>Estimated No of units</b> (patients/subjects participating in the study)			XX	XX	XX	XX	
<b>Total unit cost for beneficiary/linked third party</b> (total cost per unit x estimated no of units)			XX EUR	XX EUR	XX EUR		

\* Use costs of third party making in-kind contribution.

\*\* Capped at payment to third party, if any.

## **ACCESSION FORM FOR BENEFICIARIES**

**LEXICAL COMPUTING CZ SRO (LC)**, established in BOTANICKA 554/68A, BRNO 602 00, Czech Republic, VAT number: CZ29295491, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

**hereby agrees**

**to become beneficiary** No ('2')

**in Grant Agreement No** 731015 ('the Agreement')

**between** INSTITUT JOZEF STEFAN **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

**for the action entitled** 'European Lexicographic Infrastructure (ELEXIS)'.

**and mandates**

**the coordinator** to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary



## **ACCESSION FORM FOR BENEFICIARIES**

**STICHTING INSTITUUT VOOR DE NEDERLANDSE TAAL (IVDNT)**, established in MATTHIAS DE VRIESHOFF 2-3, LEIDEN 2311 BZ, Netherlands, VAT number: NL002818267B01, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

**hereby agrees**

**to become beneficiary No ('3')**

**in Grant Agreement No 731015 ('the Agreement')**

**between INSTITUT JOZEF STEFAN and the European Union ('the EU')**, represented by the European Commission ('the Commission'),

**for the action entitled 'European Lexicographic Infrastructure (ELEXIS)'**.

**and mandates**

**the coordinator** to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

## **ACCESSION FORM FOR BENEFICIARIES**

**UNIVERSITA DEGLI STUDI DI ROMA LA SAPIENZA (UNIROMA1)**, established in Piazzale Aldo Moro 5, ROMA 00185, Italy, VAT number: IT02133771002, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

**hereby agrees**

**to become beneficiary** No ('4')

**in Grant Agreement No** 731015 ('the Agreement')

**between** INSTITUT JOZEF STEFAN **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

**for the action entitled** 'European Lexicographic Infrastructure (ELEXIS)'.

**and mandates**

**the coordinator** to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

## **ACCESSION FORM FOR BENEFICIARIES**

**NATIONAL UNIVERSITY OF IRELAND GALWAY (NUI GALWAY)**, established in UNIVERSITY ROAD, GALWAY, Ireland, VAT number: IE0022578J, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

**hereby agrees**

**to become beneficiary** No ('5')

**in Grant Agreement No** 731015 ('the Agreement')

**between** INSTITUT JOZEF STEFAN **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

**for the action entitled** 'European Lexicographic Infrastructure (ELEXIS)'.

**and mandates**

**the coordinator** to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

## **ACCESSION FORM FOR BENEFICIARIES**

**OESTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN (OEAW)**, established in DR. IGNAZ SEIPEL-PLATZ 2, WIEN 1010, Austria, VAT number: ATU37116303, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

**hereby agrees**

**to become beneficiary** No ('6')

**in Grant Agreement No** 731015 ('the Agreement')

**between** INSTITUT JOZEF STEFAN **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

**for the action entitled** 'European Lexicographic Infrastructure (ELEXIS)'.

**and mandates**

**the coordinator** to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

## **ACCESSION FORM FOR BENEFICIARIES**

**CENTAR ZA DIGITALNE HUMANISTICKE NAUKE (BCDH)**, established in DZORDZA VASINGTONA 28A, BEOGRAD 11000, Serbia, VAT number: RS105604641, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

**hereby agrees**

**to become beneficiary** No ('7')

**in Grant Agreement No** 731015 ('the Agreement')

**between** INSTITUT JOZEF STEFAN **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

**for the action entitled** 'European Lexicographic Infrastructure (ELEXIS)'.

**and mandates**

**the coordinator** to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

## **ACCESSION FORM FOR BENEFICIARIES**

**MAGYAR TUDOMANYOS AKADEMIA, NYELVTUDOMANYI INTEZET (MTANYTI)**, established in BENCZUR UTCA 33, BUDAPEST 1068, Hungary, VAT number: HU15300571, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

**hereby agrees**

**to become beneficiary** No ('8')

**in Grant Agreement No** 731015 ('the Agreement')

**between** INSTITUT JOZEF STEFAN **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

**for the action entitled** 'European Lexicographic Infrastructure (ELEXIS)'.

**and mandates**

**the coordinator** to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

## **ACCESSION FORM FOR BENEFICIARIES**

**INSTITUTE FOR BULGARIAN LANGUAGE PROF LYUBOMIR ANDREYCHIN (IBL)**, established in UL. SHIPCHENSKI PROHOD 52 BL 17, Sofia 1113, Bulgaria, VAT number: BG000665498, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

**hereby agrees**

**to become beneficiary** No ('9')

**in Grant Agreement No** 731015 ('the Agreement')

**between** INSTITUT JOZEF STEFAN **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

**for the action entitled** 'European Lexicographic Infrastructure (ELEXIS)'.

**and mandates**

**the coordinator** to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

## **ACCESSION FORM FOR BENEFICIARIES**

**UNIVERSIDADE NOVA DE LISBOA (FCSH-UNL)**, established in CAMPUS DE CAMPOLIDE, LISBOA 1099 085, Portugal, VAT number: PT501559094, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

**hereby agrees**

**to become beneficiary** No ('10')

**in Grant Agreement No** 731015 ('the Agreement')

**between** INSTITUT JOZEF STEFAN **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

**for the action entitled** 'European Lexicographic Infrastructure (ELEXIS)'.

**and mandates**

**the coordinator** to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary



## **ACCESSION FORM FOR BENEFICIARIES**

**K DICTIONARIES LTD (K Dictionaries)**, established in 8 NAHUM HANAVI STREET, TEL AVIV 6350310, Israel, VAT number: IL511783672, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

**hereby agrees**

**to become beneficiary** No ('11')

**in Grant Agreement No** 731015 ('the Agreement')

**between** INSTITUT JOZEF STEFAN **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

**for the action entitled** 'European Lexicographic Infrastructure (ELEXIS)'.

**and mandates**

**the coordinator** to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

## **ACCESSION FORM FOR BENEFICIARIES**

**CONSIGLIO NAZIONALE DELLE RICERCHE (CNR)**, established in PIAZZALE ALDO MORO 7, ROMA 00185, Italy, VAT number: IT02118311006, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

**hereby agrees**

**to become beneficiary** No ('12')

**in Grant Agreement No** 731015 ('the Agreement')

**between** INSTITUT JOZEF STEFAN **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

**for the action entitled** 'European Lexicographic Infrastructure (ELEXIS)'.

**and mandates**

**the coordinator** to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

## **ACCESSION FORM FOR BENEFICIARIES**

**DET DANSKE SPROG- OG LITTERATURELSKAB (DSL)**, established in CHRISTIANS BRYGGE 1, KOBENHAVN 1219, Denmark, VAT number: DK59188917, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

**hereby agrees**

**to become beneficiary** No ('13')

**in Grant Agreement No** 731015 ('the Agreement')

**between** INSTITUT JOZEF STEFAN **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

**for the action entitled** 'European Lexicographic Infrastructure (ELEXIS)'.

**and mandates**

**the coordinator** to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

## **ACCESSION FORM FOR BENEFICIARIES**

**KOBENHAVNS UNIVERSITET (UCPH)**, established in NORREGADE 10, KOBENHAVN 1165, Denmark, VAT number: DK29979812, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

**hereby agrees**

**to become beneficiary** No ('14')

**in Grant Agreement No** 731015 ('the Agreement')

**between** INSTITUT JOZEF STEFAN **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

**for the action entitled** 'European Lexicographic Infrastructure (ELEXIS)'.

**and mandates**

**the coordinator** to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

## **ACCESSION FORM FOR BENEFICIARIES**

**UNIVERSITAT TRIER (UT)**, established in UNIVERSITATSRING 15, TRIER 54296, Germany, VAT number: DE149881695, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

**hereby agrees**

**to become beneficiary** No ('15')

**in Grant Agreement No** 731015 ('the Agreement')

**between** INSTITUT JOZEF STEFAN **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

**for the action entitled** 'European Lexicographic Infrastructure (ELEXIS)'.

**and mandates**

**the coordinator** to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

## **ACCESSION FORM FOR BENEFICIARIES**

**EESTI KEELE INSTITUUT (EKI)**, established in ROOSIKRANTSI 6, TALLINN 10119, Estonia, VAT number: EE100252305, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

**hereby agrees**

**to become beneficiary** No ('16')

**in Grant Agreement No** 731015 ('the Agreement')

**between** INSTITUT JOZEF STEFAN **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

**for the action entitled** 'European Lexicographic Infrastructure (ELEXIS)'.

**and mandates**

**the coordinator** to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

## **ACCESSION FORM FOR BENEFICIARIES**

**REAL ACADEMIA ESPANOLA (RAE)**, established in CALLE FELIPE IV 4, MADRID 28014, Spain, VAT number: ESQ2868010F, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

**hereby agrees**

**to become beneficiary** No ('17')

**in Grant Agreement No** 731015 ('the Agreement')

**between** INSTITUT JOZEF STEFAN **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

**for the action entitled** 'European Lexicographic Infrastructure (ELEXIS)'.

**and mandates**

**the coordinator** to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

print  
format A4  
landscape

MODEL ANNEX 4 FOR H2020 GENERAL MGA — MULTI

FINANCIAL STATEMENT FOR [BENEFICIARY [name]/ LINKED THIRD PARTY [name]] FOR REPORTING PERIOD [reporting period]

Eligible <sup>1</sup> costs (per budget category)													Receipts	EU contribution			Additional information	
A. Direct personnel costs		B. Direct costs of subcontracting		[C. Direct costs of fin. support]		D. Other direct costs		E. Indirect costs <sup>2</sup>		[F. Costs of ... ]		Total costs	Receipts	Reimbursement rate %	Maximum EU contribution <sup>3</sup>	Requested EU contribution		
A.1 Employees (or equivalent)		A.4 SME owners without salary		[C.1 Financial support]		D.1 Travel		[D.4 Costs of large research infrastructure]		[F.1 Costs of ...]		Receipts of the action, to be reported in the last reporting period, according to Article 5.3.3						
A.2 Natural persons under direct contract		A.5 Beneficiaries that are natural persons without salary		[C.2 Prizes]		D.2 Equipment												
A.3 Seconded persons [A.6 Personnel for providing access]						D.3 Other goods and												
Form of costs <sup>4</sup>	Actual	Unit		Actual	Actual	Actual	Actual	Flat-rate <sup>5</sup>	[Unit][Lump sum]		Unit	Receipts of the action, to be reported in the last reporting period, according to Article 5.3.3						
	a	Total b	No hours	Total c	d	[e]	f	[g]	h=0,25 x (a+b+c+f+g)	No units	Total [i1]					Total [i2]	j=a+b+c+d+h+e+f+g	k
[short name beneficiary /linked third party]																		

**The beneficiary/linked third party hereby confirms that:**  
 The information provided is complete, reliable and true.  
 The costs declared are eligible (see Article 6).  
 The costs can be substantiated by adequate records and supporting documentation that will be produced upon request or in the context of checks, reviews, audits and investigations (see Articles 17, 18 and 22).  
 For the last reporting period: that all the receipts have been declared (see Article 5.3.3).

Please declare all eligible costs, even if they exceed the amounts indicated in the estimated budget (see Annex 2). Only amounts that were declared in your individual financial statements can be taken into account later on, in order to replace other costs that are found to be ineligible.

<sup>1</sup> See Article 6 for the eligibility conditions

<sup>2</sup> The indirect costs claimed must be free of any amounts covered by an operating grant (received under any EU or Euratom funding programme; see Article 6.2.E). If you have received an operating grant during this reporting period, you cannot claim any indirect costs.

<sup>3</sup> This is the *theoretical* amount of EU contribution that the system calculates automatically (by multiplying the reimbursement rate by the total costs declared). The amount you request (in the column 'requested EU contribution') may be less

<sup>4</sup> See Article 5 for the form of costs

<sup>5</sup> Flat rate : 25% of eligible direct costs, from which are excluded: direct costs of subcontracting, costs of in-kind contributions not used on premises, direct costs of financial support, and unit costs declared under budget category F if they include indirect costs (see Article 6.2.E)

<sup>6</sup> Only specific unit costs that do not include indirect costs



## ANNEX 5

### MODEL FOR THE CERTIFICATE ON THE FINANCIAL STATEMENTS

- For options [*in italics in square brackets*]: choose the applicable option. Options not chosen should be deleted.
- For fields in [grey in square brackets]: enter the appropriate data

#### TABLE OF CONTENTS

TERMS OF REFERENCE FOR AN INDEPENDENT REPORT OF FACTUAL FINDINGS ON COSTS DECLARED UNDER A GRANT AGREEMENT FINANCED UNDER THE HORIZON 2020 RESEARCH FRAMEWORK PROGRAMME

INDEPENDENT REPORT OF FACTUAL FINDINGS ON COSTS DECLARED UNDER A GRANT AGREEMENT FINANCED UNDER THE HORIZON 2020 RESEARCH FRAMEWORK PROGRAMME

## Terms of Reference for an Independent Report of Factual Findings on costs declared under a Grant Agreement financed under the Horizon 2020 Research and Innovation Framework Programme

This document sets out the ‘Terms of Reference (ToR)’ under which

[OPTION 1: [insert name of the beneficiary] (‘the Beneficiary’)] [OPTION 2: [insert name of the linked third party] (‘the Linked Third Party’), third party linked to the Beneficiary [insert name of the beneficiary] (‘the Beneficiary’)]

agrees to engage

[insert legal name of the auditor] (‘the Auditor’)

to produce an independent report of factual findings (‘the Report’) concerning the Financial Statement(s)<sup>1</sup> drawn up by the [Beneficiary] [Linked Third Party] for the Horizon 2020 grant agreement [insert number of the grant agreement, title of the action, acronym and duration from/to] (‘the Agreement’), and

to issue a Certificate on the Financial Statements’ (‘CFS’) referred to in Article 20.4 of the Agreement based on the compulsory reporting template stipulated by the Commission.

The Agreement has been concluded under the Horizon 2020 Research and Innovation Framework Programme (H2020) between the Beneficiary and [OPTION 1: the European Union, represented by the European Commission (‘the Commission’)] [OPTION 2: the European Atomic Energy Community (Euratom,) represented by the European Commission (‘the Commission’)] [OPTION 3: the [Research Executive Agency (REA)] [European Research Council Executive Agency (ERCEA)] [Innovation and Networks Executive Agency (INEA)] [Executive Agency for Small and Medium-sized Enterprises (EASME)] (‘the Agency’), under the powers delegated by the European Commission (‘the Commission’).]

The [Commission] [Agency] is mentioned as a signatory of the Agreement with the Beneficiary only. The [European Union][Euratom][Agency] is not a party to this engagement.

### 1.1 Subject of the engagement

The coordinator must submit to the [Commission][Agency] the final report within 60 days following the end of the last reporting period which should include, amongst other documents, a CFS for each beneficiary and for each linked third party that requests a total contribution of EUR 325 000 or more, as reimbursement of actual costs and unit costs calculated on the basis of its usual cost accounting practices (see Article 20.4 of the Agreement). The CFS must cover all reporting periods of the beneficiary or linked third party indicated above.

The Beneficiary must submit to the coordinator the CFS for itself and for its linked third party(ies), if the CFS must be included in the final report according to Article 20.4 of the Agreement..

The CFS is composed of two separate documents:

- The Terms of Reference (‘the ToR’) to be signed by the [Beneficiary] [Linked Third Party] and the Auditor;

---

<sup>1</sup> By which costs under the Agreement are declared (see template ‘Model Financial Statements’ in Annex 4 to the Grant Agreement).

- The Auditor's Independent Report of Factual Findings ('the Report') to be issued on the Auditor's letterhead, dated, stamped and signed by the Auditor (or the competent public officer) which includes the agreed-upon procedures ('the Procedures') to be performed by the Auditor, and the standard factual findings ('the Findings') to be confirmed by the Auditor.

If the CFS must be included in the final report according to Article 20.4 of the Agreement, the request for payment of the balance relating to the Agreement cannot be made without the CFS. However, the payment for reimbursement of costs covered by the CFS does not preclude the [Commission,] [Agency,] the European Anti-Fraud Office and the European Court of Auditors from carrying out checks, reviews, audits and investigations in accordance with Article 22 of the Agreement.

## 1.2 Responsibilities

The [Beneficiary] [Linked Third Party]:

- must draw up the Financial Statement(s) for the action financed by the Agreement in compliance with the obligations under the Agreement. The Financial Statement(s) must be drawn up according to the [Beneficiary's] [Linked Third Party's] accounting and book-keeping system and the underlying accounts and records;
- must send the Financial Statement(s) to the Auditor;
- is responsible and liable for the accuracy of the Financial Statement(s);
- is responsible for the completeness and accuracy of the information provided to enable the Auditor to carry out the Procedures. It must provide the Auditor with a written representation letter supporting these statements. The written representation letter must state the period covered by the statements and must be dated;
- accepts that the Auditor cannot carry out the Procedures unless it is given full access to the [Beneficiary's] [Linked Third Party's] staff and accounting as well as any other relevant records and documentation.

The Auditor:

- [Option 1 by default: is qualified to carry out statutory audits of accounting documents in accordance with Directive 2006/43/EC of the European Parliament and of the Council of 17 May 2006 on statutory audits of annual accounts and consolidated accounts, amending Council Directives 78/660/EEC and 83/349/EEC and repealing Council Directive 84/253/EEC or similar national regulations].
- [Option 2 if the Beneficiary or Linked Third Party has an independent Public Officer: is a competent and independent Public Officer for which the relevant national authorities have established the legal capacity to audit the Beneficiary].
- [Option 3 if the Beneficiary or Linked Third Party is an international organisation: is an [internal] [external] auditor in accordance with the internal financial regulations and procedures of the international organisation].

The Auditor:

- must be independent from the Beneficiary [and the Linked Third Party], in particular, it must not have been involved in preparing the [Beneficiary's] [Linked Third Party's] Financial Statement(s);
- must plan work so that the Procedures may be carried out and the Findings may be assessed;
- must adhere to the Procedures laid down and the compulsory report format;
- must carry out the engagement in accordance with this ToR;
- must document matters which are important to support the Report;
- must base its Report on the evidence gathered;
- must submit the Report to the [Beneficiary] [Linked Third Party].

The Commission sets out the Procedures to be carried out by the Auditor. The Auditor is not responsible for their suitability or pertinence. As this engagement is not an assurance engagement, the Auditor does not provide an audit opinion or a statement of assurance.

### 1.3 Applicable Standards

The Auditor must comply with these Terms of Reference and with<sup>2</sup>:

- the International Standard on Related Services ('ISRS') 4400 *Engagements to perform Agreed-upon Procedures regarding Financial Information* as issued by the International Auditing and Assurance Standards Board (IAASB);
- the *Code of Ethics for Professional Accountants* issued by the International Ethics Standards Board for Accountants (IESBA). Although ISRS 4400 states that independence is not a requirement for engagements to carry out agreed-upon procedures, the [Commission][Agency] requires that the Auditor also complies with the Code's independence requirements.

The Auditor's Report must state that there is no conflict of interests in establishing this Report between the Auditor and the Beneficiary [and the Linked Third Party], and must specify - if the service is invoiced - the total fee paid to the Auditor for providing the Report.

### 1.4 Reporting

The Report must be written in the language of the Agreement (see Article 20.7).

Under Article 22 of the Agreement, the Commission[, the Agency], the European Anti-Fraud Office and the Court of Auditors have the right to audit any work that is carried out under the action and for which costs are declared from [the European Union] [Euratom] budget. This includes work related to this engagement. The Auditor must provide access to all working papers (e.g. recalculation of hourly rates, verification of the time declared for the action) related to this assignment if the Commission [, the Agency], the European Anti-Fraud Office or the European Court of Auditors requests them.

### 1.5 Timing

The Report must be provided by [dd Month yyyy].

### 1.6 Other terms

[The [Beneficiary] [Linked Third Party] and the Auditor can use this section to agree other specific terms, such as the Auditor's fees, liability, applicable law, etc. Those specific terms must not contradict the terms specified above.]

[legal name of the Auditor]	[legal name of the [Beneficiary][Linked Third Party]]
[name & function of authorised representative]	[name & function of authorised representative]
[dd Month yyyy]	[dd Month yyyy]
Signature of the Auditor	Signature of the [Beneficiary][Linked Third Party]

<sup>2</sup> Supreme Audit Institutions applying INTOSAI-standards may carry out the Procedures according to the corresponding International Standards of Supreme Audit Institutions and code of ethics issued by INTOSAI instead of the International Standard on Related Services ('ISRS') 4400 and the Code of Ethics for Professional Accountants issued by the IAASB and the IESBA.

**Independent Report of Factual Findings on costs declared  
under Horizon 2020 Research and Innovation Framework Programme**

*(To be printed on the Auditor's letterhead)*

To  
[ name of contact person(s)], [Position]  
[ [Beneficiary's] [Linked Third Party's] name ]  
[ Address]  
[ dd Month yyyy]

Dear [Name of contact person(s)],

As agreed under the terms of reference dated [dd Month yyyy]

with [OPTION 1: [insert name of the beneficiary] ('the Beneficiary')] [OPTION 2: [insert name of the linked third party] ('the Linked Third Party'), third party linked to the Beneficiary [insert name of the beneficiary] ('the Beneficiary')],

we

[name of the auditor] ('the Auditor'),

established at

[full address/city/state/province/country],

represented by

[name and function of an authorised representative],

have carried out the procedures agreed with you regarding the costs declared in the Financial Statement(s)<sup>3</sup> of the [Beneficiary] [Linked Third Party] concerning the grant agreement [insert grant agreement reference: number, title of the action and acronym] ('the Agreement'),

with a total cost declared of  
[total amount] EUR,

and a total of actual costs and 'direct personnel costs declared as unit costs calculated in accordance with the [Beneficiary's] [Linked Third Party's] usual cost accounting practices' declared of

[sum of total actual costs and total direct personnel costs declared as unit costs calculated in accordance with the [Beneficiary's] [Linked Third Party's] usual cost accounting practices] EUR

and **hereby provide our Independent Report of Factual Findings ('the Report')** using the compulsory report format agreed with you.

**The Report**

Our engagement was carried out in accordance with the terms of reference ('the ToR') appended to this Report. The Report includes the agreed-upon procedures ('the Procedures') carried out and the standard factual findings ('the Findings') examined.

---

<sup>3</sup> By which the Beneficiary declares costs under the Agreement (see template 'Model Financial Statement' in Annex 4 to the Agreement).

The Procedures were carried out solely to assist the [Commission] [Agency] in evaluating whether the [Beneficiary's] [Linked Third Party's] costs in the accompanying Financial Statement(s) were declared in accordance with the Agreement. The [Commission] [Agency] draws its own conclusions from the Report and any additional information it may require.

The scope of the Procedures was defined by the Commission. Therefore, the Auditor is not responsible for their suitability or pertinence. Since the Procedures carried out constitute neither an audit nor a review made in accordance with International Standards on Auditing or International Standards on Review Engagements, the Auditor does not give a statement of assurance on the Financial Statements.

Had the Auditor carried out additional procedures or an audit of the [Beneficiary's] [Linked Third Party's] Financial Statements in accordance with International Standards on Auditing or International Standards on Review Engagements, other matters might have come to its attention and would have been included in the Report.

### **Not applicable Findings**

We examined the Financial Statement(s) stated above and considered the following Findings not applicable:

*Explanation (to be removed from the Report):*

*If a Finding was not applicable, it must be marked as 'N.A.' ('Not applicable') in the corresponding row on the right-hand column of the table and means that the Finding did not have to be corroborated by the Auditor and the related Procedure(s) did not have to be carried out.*

*The reasons of the non-application of a certain Finding must be obvious i.e.*

- i) if no cost was declared under a certain category then the related Finding(s) and Procedure(s) are not applicable;*
- ii) if the condition set to apply certain Procedure(s) are not met the related Finding(s) and those Procedure(s) are not applicable. For instance, for 'beneficiaries with accounts established in a currency other than euro' the Procedure and Finding related to 'beneficiaries with accounts established in euro' are not applicable. Similarly, if no additional remuneration is paid, the related Finding(s) and Procedure(s) for additional remuneration are not applicable.*

**List here all Findings considered not applicable for the present engagement and explain the reasons of the non-applicability.**

....

### **Exceptions**

Apart from the exceptions listed below, the [Beneficiary] [Linked Third Party] provided the Auditor all the documentation and accounting information needed by the Auditor to carry out the requested Procedures and evaluate the Findings.

*Explanation (to be removed from the Report):*

- If the Auditor was not able to successfully complete a procedure requested, it must be marked as 'E' ('Exception') in the corresponding row on the right-hand column of the table. The reason such as the inability to reconcile key information or the unavailability of data that prevents the Auditor from carrying out the Procedure must be indicated below.*
- If the Auditor cannot corroborate a standard finding after having carried out the corresponding procedure, it must also be marked as 'E' ('Exception') and, where possible, the reasons why the Finding was not fulfilled and its possible impact must be explained here below.*

**List here any exceptions and add any information on the cause and possible consequences of each exception, if known. If the exception is quantifiable, include the corresponding amount.**

....



**Example (to be removed from the Report):**

1. *The Beneficiary was unable to substantiate the Finding number 1 on ... because ....*
2. *Finding number 30 was not fulfilled because the methodology used by the Beneficiary to calculate unit costs was different from the one approved by the Commission. The differences were as follows: ...*
3. *After carrying out the agreed procedures to confirm the Finding number 31, the Auditor found a difference of \_\_\_\_\_ EUR. The difference can be explained by ...*

**Further Remarks**

In addition to reporting on the results of the specific procedures carried out, the Auditor would like to make the following general remarks:

**Example (to be removed from the Report):**

1. *Regarding Finding number 8 the conditions for additional remuneration were considered as fulfilled because ...*
2. *In order to be able to confirm the Finding number 15 we carried out the following additional procedures: ....*

**Use of this Report**

This Report may be used only for the purpose described in the above objective. It was prepared solely for the confidential use of the [Beneficiary] [Linked Third Party] and the [Commission] [Agency], and only to be submitted to the [Commission] [Agency] in connection with the requirements set out in Article 20.4 of the Agreement. The Report may not be used by the [Beneficiary] [Linked Third Party] or by the [Commission] [Agency] for any other purpose, nor may it be distributed to any other parties. The [Commission] [Agency] may only disclose the Report to authorised parties, in particular to the European Anti-Fraud Office (OLAF) and the European Court of Auditors.

This Report relates only to the Financial Statement(s) submitted to the [Commission] [Agency] by the [Beneficiary] [Linked Third Party] for the Agreement. Therefore, it does not extend to any other of the [Beneficiary's] [Linked Third Party's] Financial Statement(s).

There was no conflict of interest<sup>4</sup> between the Auditor and the Beneficiary [and Linked Third Party] in establishing this Report. The total fee paid to the Auditor for providing the Report was EUR [ ] (including EUR [ ] of deductible VAT).

We look forward to discussing our Report with you and would be pleased to provide any further information or assistance.

[legal name of the Auditor]

[name and function of an authorised representative]

[dd Month yyyy]

Signature of the Auditor

<sup>4</sup> A conflict of interest arises when the Auditor's objectivity to establish the certificate is compromised in fact or in appearance when the Auditor for instance:

- was involved in the preparation of the Financial Statements;
- stands to benefit directly should the certificate be accepted;
- has a close relationship with any person representing the beneficiary;
- is a director, trustee or partner of the beneficiary; or
- is in any other situation that compromises his or her independence or ability to establish the certificate impartially.

**Agreed-upon procedures to be performed and standard factual findings to be confirmed by the Auditor**

The European Commission reserves the right to i) provide the auditor with additional guidance regarding the procedures to be followed or the facts to be ascertained and the way in which to present them (this may include sample coverage and findings) or to ii) change the procedures, by notifying the Beneficiary in writing. The procedures carried out by the auditor to confirm the standard factual finding are listed in the table below.

If this certificate relates to a Linked Third Party, any reference here below to ‘the Beneficiary’ is to be considered as a reference to ‘the Linked Third Party’.

The ‘result’ column has three different options: ‘C’, ‘E’ and ‘N.A.’:

- ‘C’ stands for ‘confirmed’ and means that the auditor can confirm the ‘standard factual finding’ and, therefore, there is no exception to be reported.
- ‘E’ stands for ‘exception’ and means that the Auditor carried out the procedures but cannot confirm the ‘standard factual finding’, or that the Auditor was not able to carry out a specific procedure (e.g. because it was impossible to reconcile key information or data were unavailable),
- ‘N.A.’ stands for ‘not applicable’ and means that the Finding did not have to be examined by the Auditor and the related Procedure(s) did not have to be carried out. The reasons of the non-application of a certain Finding must be obvious i.e. i) if no cost was declared under a certain category then the related Finding(s) and Procedure(s) are not applicable; ii) if the condition set to apply certain Procedure(s) are not met then the related Finding(s) and Procedure(s) are not applicable. For instance, for ‘beneficiaries with accounts established in a currency other than the euro’ the Procedure related to ‘beneficiaries with accounts established in euro’ is not applicable. Similarly, if no additional remuneration is paid, the related Finding(s) and Procedure(s) for additional remuneration are not applicable.

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
A	<b>ACTUAL PERSONNEL COSTS AND UNIT COSTS CALCULATED BY THE BENEFICIARY IN ACCORDANCE WITH ITS USUAL COST ACCOUNTING PRACTICE</b>		
	<p>The Auditor draws a sample of persons whose costs were declared in the Financial Statement(s) to carry out the procedures indicated in the consecutive points of this section A.</p> <p><i>(The sample should be selected randomly so that it is representative. Full coverage is required if there are fewer than 10 people (including employees, natural persons working under a direct contract and personnel seconded by a third party), otherwise the sample should have a minimum of 10 people, or 10% of the total, whichever number is the highest)</i></p> <p>The Auditor sampled [ ] people out of the total of [ ] people.</p>		



Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
A.1	<p><b>PERSONNEL COSTS</b></p> <p><u>For the persons included in the sample and working under an employment contract or equivalent act (general procedures for individual actual personnel costs and personnel costs declared as unit costs)</u></p> <p>To confirm standard factual findings 1-5 listed in the next column, the Auditor reviewed following information/documents provided by the Beneficiary:</p> <ul style="list-style-type: none"> <li>○ a list of the persons included in the sample indicating the period(s) during which they worked for the action, their position (classification or category) and type of contract;</li> <li>○ the payslips of the employees included in the sample;</li> <li>○ reconciliation of the personnel costs declared in the Financial Statement(s) with the accounting system (project accounting and general ledger) and payroll system;</li> <li>○ information concerning the employment status and employment conditions of personnel included in the sample, in particular their employment contracts or equivalent;</li> <li>○ the Beneficiary’s usual policy regarding payroll matters (e.g. salary policy, overtime policy, variable pay);</li> <li>○ applicable national law on taxes, labour and social security and</li> <li>○ any other document that supports the personnel costs declared.</li> </ul> <p>The Auditor also verified the eligibility of all components of the retribution (see Article 6 GA) and recalculated the personnel costs for employees included in the sample.</p>	1) The employees were i) directly hired by the Beneficiary in accordance with its national legislation, ii) under the Beneficiary’s sole technical supervision and responsibility and iii) remunerated in accordance with the Beneficiary’s usual practices.	
		2) Personnel costs were recorded in the Beneficiary's accounts/payroll system.	
		3) Costs were adequately supported and reconciled with the accounts and payroll records.	
		4) Personnel costs did not contain any ineligible elements.	
		5) There were no discrepancies between the personnel costs charged to the action and the costs recalculated by the Auditor.	
	<p><i>Further procedures if ‘additional remuneration’ is paid</i></p> <p>To confirm standard factual findings 6-9 listed in the next column, the Auditor:</p> <ul style="list-style-type: none"> <li>○ reviewed relevant documents provided by the Beneficiary (legal form, legal/statutory</li> </ul>	6) The Beneficiary paying “additional remuneration” was a non-profit legal entity.	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	<p>obligations, the Beneficiary’s usual policy on additional remuneration, criteria used for its calculation...);</p> <ul style="list-style-type: none"> <li>○ recalculated the amount of additional remuneration eligible for the action based on the supporting documents received (full-time or part-time work, exclusive or non-exclusive dedication to the action, etc.) to arrive at the applicable FTE/year and pro-rata rate (see data collected in the course of carrying out the procedures under A.2 ‘Productive hours’ and A.4 ‘Time recording system’).</li> </ul> <p><i>IF ANY PART OF THE REMUNERATION PAID TO THE EMPLOYEE IS NOT MANDATORY ACCORDING TO THE NATIONAL LAW OR THE EMPLOYMENT CONTRACT ("ADDITIONAL REMUNERATION") AND IS ELIGIBLE UNDER THE PROVISIONS OF ARTICLE 6.2.A.1, THIS CAN BE CHARGED AS ELIGIBLE COST TO THE ACTION UP TO THE FOLLOWING AMOUNT:</i></p> <p>(A) <i>IF THE PERSON WORKS FULL TIME AND EXCLUSIVELY ON THE ACTION DURING THE FULL YEAR: UP TO EUR 8 000/YEAR;</i></p> <p>(B) <i>IF THE PERSON WORKS EXCLUSIVELY ON THE ACTION BUT NOT FULL-TIME OR NOT FOR THE FULL YEAR: UP TO THE CORRESPONDING PRO-RATA AMOUNT OF EUR 8 000, OR</i></p> <p>(C) <i>IF THE PERSON DOES NOT WORK EXCLUSIVELY ON THE ACTION: UP TO A PRO-RATA AMOUNT CALCULATED IN ACCORDANCE TO ARTICLE 6.2.A.1.</i></p>	<p>7) The amount of additional remuneration paid corresponded to the Beneficiary’s usual remuneration practices and was consistently paid whenever the same kind of work or expertise was required.</p>	
		<p>8) The criteria used to calculate the additional remuneration were objective and generally applied by the Beneficiary regardless of the source of funding used.</p>	
		<p>9) The amount of additional remuneration included in the personnel costs charged to the action was capped at EUR 8,000 per FTE/year (up to the equivalent pro-rata amount if the person did not work on the action full-time during the year or did not work exclusively on the action).</p>	
	<p><i>Additional procedures in case “unit costs calculated by the Beneficiary in accordance with its usual cost accounting practices” is applied:</i></p> <p>Apart from carrying out the procedures indicated above to confirm standard factual findings 1-5 and, if applicable, also 6-9, the Auditor carried out following procedures to confirm standard</p>	<p>10) The personnel costs included in the Financial Statement were calculated in accordance with the Beneficiary's usual cost accounting practice. This methodology was consistently</p>	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	<p>factual findings 10-13 listed in the next column:</p> <ul style="list-style-type: none"> <li>○ obtained a description of the Beneficiary's usual cost accounting practice to calculate unit costs;</li> <li>○ reviewed whether the Beneficiary's usual cost accounting practice was applied for the Financial Statements subject of the present CFS;</li> <li>○ verified the employees included in the sample were charged under the correct category (in accordance with the criteria used by the Beneficiary to establish personnel categories) by reviewing the contract/HR-record or analytical accounting records;</li> <li>○ verified that there is no difference between the total amount of personnel costs used in calculating the cost per unit and the total amount of personnel costs recorded in the statutory accounts;</li> <li>○ verified whether actual personnel costs were adjusted on the basis of budgeted or estimated elements and, if so, verified whether those elements used are actually relevant for the calculation, objective and supported by documents.</li> </ul>	<p>used in all H2020 actions.</p> <p>11) The employees were charged under the correct category.</p> <p>12) Total personnel costs used in calculating the unit costs were consistent with the expenses recorded in the statutory accounts.</p> <p>13) Any estimated or budgeted element used by the Beneficiary in its unit-cost calculation were relevant for calculating personnel costs and corresponded to objective and verifiable information.</p>	
	<p><u>For natural persons included in the sample and working with the Beneficiary under a direct contract other than an employment contract, such as consultants (no subcontractors).</u></p> <p>To confirm standard factual findings 14-18 listed in the next column the Auditor reviewed following information/documents provided by the Beneficiary:</p> <ul style="list-style-type: none"> <li>○ the contracts, especially the cost, contract duration, work description, place of work, ownership of the results and reporting obligations to the Beneficiary;</li> <li>○ the employment conditions of staff in the same category to compare costs and;</li> <li>○ any other document that supports the costs declared and its registration (e.g. invoices,</li> </ul>	<p>14) The natural persons reported to the Beneficiary (worked under the Beneficiary's instructions).</p> <p>15) They worked on the Beneficiary's premises (unless otherwise agreed with the Beneficiary).</p> <p>16) The results of work carried out belong to the Beneficiary.</p>	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	accounting records, etc.).	17) Their costs were not significantly different from those for staff who performed similar tasks under an employment contract with the Beneficiary.	
		18) The costs were supported by audit evidence and registered in the accounts.	
	<p><u>For personnel seconded by a third party and included in the sample (not subcontractors)</u></p> <p>To confirm standard factual findings 19-22 listed in the next column, the Auditor reviewed following information/documents provided by the Beneficiary:</p> <ul style="list-style-type: none"> <li>○ their secondment contract(s) notably regarding costs, duration, work description, place of work and ownership of the results;</li> <li>○ if there is reimbursement by the Beneficiary to the third party for the resource made available (in-kind contribution against payment): any documentation that supports the costs declared (e.g. contract, invoice, bank payment, and proof of registration in its accounting/payroll, etc.) and reconciliation of the Financial Statement(s) with the accounting system (project accounting and general ledger) as well as any proof that the amount invoiced by the third party did not include any profit;</li> <li>○ if there is no reimbursement by the Beneficiary to the third party for the resource made available (in-kind contribution free of charge): a proof of the actual cost borne by the Third Party for the resource made available free of charge to the Beneficiary such as a statement of costs incurred by the Third Party and proof of the registration in the Third Party's accounting/payroll;</li> <li>○ any other document that supports the costs declared (e.g. invoices, etc.).</li> </ul>	19) Seconded personnel reported to the Beneficiary and worked on the Beneficiary's premises (unless otherwise agreed with the Beneficiary).	
		20) The results of work carried out belong to the Beneficiary.	
		<p><i>If personnel is seconded against payment:</i></p> <p>21) The costs declared were supported with documentation and recorded in the Beneficiary's accounts. The third party did not include any profit.</p>	
		<p><i>If personnel is seconded free of charge:</i></p> <p>22) The costs declared did not exceed the third party's cost as</p>	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
		recorded in the accounts of the third party and were supported with documentation.	
A.2	<p><b>PRODUCTIVE HOURS</b></p> <p>To confirm standard factual findings 23-28 listed in the next column, the Auditor reviewed relevant documents, especially national legislation, labour agreements and contracts and time records of the persons included in the sample, to verify that:</p> <ul style="list-style-type: none"> <li>○ the annual productive hours applied were calculated in accordance with one of the methods described below,</li> <li>○ the full-time equivalent (FTEs) ratios for employees not working full-time were correctly calculated.</li> </ul> <p>If the Beneficiary applied method B, the auditor verified that the correctness in which the total number of hours worked was calculated and that the contracts specified the annual workable hours.</p> <p>If the Beneficiary applied method C, the auditor verified that the ‘annual productive hours’ applied when calculating the hourly rate were equivalent to at least 90 % of the ‘standard annual workable hours’. The Auditor can only do this if the calculation of the standard annual workable hours can be supported by records, such as national legislation, labour agreements, and contracts.</p> <p><i>BENEFICIARY'S PRODUCTIVE HOURS' FOR PERSONS WORKING FULL TIME SHALL BE ONE OF THE FOLLOWING METHODS:</i></p> <p><i>A. 1720 ANNUAL PRODUCTIVE HOURS (PRO-RATA FOR PERSONS NOT WORKING FULL-TIME)</i></p> <p><i>B. THE TOTAL NUMBER OF HOURS WORKED BY THE PERSON FOR THE BENEFICIARY IN THE YEAR (THIS METHOD IS ALSO REFERRED TO AS ‘TOTAL NUMBER OF HOURS WORKED’ IN THE NEXT COLUMN). THE CALCULATION OF THE TOTAL NUMBER OF HOURS WORKED WAS DONE AS FOLLOWS: ANNUAL WORKABLE HOURS OF THE PERSON ACCORDING TO THE EMPLOYMENT</i></p>	<p>23) The Beneficiary applied method [<i>choose one option and delete the others</i>] [A: 1720 hours] [B: the ‘total number of hours worked’] [C: ‘standard annual productive hours’ used correspond to usual accounting practices]</p> <p>24) Productive hours were calculated annually.</p> <p>25) For employees not working full-time the full-time equivalent (FTE) ratio was correctly applied.</p> <p><i>If the Beneficiary applied method B.</i></p> <p>26) The calculation of the number of ‘annual workable hours’, overtime and absences was verifiable based on the documents provided by the Beneficiary.</p>	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	<p><i>CONTRACT, APPLICABLE LABOUR AGREEMENT OR NATIONAL LAW PLUS OVERTIME WORKED MINUS ABSENCES (SUCH AS SICK LEAVE OR SPECIAL LEAVE).</i></p> <p><i>C. THE STANDARD NUMBER OF ANNUAL HOURS GENERALLY APPLIED BY THE BENEFICIARY FOR ITS PERSONNEL IN ACCORDANCE WITH ITS USUAL COST ACCOUNTING PRACTICES (THIS METHOD IS ALSO REFERRED TO AS 'STANDARD ANNUAL PRODUCTIVE HOURS' IN THE NEXT COLUMN). THIS NUMBER MUST BE AT LEAST 90% OF THE STANDARD ANNUAL WORKABLE HOURS.</i></p> <p><i>'ANNUAL WORKABLE HOURS' MEANS THE PERIOD DURING WHICH THE PERSONNEL MUST BE WORKING, AT THE EMPLOYER'S DISPOSAL AND CARRYING OUT HIS/HER ACTIVITY OR DUTIES UNDER THE EMPLOYMENT CONTRACT, APPLICABLE COLLECTIVE LABOUR AGREEMENT OR NATIONAL WORKING TIME LEGISLATION.</i></p>	<p>26.1) The Beneficiary calculates the hourly rates per full financial year following procedure A.3 (method B is not allowed for beneficiaries calculating hourly rates per month).</p> <p><i>If the Beneficiary applied method C.</i></p> <p>27) The calculation of the number of 'standard annual workable hours' was verifiable based on the documents provided by the Beneficiary.</p> <p>28) The 'annual productive hours' used for calculating the hourly rate were consistent with the usual cost accounting practices of the Beneficiary and were equivalent to at least 90 % of the 'annual workable hours'.</p>	
A.3	<p><b>HOURLY PERSONNEL RATES</b></p> <p><u>D) For unit costs calculated in accordance to the Beneficiary's usual cost accounting practice (unit costs):</u></p> <p>If the Beneficiary has a "Certificate on Methodology to calculate unit costs " (CoMUC) approved by the Commission, the Beneficiary provides the Auditor with a description of the approved methodology and the Commission's letter of acceptance. The Auditor verified that the</p>	<p>29) The Beneficiary applied [<i>choose one option and delete the other</i>]:</p> <p>[Option I: "Unit costs (hourly rates) were calculated in accordance with the Beneficiary's usual cost</p>	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	<p>Beneficiary has indeed used the methodology approved. If so, no further verification is necessary.</p> <p>If the Beneficiary does not have a "Certificate on Methodology" (CoMUC) approved by the Commission, or if the methodology approved was not applied, then the Auditor:</p> <ul style="list-style-type: none"> <li>○ reviewed the documentation provided by the Beneficiary, including manuals and internal guidelines that explain how to calculate hourly rates;</li> <li>○ recalculated the unit costs (hourly rates) of staff included in the sample following the results of the procedures carried out in A.1 and A.2.</li> </ul> <p><u>II) For individual hourly rates:</u></p> <p>The Auditor:</p> <ul style="list-style-type: none"> <li>○ reviewed the documentation provided by the Beneficiary, including manuals and internal guidelines that explain how to calculate hourly rates;</li> <li>○ recalculated the hourly rates of staff included in the sample (recalculation of all hourly rates if the Beneficiary uses annual rates, recalculation of three months selected randomly for every year and person if the Beneficiary uses monthly rates) following the results of the procedures carried out in A.1 and A.2;</li> <li>○ (only in case of monthly rates) confirmed that the time spent on parental leave is not deducted, and that, if parts of the basic remuneration are generated over a period longer than a month, the Beneficiary has included only the share which is generated in the month.</li> </ul> <p><u>“UNIT COSTS CALCULATED BY THE BENEFICIARY IN ACCORDANCE WITH ITS USUAL COST ACCOUNTING PRACTICES”:</u></p> <p><i>IT IS CALCULATED BY DIVIDING THE TOTAL AMOUNT OF PERSONNEL COSTS OF THE CATEGORY TO WHICH THE EMPLOYEE BELONGS VERIFIED IN LINE WITH PROCEDURE A.1 BY THE NUMBER OF FTE AND THE ANNUAL TOTAL PRODUCTIVE HOURS OF THE SAME CATEGORY CALCULATED BY THE BENEFICIARY IN ACCORDANCE WITH PROCEDURE A.2.</i></p>	<p>accounting practices”]</p> <p>[Option II: Individual hourly rates were applied]</p> <p><i>For option I concerning unit costs and if the Beneficiary applies the methodology approved by the Commission (CoMUC):</i></p> <p>30) The Beneficiary used the Commission-approved methodology to calculate hourly rates. It corresponded to the organisation's usual cost accounting practices and was applied consistently for all activities irrespective of the source of funding.</p> <p><i>For option I concerning unit costs and if the Beneficiary applies a methodology not approved by the Commission:</i></p> <p>31) The unit costs re-calculated by the Auditor were the same as the rates applied by the Beneficiary.</p> <p><i>For option II concerning individual hourly rates:</i></p> <p>32) The individual rates re-</p>	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	<p><u>HOURLY RATE FOR INDIVIDUAL ACTUAL PERSONAL COSTS:</u>  <i>IT IS CALCULATED FOLLOWING ONE OF THE TWO OPTIONS BELOW:</i></p> <p>A) [OPTION BY DEFAULT] BY DIVIDING THE ACTUAL ANNUAL AMOUNT OF PERSONNEL COSTS OF AN EMPLOYEE VERIFIED IN LINE WITH PROCEDURE A.1 BY THE NUMBER OF ANNUAL PRODUCTIVE HOURS VERIFIED IN LINE WITH PROCEDURE A.2 (FULL FINANCIAL YEAR HOURLY RATE);</p> <p>B) BY DIVIDING THE ACTUAL MONTHLY AMOUNT OF PERSONNEL COSTS OF AN EMPLOYEE VERIFIED IN LINE WITH PROCEDURE A.1 BY 1/12 OF THE NUMBER OF ANNUAL PRODUCTIVE HOURS VERIFIED IN LINE WITH PROCEDURE A.2.(MONTHLY HOURLY RATE).</p>	<p>calculated by the Auditor were the same as the rates applied by the Beneficiary.</p> <p>32.1) The Beneficiary used only one option (per full financial year or per month) throughout each financial year examined.</p>	
A.4	<p><b>TIME RECORDING SYSTEM</b></p> <p>To verify that the time recording system ensures the fulfilment of all minimum requirements and that the hours declared for the action were correct, accurate and properly authorised and supported by documentation, the Auditor made the following checks for the persons included in the sample that declare time as worked for the action on the basis of time records:</p> <ul style="list-style-type: none"> <li>○ description of the time recording system provided by the Beneficiary (registration, authorisation, processing in the HR-system);</li> <li>○ its actual implementation;</li> <li>○ time records were signed at least monthly by the employees (on paper or electronically) and authorised by the project manager or another manager;</li> <li>○ the hours declared were worked within the project period;</li> <li>○ there were no hours declared as worked for the action if HR-records showed absence due to holidays or sickness (further cross-checks with travels are carried out in B.1 below) ;</li> <li>○ the hours charged to the action matched those in the time recording system.</li> </ul>	<p>33) All persons recorded their time dedicated to the action on a <b>daily/ weekly/ monthly</b> basis using a <b>paper/computer-based</b> system. <i>(delete the answers that are not applicable)</i></p> <p>34) Their time-records were authorised at least monthly by the project manager or other superior.</p> <p>35) Hours declared were worked within the project period and were consistent with the presences/absences recorded in HR-records.</p>	



Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	<p><i>ONLY THE HOURS WORKED ON THE ACTION CAN BE CHARGED. ALL WORKING TIME TO BE CHARGED SHOULD BE RECORDED THROUGHOUT THE DURATION OF THE PROJECT, ADEQUATELY SUPPORTED BY EVIDENCE OF THEIR REALITY AND RELIABILITY (SEE SPECIFIC PROVISIONS BELOW FOR PERSONS WORKING EXCLUSIVELY FOR THE ACTION WITHOUT TIME RECORDS).</i></p>	36) There were no discrepancies between the number of hours charged to the action and the number of hours recorded.	
	<p><u>If the persons are working exclusively for the action and without time records</u></p> <p>For the persons selected that worked exclusively for the action without time records, the Auditor verified evidence available demonstrating that they were in reality exclusively dedicated to the action and that the Beneficiary signed a declaration confirming that they have worked exclusively for the action.</p>	37) The exclusive dedication is supported by a declaration signed by the Beneficiary's and by any other evidence gathered.	
<b>B</b>	<b>COSTS OF SUBCONTRACTING</b>		
<b>B.1</b>	<p><b>The Auditor obtained the detail/breakdown of subcontracting costs and sampled cost items selected randomly</b> <i>(full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is highest).</i></p> <p>To confirm standard factual findings 38-42 listed in the next column, the Auditor reviewed the following for the items included in the sample:</p> <ul style="list-style-type: none"> <li>○ the use of subcontractors was foreseen in Annex 1;</li> <li>○ subcontracting costs were declared in the subcontracting category of the Financial Statement;</li> <li>○ supporting documents on the selection and award procedure were followed;</li> <li>○ the Beneficiary ensured best value for money (key elements to appreciate the respect of this principle are the award of the subcontract to the bid offering best price-quality ratio, under conditions of transparency and equal treatment. In case an existing framework contract was used the Beneficiary ensured it was established on the basis of the principle of best value for money under conditions of transparency and equal treatment).</li> </ul>	<p>38) The use of claimed subcontracting costs was foreseen in Annex 1 and costs were declared in the Financial Statements under the subcontracting category.</p> <p>39) There were documents of requests to different providers, different offers and assessment of the offers before selection of the provider in line with internal procedures and procurement rules. Subcontracts were awarded in accordance with the principle of best value for money.</p> <p><i>(When different offers were not collected the Auditor explains</i></p>	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	<p>In particular,</p> <ul style="list-style-type: none"> <li>i. if the Beneficiary acted as a contracting authority within the meaning of Directive 2004/18/EC (or 2014/24/EU) or of Directive 2004/17/EC (or 2014/25/EU), the Auditor verified that the applicable national law on public procurement was followed and that the subcontracting complied with the Terms and Conditions of the Agreement.</li> <li>ii. if the Beneficiary did not fall under the above-mentioned category the Auditor verified that the Beneficiary followed their usual procurement rules and respected the Terms and Conditions of the Agreement..</li> </ul> <p>For the items included in the sample the Auditor also verified that:</p> <ul style="list-style-type: none"> <li>o the subcontracts were not awarded to other Beneficiaries in the consortium;</li> <li>o there were signed agreements between the Beneficiary and the subcontractor;</li> <li>o there was evidence that the services were provided by subcontractor;</li> </ul>	<p><i>the reasons provided by the Beneficiary under the caption “Exceptions” of the Report. The Commission will analyse this information to evaluate whether these costs might be accepted as eligible)</i></p>	
		40) The subcontracts were not awarded to other Beneficiaries of the consortium.	
		41) All subcontracts were supported by signed agreements between the Beneficiary and the subcontractor.	
		42) There was evidence that the services were provided by the subcontractors.	
<b>C</b>	<b>COSTS OF PROVIDING FINANCIAL SUPPORT TO THIRD PARTIES</b>		
<b>C.1</b>	<p><b>The Auditor obtained the detail/breakdown of the costs of providing financial support to third parties and sampled [redacted] cost items selected randomly (full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is highest).</b></p> <p>The Auditor verified that the following minimum conditions were met:</p> <ul style="list-style-type: none"> <li>a) the maximum amount of financial support for each third party did not exceed EUR 60 000, unless explicitly mentioned in Annex 1;</li> </ul>	43) All minimum conditions were met	

<b>Ref</b>	<b>Procedures</b>	<b>Standard factual finding</b>	<b>Result (C / E / N.A.)</b>
	b) the financial support to third parties was agreed in Annex 1 of the Agreement and the other provisions on financial support to third parties included in Annex 1 were respected.		

<b>D</b>	<b>OTHER ACTUAL DIRECT COSTS</b>		
<b>D.1</b>	<p><b>COSTS OF TRAVEL AND RELATED SUBSISTENCE ALLOWANCES</b></p> <p><b>The Auditor sampled [REDACTED] cost items selected randomly</b> (<i>full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is the highest</i>).</p> <p>The Auditor inspected the sample and verified that:</p> <ul style="list-style-type: none"> <li>○ travel and subsistence costs were consistent with the Beneficiary's usual policy for travel. In this context, the Beneficiary provided evidence of its normal policy for travel costs (e.g. use of first class tickets, reimbursement by the Beneficiary on the basis of actual costs, a lump sum or per diem) to enable the Auditor to compare the travel costs charged with this policy;</li> <li>○ travel costs are correctly identified and allocated to the action (e.g. trips are directly linked to the action) by reviewing relevant supporting documents such as minutes of meetings, workshops or conferences, their registration in the correct project account, their consistency with time records or with the dates/duration of the workshop/conference;</li> <li>○ no ineligible costs or excessive or reckless expenditure was declared.</li> </ul>	44) Costs were incurred, approved and reimbursed in line with the Beneficiary's usual policy for travels.	
		45) There was a link between the trip and the action.	
		46) The supporting documents were consistent with each other regarding subject of the trip, dates, duration and reconciled with time records and accounting.	
		47) No ineligible costs or excessive or reckless expenditure was declared.	
<b>D.2</b>	<p><b>DEPRECIATION COSTS FOR EQUIPMENT, INFRASTRUCTURE OR OTHER ASSETS</b></p> <p><b>The Auditor sampled [REDACTED] cost items selected randomly</b> (<i>full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is the highest</i>).</p> <p>For “equipment, infrastructure or other assets” [from now on called “asset(s)”] selected in the sample the Auditor verified that:</p> <ul style="list-style-type: none"> <li>○ the assets were acquired in conformity with the Beneficiary's internal guidelines and procedures;</li> <li>○ they were correctly allocated to the action (with supporting documents such as delivery</li> </ul>	48) Procurement rules, principles and guides were followed.	
		49) There was a link between the grant agreement and the asset charged to the action.	
		50) The asset charged to the action was traceable to the accounting records and the underlying documents.	

	<p>note invoice or any other proof demonstrating the link to the action)</p> <ul style="list-style-type: none"> <li>○ they were entered in the accounting system;</li> <li>○ the extent to which the assets were used for the action (as a percentage) was supported by reliable documentation (e.g. usage overview table);</li> </ul> <p>The Auditor recalculated the depreciation costs and verified that they were in line with the applicable rules in the Beneficiary's country and with the Beneficiary's usual accounting policy (e.g. depreciation calculated on the acquisition value).</p> <p>The Auditor verified that no ineligible costs such as deductible VAT, exchange rate losses, excessive or reckless expenditure were declared (see Article 6.5 GA).</p>	51) The depreciation method used to charge the asset to the action was in line with the applicable rules of the Beneficiary's country and the Beneficiary's usual accounting policy.	
		52) The amount charged corresponded to the actual usage for the action.	
		53) No ineligible costs or excessive or reckless expenditure were declared.	
<b>D.3</b>	<p><b>COSTS OF OTHER GOODS AND SERVICES</b></p> <p><b>The Auditor sampled [REDACTED] cost items selected randomly</b> (<i>full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 items, or 10% of the total, whichever number is highest</i>).</p> <p>For the purchase of goods, works or services included in the sample the Auditor verified that:</p> <ul style="list-style-type: none"> <li>○ the contracts did not cover tasks described in Annex 1;</li> <li>○ they were correctly identified, allocated to the proper action, entered in the accounting system (traceable to underlying documents such as purchase orders, invoices and accounting);</li> <li>○ the goods were not placed in the inventory of durable equipment;</li> <li>○ the costs charged to the action were accounted in line with the Beneficiary's usual accounting practices;</li> <li>○ no ineligible costs or excessive or reckless expenditure were declared (see Article 6 GA).</li> </ul> <p>In addition, the Auditor verified that these goods and services were acquired in conformity with the Beneficiary's internal guidelines and procedures, in particular:</p> <ul style="list-style-type: none"> <li>○ if Beneficiary acted as a contracting authority within the meaning of Directive</li> </ul>	54) Contracts for works or services did not cover tasks described in Annex 1.	
		55) Costs were allocated to the correct action and the goods were not placed in the inventory of durable equipment.	
		56) The costs were charged in line with the Beneficiary's accounting policy and were adequately supported.	
		57) No ineligible costs or excessive or reckless expenditure were declared. For internal invoices/charges only the cost element was charged, without any mark-ups.	

	<p>2004/18/EC (or 2014/24/EU) or of Directive 2004/17/EC (or 2014/25/EU), the Auditor verified that the applicable national law on public procurement was followed and that the procurement contract complied with the Terms and Conditions of the Agreement.</p> <ul style="list-style-type: none"> <li>○ if the Beneficiary did not fall into the category above, the Auditor verified that the Beneficiary followed their usual procurement rules and respected the Terms and Conditions of the Agreement.</li> </ul> <p>For the items included in the sample the Auditor also verified that:</p> <ul style="list-style-type: none"> <li>○ the Beneficiary ensured best value for money (key elements to appreciate the respect of this principle are the award of the contract to the bid offering best price-quality ratio, under conditions of transparency and equal treatment. In case an existing framework contract was used the Auditor also verified that the Beneficiary ensured it was established on the basis of the principle of best value for money under conditions of transparency and equal treatment);</li> </ul> <p><i>SUCH GOODS AND SERVICES INCLUDE, FOR INSTANCE, CONSUMABLES AND SUPPLIES, DISSEMINATION (INCLUDING OPEN ACCESS), PROTECTION OF RESULTS, SPECIFIC EVALUATION OF THE ACTION IF IT IS REQUIRED BY THE AGREEMENT, CERTIFICATES ON THE FINANCIAL STATEMENTS IF THEY ARE REQUIRED BY THE AGREEMENT AND CERTIFICATES ON THE METHODOLOGY, TRANSLATIONS, REPRODUCTION.</i></p>	<p>58) Procurement rules, principles and guides were followed. There were documents of requests to different providers, different offers and assessment of the offers before selection of the provider in line with internal procedures and procurement rules. The purchases were made in accordance with the principle of best value for money.</p> <p><i>(When different offers were not collected the Auditor explains the reasons provided by the Beneficiary under the caption “Exceptions” of the Report. The Commission will analyse this information to evaluate whether these costs might be accepted as eligible)</i></p>	
<p><b>D.4</b></p>	<p><b>AGGREGATED CAPITALISED AND OPERATING COSTS OF RESEARCH INFRASTRUCTURE</b></p> <p>The Auditor ensured the existence of a positive ex-ante assessment (issued by the EC Services) of the cost accounting methodology of the Beneficiary allowing it to apply the guidelines on direct costing for large research infrastructures in Horizon 2020.</p> <p><i>In the cases that a positive ex-ante assessment has been issued (see the standard factual findings 59-60 on the next column),</i></p>	<p>59) The costs declared as direct costs for Large Research Infrastructures (in the appropriate line of the Financial Statement) comply with the methodology described in the positive ex-ante assessment report.</p>	

	<p>The Auditor ensured that the beneficiary has applied consistently the methodology that is explained and approved in the positive ex ante assessment;</p> <p><i>In the cases that a positive ex-ante assessment has NOT been issued (see the standard factual findings 61 on the next column),</i> The Auditor verified that no costs of Large Research Infrastructure have been charged as direct costs in any costs category;</p> <p><i>In the cases that a draft ex-ante assessment report has been issued with recommendation for further changes (see the standard factual findings 61 on the next column),</i></p> <ul style="list-style-type: none"> <li>The Auditor followed the same procedure as above (when a positive ex-ante assessment has NOT yet been issued) and paid particular attention (testing reinforced) to the cost items for which the draft ex-ante assessment either rejected the inclusion as direct costs for Large Research Infrastructures or issued recommendations.</li> </ul>	<p>60) Any difference between the methodology applied and the one positively assessed was extensively described and adjusted accordingly.</p>	
		<p>61) The direct costs declared were free from any indirect costs items related to the Large Research Infrastructure.</p>	
<p><b>E</b></p>	<p><b>USE OF EXCHANGE RATES</b></p>		
<p><b>E.1</b></p>	<p><u>a) For Beneficiaries with accounts established in a currency other than euros</u></p> <p><b>The Auditor sampled [REDACTED] cost items selected randomly and verified that the exchange rates used for converting other currencies into euros were in accordance with the following rules established in the Agreement ( full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is highest):</b></p> <p><i>COSTS RECORDED IN THE ACCOUNTS IN A CURRENCY OTHER THAN EURO SHALL BE CONVERTED INTO EURO AT THE AVERAGE OF THE DAILY EXCHANGE RATES PUBLISHED IN THE C SERIES OF OFFICIAL JOURNAL OF THE EUROPEAN UNION (<a href="https://www.ecb.int/stats/exchange/eurofxref/html/index.en.html">https://www.ecb.int/stats/exchange/eurofxref/html/index.en.html</a> ), DETERMINED OVER THE CORRESPONDING REPORTING PERIOD.</i></p> <p><i>IF NO DAILY EURO EXCHANGE RATE IS PUBLISHED IN THE OFFICIAL JOURNAL OF THE EUROPEAN UNION FOR THE CURRENCY IN QUESTION, CONVERSION SHALL BE MADE AT THE AVERAGE OF THE MONTHLY ACCOUNTING RATES ESTABLISHED BY THE COMMISSION AND PUBLISHED ON ITS WEBSITE (<a href="http://ec.europa.eu/budget/contracts_grants/info_contracts/inforeuro/inforeuro_en.cfm">http://ec.europa.eu/budget/contracts_grants/info_contracts/inforeuro/inforeuro_en.cfm</a> ), DETERMINED OVER THE CORRESPONDING REPORTING PERIOD.</i></p>	<p>62) The exchange rates used to convert other currencies into Euros were in accordance with the rules established of the Grant Agreement and there was no difference in the final figures.</p>	

	<p>b) <u>For Beneficiaries with accounts established in euros</u></p> <p><b>The Auditor sampled [REDACTED] cost items selected randomly and verified that the exchange rates used for converting other currencies into euros were in accordance with the following rules established in the Agreement ( full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is highest):</b></p> <p><i>COSTS INCURRED IN ANOTHER CURRENCY SHALL BE CONVERTED INTO EURO BY APPLYING THE BENEFICIARY'S USUAL ACCOUNTING PRACTICES.</i></p>	<p>63) The Beneficiary applied its usual accounting practices.</p>	
--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------	--

*[legal name of the audit firm]*

*[name and function of an authorised representative]*

*[dd Month yyyy]*

*<Signature of the Auditor>*



## ANNEX 6

### MODEL FOR THE CERTIFICATE ON THE METHODOLOGY

- For options [*in italics in square brackets*]: choose the applicable option. Options not chosen should be deleted.
- For fields in [grey in square brackets]: enter the appropriate data.

#### TABLE OF CONTENTS

TERMS OF REFERENCE FOR AN AUDIT ENGAGEMENT FOR A METHODOLOGY CERTIFICATE IN CONNECTION WITH ONE OR MORE GRANT AGREEMENTS FINANCED UNDER THE HORIZON 2020 RESEARCH AND INNOVATION FRAMEWORK PROGRAMME

INDEPENDENT REPORT OF FACTUAL FINDINGS ON THE METHODOLOGY CONCERNING GRANT AGREEMENTS FINANCED UNDER THE HORIZON 2020 RESEARCH AND INNOVATION FRAMEWORK PROGRAMME

**Terms of reference for an audit engagement for a methodology certificate  
in connection with one or more grant agreements financed  
under the Horizon 2020 Research and Innovation Framework Programme**

This document sets out the ‘**Terms of Reference (ToR)**’ under which

[OPTION 1: [insert name of the beneficiary] (*‘the Beneficiary’*)] [OPTION 2: [insert name of the linked third party] (*‘the Linked Third Party’*), third party linked to the Beneficiary [insert name of the beneficiary] (*‘the Beneficiary’*)]

agrees to engage

[insert legal name of the auditor] (*‘the Auditor’*)

to produce an independent report of factual findings (*‘the Report’*) concerning the [Beneficiary’s] [Linked Third Party’s] usual accounting practices for calculating and claiming direct personnel costs declared as unit costs (*‘the Methodology’*) in connection with grant agreements financed under the Horizon 2020 Research and Innovation Framework Programme.

The procedures to be carried out for the assessment of the methodology will be based on the grant agreement(s) detailed below:

[title and number of the grant agreement(s)] (*‘the Agreement(s)’*)

The Agreement(s) has(have) been concluded between the Beneficiary and [OPTION 1: *the European Union, represented by the European Commission (‘the Commission’)*][ OPTION 2: *the European Atomic Energy Community (Euratom,) represented by the European Commission (‘the Commission’)*][OPTION 3: *the [Research Executive Agency (REA)] [European Research Council Executive Agency (ERCEA)] [Innovation and Networks Executive Agency (INEA)] [Executive Agency for Small and Medium-sized Enterprises (EASME)] (‘the Agency’), under the powers delegated by the European Commission (‘the Commission’).*].

The [Commission] [Agency] is mentioned as a signatory of the Agreement with the Beneficiary only. The [European Union] [Euratom] [Agency] is not a party to this engagement.

### **1.1 Subject of the engagement**

According to Article 18.1.2 of the Agreement, beneficiaries [and linked third parties] that declare direct personnel costs as unit costs calculated in accordance with their usual cost accounting practices may submit to the [Commission] [Agency], for approval, a certificate on the methodology (*‘CoMUC’*) stating that there are adequate records and documentation to prove that their cost accounting practices used comply with the conditions set out in Point A of Article 6.2.

The subject of this engagement is the CoMUC which is composed of two separate documents:

- the Terms of Reference (*‘the ToR’*) to be signed by the [Beneficiary] [Linked Third Party] and the Auditor;
- the Auditor’s Independent Report of Factual Findings (*‘the Report’*) issued on the Auditor’s letterhead, dated, stamped and signed by the Auditor which includes; the standard statements (*‘the Statements’*) evaluated and signed by the [Beneficiary] [Linked Third Party], the agreed-upon procedures (*‘the Procedures’*) performed by the Auditor and the standard factual findings

(‘the Findings’) assessed by the Auditor. The Statements, Procedures and Findings are summarised in the table that forms part of the Report.

The information provided through the Statements, the Procedures and the Findings will enable the Commission to draw conclusions regarding the existence of the *[Beneficiary’s] [Linked Third Party’s]* usual cost accounting practice and its suitability to ensure that direct personnel costs claimed on that basis comply with the provisions of the Agreement. The Commission draws its own conclusions from the Report and any additional information it may require.

## 1.2 Responsibilities

The parties to this agreement are the *[Beneficiary] [Linked Third Party]* and the Auditor.

The *[Beneficiary] [Linked Third Party]*:

- is responsible for preparing financial statements for the Agreement(s) (‘the Financial Statements’) in compliance with those Agreements;
- is responsible for providing the Financial Statement(s) to the Auditor and enabling the Auditor to reconcile them with the *[Beneficiary’s] [Linked Third Party’s]* accounting and bookkeeping system and the underlying accounts and records. The Financial Statement(s) will be used as a basis for the procedures which the Auditor will carry out under this ToR;
- is responsible for its Methodology and liable for the accuracy of the Financial Statement(s);
- is responsible for endorsing or refuting the Statements indicated under the heading ‘Statements to be made by the Beneficiary/ Linked Third Party’ in the first column of the table that forms part of the Report;
- must provide the Auditor with a signed and dated representation letter;
- accepts that the ability of the Auditor to carry out the Procedures effectively depends upon the *[Beneficiary] [Linked Third Party]* providing full and free access to the *[Beneficiary’s] [Linked Third Party’s]* staff and to its accounting and other relevant records.

The Auditor:

- *[Option 1 by default: is qualified to carry out statutory audits of accounting documents in accordance with Directive 2006/43/EC of the European Parliament and of the Council of 17 May 2006 on statutory audits of annual accounts and consolidated accounts, amending Council Directives 78/660/EEC and 83/349/EEC and repealing Council Directive 84/253/EEC or similar national regulations].*
- *[Option 2 if the Beneficiary or Linked Third Party has an independent Public Officer: is a competent and independent Public Officer for which the relevant national authorities have established the legal capacity to audit the Beneficiary].*
- *[Option 3 if the Beneficiary or Linked Third Party is an international organisation: is an [internal] [external] auditor in accordance with the internal financial regulations and procedures of the international organisation].*

The Auditor:

- must be independent from the Beneficiary *[and the Linked Third Party]*, in particular, it must not have been involved in preparing the Beneficiary’s *[and Linked Third Party’s]* Financial Statement(s);
- must plan work so that the Procedures may be carried out and the Findings may be assessed;
- must adhere to the Procedures laid down and the compulsory report format;
- must carry out the engagement in accordance with these ToR;
- must document matters which are important to support the Report;
- must base its Report on the evidence gathered;
- must submit the Report to the *[Beneficiary] [Linked Third Party]*.

The Commission sets out the Procedures to be carried out and the Findings to be endorsed by the Auditor. The Auditor is not responsible for their suitability or pertinence. As this engagement is not an assurance engagement the Auditor does not provide an audit opinion or a statement of assurance.

### 1.3 Applicable Standards

The Auditor must comply with these Terms of Reference and with<sup>1</sup>:

- the International Standard on Related Services (‘ISRS’) 4400 *Engagements to perform Agreed-upon Procedures regarding Financial Information* as issued by the International Auditing and Assurance Standards Board (IAASB);
- the *Code of Ethics for Professional Accountants* issued by the International Ethics Standards Board for Accountants (IESBA). Although ISRS 4400 states that independence is not a requirement for engagements to carry out agreed-upon procedures, the Commission requires that the Auditor also complies with the Code’s independence requirements.

The Auditor’s Report must state that there was no conflict of interests in establishing this Report between the Auditor and the Beneficiary [*and the Linked Third Party*] that could have a bearing on the Report, and must specify – if the service is invoiced - the total fee paid to the Auditor for providing the Report.

### 1.4 Reporting

The Report must be written in the language of the Agreement (see Article 20.7 of the Agreement).

Under Article 22 of the Agreement, the Commission, [*the Agency*], the European Anti-Fraud Office and the Court of Auditors have the right to audit any work that is carried out under the action and for which costs are declared from [*the European Union*] [*Euratom*] budget. This includes work related to this engagement. The Auditor must provide access to all working papers related to this assignment if the Commission[, *the Agency*], the European Anti-Fraud Office or the European Court of Auditors requests them.

### 1.5 Timing

The Report must be provided by [dd Month yyyy].

### 1.6 Other Terms

*[The [Beneficiary] [Linked Third Party] and the Auditor can use this section to agree other specific terms, such as the Auditor’s fees, liability, applicable law, etc. Those specific terms must not contradict the terms specified above.]*

[legal name of the Auditor]

[name & title of authorised representative]

[dd Month yyyy]

Signature of the Auditor

[legal name of the [Beneficiary] [Linked Third Party]]

[name & title of authorised representative]

[dd Month yyyy]

Signature of the [*Beneficiary*] [*Linked Third Party*]

<sup>1</sup> Supreme Audit Institutions applying INTOSAI-standards may carry out the Procedures according to the corresponding International Standards of Supreme Audit Institutions and code of ethics issued by INTOSAI instead of the International Standard on Related Services (‘ISRS’) 4400 and the Code of Ethics for Professional Accountants issued by the IAASB and the IESBA.

## **Independent report of factual findings on the methodology concerning grant agreements financed under the Horizon 2020 Research and Innovation Framework Programme**

*(To be printed on letterhead paper of the auditor)*

To

[ name of contact person(s)], [Position]  
[[Beneficiary's] [Linked Third Party's] name]  
[ Address]  
[ dd Month yyyy]

Dear [Name of contact person(s)],

As agreed under the terms of reference dated [dd Month yyyy]

with [OPTION 1: [insert name of the beneficiary] ('the Beneficiary')] [OPTION 2: [insert name of the linked third party] ('the Linked Third Party'), third party linked to the Beneficiary [insert name of the beneficiary] ('the Beneficiary')],

we

[ name of the auditor] ('the Auditor'),  
established at [full address/city/state/province/country],  
represented by [name and function of an authorised representative],

have carried out the agreed-upon procedures ('the Procedures') and provide hereby our Independent Report of Factual Findings ('the Report'), concerning the [Beneficiary's] [Linked Third Party's] usual accounting practices for calculating and declaring direct personnel costs declared as unit costs ('the Methodology').

You requested certain procedures to be carried out in connection with the grant(s)

[title and number of the grant agreement(s)] ('the Agreement(s)').

### **The Report**

Our engagement was carried out in accordance with the terms of reference ('the ToR') appended to this Report. The Report includes: the standard statements ('the Statements') made by the [Beneficiary] [Linked Third Party], the agreed-upon procedures ('the Procedures') carried out and the standard factual findings ('the Findings') confirmed by us.

The engagement involved carrying out the Procedures and assessing the Findings and the documentation requested appended to this Report, the results of which the Commission uses to draw conclusions regarding the acceptability of the Methodology applied by the [Beneficiary] [Linked Third Party].

The Report covers the methodology used from [dd Month yyyy]. In the event that the [Beneficiary] [Linked Third Party] changes this methodology, the Report will not be applicable to any Financial Statement<sup>1</sup> submitted thereafter.

The scope of the Procedures and the definition of the standard statements and findings were determined solely by the Commission. Therefore, the Auditor is not responsible for their suitability or pertinence.

Since the Procedures carried out constitute neither an audit nor a review made in accordance with International Standards on Auditing or International Standards on Review Engagements, we do not give a statement of assurance on the costs declared on the basis of the [Beneficiary's] [Linked Third Party's] Methodology. Had we carried out additional procedures or had we performed an audit or review in accordance with these standards, other matters might have come to its attention and would have been included in the Report.

### Exceptions

Apart from the exceptions listed below, the [Beneficiary] [Linked Third Party] agreed with the standard Statements and provided the Auditor all the documentation and accounting information needed by the Auditor to carry out the requested Procedures and corroborate the standard Findings.

**List here any exception and add any information on the cause and possible consequences of each exception, if known. If the exception is quantifiable, also indicate the corresponding amount.**

.....

*Explanation of possible exceptions in the form of examples (to be removed from the Report):*

- i. the [Beneficiary] [Linked Third Party] did not agree with the standard Statement number ... because...;*
- ii. the Auditor could not carry out the procedure ... established because .... (e.g. due to the inability to reconcile key information or the unavailability or inconsistency of data);*
- iii. the Auditor could not confirm or corroborate the standard Finding number ... because ....*

### Remarks

We would like to add the following remarks relevant for the proper understanding of the Methodology applied by the [Beneficiary] [Linked Third Party] or the results reported:

*Example (to be removed from the Report):*

*Regarding the methodology applied to calculate hourly rates ...*

*Regarding standard Finding 15 it has to be noted that ...*

*The [Beneficiary] [Linked Third Party] explained the deviation from the benchmark statement XXIV concerning time recording for personnel with no exclusive dedication to the action in the following manner:*

...

### Annexes

Please provide the following documents to the auditor and annex them to the report when submitting this CoMUC to the Commission:

<sup>1</sup> Financial Statement in this context refers solely to Annex 4 of the Agreement by which the Beneficiary declares costs under the Agreement.

1. Brief description of the methodology for calculating personnel costs, productive hours and hourly rates;
2. Brief description of the time recording system in place;
3. An example of the time records used by the [Beneficiary] [Linked Third Party];
4. Description of any budgeted or estimated elements applied, together with an explanation as to why they are relevant for calculating the personnel costs and how they are based on objective and verifiable information;
5. A summary sheet with the hourly rate for direct personnel declared by the [Beneficiary] [Linked Third Party] and recalculated by the Auditor for each staff member included in the sample (the names do not need to be reported);
6. A comparative table summarising for each person selected in the sample a) the time claimed by the [Beneficiary] [Linked Third Party] in the Financial Statement(s) and b) the time according to the time record verified by the Auditor;
7. A copy of the letter of representation provided to the Auditor.

### Use of this Report

This Report has been drawn up solely for the purpose given under Point 1.1 Reasons for the engagement.

#### The Report:

- is confidential and is intended to be submitted to the Commission by the [Beneficiary] [Linked Third Party] in connection with Article 18.1.2 of the Agreement;
- may not be used by the [Beneficiary] [Linked Third Party] or by the Commission for any other purpose, nor distributed to any other parties;
- may be disclosed by the Commission only to authorised parties, in particular the European Anti-Fraud Office (OLAF) and the European Court of Auditors.
- relates only to the usual cost accounting practices specified above and does not constitute a report on the Financial Statements of the [Beneficiary] [Linked Third Party].

No conflict of interest<sup>2</sup> exists between the Auditor and the Beneficiary [and the Linked Third Party] that could have a bearing on the Report. The total fee paid to the Auditor for producing the Report was EUR [ ] (including EUR [ ] of deductible VAT).

We look forward to discussing our Report with you and would be pleased to provide any further information or assistance which may be required.

Yours sincerely

[legal name of the Auditor]  
[name and title of the authorised representative]  
[dd Month yyyy]  
Signature of the Auditor

---

<sup>2</sup> A conflict of interest arises when the Auditor's objectivity to establish the certificate is compromised in fact or in appearance when the Auditor for instance:

- was involved in the preparation of the Financial Statements;
- stands to benefit directly should the certificate be accepted;
- has a close relationship with any person representing the beneficiary;
- is a director, trustee or partner of the beneficiary; or
- is in any other situation that compromises his or her independence or ability to establish the certificate impartially.



**Statements to be made by the Beneficiary/Linked Third Party ('the Statements') and Procedures to be carried out by the Auditor ('the Procedures') and standard factual findings ('the Findings') to be confirmed by the Auditor**

The Commission reserves the right to provide the auditor with guidance regarding the Statements to be made, the Procedures to be carried out or the Findings to be ascertained and the way in which to present them. The Commission reserves the right to vary the Statements, Procedures or Findings by written notification to the Beneficiary/Linked Third Party to adapt the procedures to changes in the grant agreement(s) or to any other circumstances.

If this methodology certificate relates to the Linked Third Party's usual accounting practices for calculating and claiming direct personnel costs declared as unit costs any reference here below to 'the Beneficiary' is to be considered as a reference to 'the Linked Third Party'.

<i>Please explain any discrepancies in the body of the Report.</i>	
<b>Statements to be made by Beneficiary</b>	<b>Procedures to be carried out and Findings to be confirmed by the Auditor</b>
<p><b>A. Use of the Methodology</b></p> <p>I. The cost accounting practice described below has been in use since /dd Month yyyy/.</p> <p>II. The next planned alteration to the methodology used by the Beneficiary will be from [dd Month yyyy/].</p>	<p><b>Procedure:</b></p> <p>✓ The Auditor checked these dates against the documentation the Beneficiary has provided.</p> <p><b>Factual finding:</b></p> <p>1. The dates provided by the Beneficiary were consistent with the documentation.</p>
<p><b>B. Description of the Methodology</b></p> <p>III. The methodology to calculate unit costs is being used in a consistent manner and is reflected in the relevant procedures.</p> <p><i>[Please describe the methodology your entity uses to calculate <u>personnel costs</u>, productive hours and hourly rates, present your description to the Auditor and annex it to this certificate]</i></p> <p><i>[If the statement of section "B. Description of the methodology" cannot be endorsed by the Beneficiary or there is no written methodology to calculate unit costs it should be listed here below and reported as exception by the Auditor in the main Report of Factual Findings:</i></p> <p>- ...]</p>	<p><b>Procedure:</b></p> <p>✓ The Auditor reviewed the description, the relevant manuals and/or internal guidance documents describing the methodology.</p> <p><b>Factual finding:</b></p> <p>2. The brief description was consistent with the relevant manuals, internal guidance and/or other documentary evidence the Auditor has reviewed.</p> <p>3. The methodology was generally applied by the Beneficiary as part of its usual costs accounting practices.</p>
<p><b>C. Personnel costs</b></p> <p><u>General</u></p>	<p><b>Procedure:</b></p> <p><i>The Auditor draws a sample of employees to carry out the procedures indicated in</i></p>



<b><i>Please explain any discrepancies in the body of the Report.</i></b>	
<b>Statements to be made by Beneficiary</b>	<b>Procedures to be carried out and Findings to be confirmed by the Auditor</b>
<p>IV. The unit costs (hourly rates) are limited to salaries including during parental leave, social security contributions, taxes and other costs included in the remuneration required under national law and the employment contract or equivalent appointing act;</p> <p>V. Employees are hired directly by the Beneficiary in accordance with national law, and work under its sole supervision and responsibility;</p> <p>VI. The Beneficiary remunerates its employees in accordance with its usual practices. This means that personnel costs are charged in line with the Beneficiary’s usual payroll policy (e.g. salary policy, overtime policy, variable pay) and no special conditions exist for employees assigned to tasks relating to the European Union or Euratom, unless explicitly provided for in the grant agreement(s);</p> <p>VII. The Beneficiary allocates its employees to the relevant group/category/cost centre for the purpose of the unit cost calculation in line with the usual cost accounting practice;</p> <p>VIII. Personnel costs are based on the payroll system and accounting system.</p> <p>IX. Any exceptional adjustments of actual personnel costs resulted from relevant budgeted or estimated elements and were based on objective and verifiable information. <i>[Please describe the ‘budgeted or estimated elements’ and their relevance to personnel costs, and explain how they were reasonable and based on objective and verifiable information, present your explanation to the Auditor and annex it to this certificate].</i></p> <p>X. Personnel costs claimed do not contain any of the following ineligible costs: costs related to return on capital; debt and debt service charges; provisions for future losses or debts; interest owed; doubtful debts; currency exchange losses; bank costs charged by the Beneficiary’s bank for transfers from the Commission/Agency; excessive or reckless expenditure; deductible VAT or costs incurred during suspension of the implementation of the action.</p> <p>XI. Personnel costs were not declared under another EU or Euratom grant (including grants awarded by a Member State and financed by the EU budget and grants awarded by bodies other than the Commission/Agency for the purpose of implementing the EU budget).</p>	<p><i>this section C and the following sections D to F.</i>  <i>[The Auditor has drawn a random sample of 10 full-time equivalents made up of employees assigned to the action(s). If fewer than 10 full-time equivalents are assigned to the action(s), the Auditor has selected a sample of 10 full-time equivalents consisting of all employees assigned to the action(s), complemented by other employees irrespective of their assignments.].</i> For this sample:</p> <ul style="list-style-type: none"> <li>✓ the Auditor reviewed all documents relating to personnel costs such as employment contracts, payslips, payroll policy (e.g. salary policy, overtime policy, variable pay policy), accounting and payroll records, applicable national tax , labour and social security law and any other documents corroborating the personnel costs claimed;</li> <li>✓ in particular, the Auditor reviewed the employment contracts of the employees in the sample to verify that: <ul style="list-style-type: none"> <li>i. they were employed directly by the Beneficiary in accordance with applicable national legislation;</li> <li>ii. they were working under the sole technical supervision and responsibility of the latter;</li> <li>iii. they were remunerated in accordance with the Beneficiary’s usual practices;</li> <li>iv. they were allocated to the correct group/category/cost centre for the purposes of calculating the unit cost in line with the Beneficiary’s usual cost accounting practices;</li> </ul> </li> <li>✓ the Auditor verified that any ineligible items or any costs claimed under other costs categories or costs covered by other types of grant or by other grants financed from the European Union budget have not been taken into account when calculating the personnel costs;</li> <li>✓ the Auditor numerically reconciled the total amount of personnel costs used to calculate the unit cost with the total amount of personnel costs recorded in the statutory accounts and the payroll system.</li> <li>✓ to the extent that actual personnel costs were adjusted on the basis of budgeted or estimated elements, the Auditor carefully examined those elements and checked the information source to confirm that they correspond to objective and verifiable information;</li> </ul>

<i>Please explain any discrepancies in the body of the Report.</i>	
<b>Statements to be made by Beneficiary</b>	<b>Procedures to be carried out and Findings to be confirmed by the Auditor</b>
<p><u>If additional remuneration as referred to in the grant agreement(s) is paid</u></p> <p>XII. The Beneficiary is a non-profit legal entity;</p> <p>XIII. The additional remuneration is part of the beneficiary’s usual remuneration practices and paid consistently whenever the relevant work or expertise is required;</p> <p>XIV. The criteria used to calculate the additional remuneration are objective and generally applied regardless of the source of funding;</p> <p>XV. The additional remuneration included in the personnel costs used to calculate the hourly rates for the grant agreement(s) is capped at EUR 8 000 per full-time equivalent (reduced proportionately if the employee is not assigned exclusively to the action).</p> <p><i>[If certain statement(s) of section “C. Personnel costs” cannot be endorsed by the Beneficiary they should be listed here below and reported as exception by the Auditor in the main Report of Factual Findings: - ...]</i></p>	<p>✓ if additional remuneration has been claimed, the Auditor verified that the Beneficiary was a non-profit legal entity, that the amount was capped at EUR 8000 per full-time equivalent and that it was reduced proportionately for employees not assigned exclusively to the action(s).</p> <p>✓ the Auditor recalculated the personnel costs for the employees in the sample.</p> <p><b>Factual finding:</b></p> <ol style="list-style-type: none"> <li>4. All the components of the remuneration that have been claimed as personnel costs are supported by underlying documentation.</li> <li>5. The employees in the sample were employed directly by the Beneficiary in accordance with applicable national law and were working under its sole supervision and responsibility.</li> <li>6. Their employment contracts were in line with the Beneficiary’s usual policy;</li> <li>7. Personnel costs were duly documented and consisted solely of salaries, social security contributions (pension contributions, health insurance, unemployment fund contributions, etc.), taxes and other statutory costs included in the remuneration (holiday pay, thirteenth month’s pay, etc.);</li> <li>8. The totals used to calculate the personnel unit costs are consistent with those registered in the payroll and accounting records;</li> <li>9. To the extent that actual personnel costs were adjusted on the basis of budgeted or estimated elements, those elements were relevant for calculating the personnel costs and correspond to objective and verifiable information. The budgeted or estimated elements used are: — (indicate the elements and their values).</li> <li>10. Personnel costs contained no ineligible elements;</li> <li>11. Specific conditions for eligibility were fulfilled when additional remuneration was paid: a) the Beneficiary is registered in the grant agreements as a non-profit legal entity; b) it was paid according to objective criteria generally applied regardless of the source of funding used and c) remuneration was capped at EUR 8 000 per full-time equivalent (or up to up to the equivalent pro-rata amount if the person did not work on the action full-time during the year or did not work exclusively on the action).</li> </ol>

<i>Please explain any discrepancies in the body of the Report.</i>	
Statements to be made by Beneficiary	Procedures to be carried out and Findings to be confirmed by the Auditor
<p><b>D. Productive hours</b></p> <p>XVI. The number of productive hours per full-time employee applied is <i>[delete as appropriate]</i>:</p> <p>A. 1720 productive hours per year for a person working full-time (corresponding pro-rata for persons not working full time).</p> <p>B. the total number of hours worked in the year by a person for the Beneficiary</p> <p>C. the standard number of annual hours generally applied by the beneficiary for its personnel in accordance with its usual cost accounting practices. This number must be at least 90% of the standard annual workable hours.</p> <p><u>If method B is applied</u></p> <p>XVII. The calculation of the total number of hours worked was done as follows: annual workable hours of the person according to the employment contract, applicable labour agreement or national law plus overtime worked minus absences (such as sick leave and special leave).</p> <p>XVIII. ‘Annual workable hours’ are hours during which the personnel must be working, at the employer’s disposal and carrying out his/her activity or duties under the employment contract, applicable collective labour agreement or national working time legislation.</p> <p>XIX. The contract (applicable collective labour agreement or national working time legislation) do specify the working time enabling to calculate the annual workable hours.</p> <p><u>If method C is applied</u></p> <p>XX. The standard number of productive hours per year is that of a full-time equivalent.</p> <p>XXI. The number of productive hours per year on which the hourly rate is based i) corresponds to the Beneficiary’s usual accounting practices; ii) is at least 90% of the standard number of workable (working) hours per year.</p> <p>XXII. Standard workable (working) hours are hours during which personnel are at</p>	<p><b>Procedure (same sample basis as for Section C: Personnel costs):</b></p> <ul style="list-style-type: none"> <li>✓ The Auditor verified that the number of productive hours applied is in accordance with method A, B or C.</li> <li>✓ The Auditor checked that the number of productive hours per full-time employee is correct.</li> <li>✓ If method B is applied the Auditor verified i) the manner in which the total number of hours worked was done and ii) that the contract specified the annual workable hours by inspecting all the relevant documents, national legislation, labour agreements and contracts.</li> <li>✓ If method C is applied the Auditor reviewed the manner in which the standard number of working hours per year has been calculated by inspecting all the relevant documents, national legislation, labour agreements and contracts and verified that the number of productive hours per year used for these calculations was at least 90% of the standard number of working hours per year.</li> </ul> <p><b>Factual finding:</b></p> <p><u>General</u></p> <p>12. The Beneficiary applied a number of productive hours consistent with method A, B or C detailed in the left-hand column.</p> <p>13. The number of productive hours per year per full-time employee was accurate.</p> <p><u>If method B is applied</u></p> <p>14. The number of ‘annual workable hours’, overtime and absences was verifiable based on the documents provided by the Beneficiary and the calculation of the total number of hours worked was accurate.</p> <p>15. The contract specified the working time enabling to calculate the annual workable hours.</p> <p><u>If method C is applied</u></p> <p>16. The calculation of the number of productive hours per year corresponded to the usual costs accounting practice of the Beneficiary.</p>

<i>Please explain any discrepancies in the body of the Report.</i>	
<b>Statements to be made by Beneficiary</b>	<b>Procedures to be carried out and Findings to be confirmed by the Auditor</b>
<p>the Beneficiary's disposal performing the duties described in the relevant employment contract, collective labour agreement or national labour legislation. The number of standard annual workable (working) hours that the Beneficiary claims is supported by labour contracts, national legislation and other documentary evidence.</p> <p><i>[If certain statement(s) of section "D. Productive hours" cannot be endorsed by the Beneficiary they should be listed here below and reported as exception by the Auditor: - ...]</i></p>	<p>17. The calculation of the standard number of workable (working) hours per year was corroborated by the documents presented by the Beneficiary.</p> <p>18. The number of productive hours per year used for the calculation of the hourly rate was at least 90 % of the number of workable (working) hours per year.</p>
<p><b>E. Hourly rates</b></p> <p>The hourly rates are correct because:</p> <p>XXIII. Hourly rates are correctly calculated since they result from dividing annual personnel costs by the productive hours of a given year and group (e.g. staff category or department or cost centre depending on the methodology applied) and they are in line with the statements made in section C. and D. above.</p> <p><i>[If the statement of section 'E. Hourly rates' cannot be endorsed by the Beneficiary they should be listed here below and reported as exception by the Auditor: - ...]</i></p>	<p><b>Procedure</b></p> <ul style="list-style-type: none"> <li>✓ The Auditor has obtained a list of all personnel rates calculated by the Beneficiary in accordance with the methodology used.</li> <li>✓ The Auditor has obtained a list of all the relevant employees, based on which the personnel rate(s) are calculated.</li> </ul> <p>For 10 full-time equivalent employees selected at random (same sample basis as Section C: Personnel costs):</p> <ul style="list-style-type: none"> <li>✓ The Auditor recalculated the hourly rates.</li> <li>✓ The Auditor verified that the methodology applied corresponds to the usual accounting practices of the organisation and is applied consistently for all activities of the organisation on the basis of objective criteria irrespective of the source of funding.</li> </ul> <p><b>Factual finding:</b></p> <p>19. No differences arose from the recalculation of the hourly rate for the employees included in the sample.</p>
<p><b>F. Time recording</b></p> <p>XXIV. Time recording is in place for all persons with no exclusive dedication to one Horizon 2020 action. At least all hours worked in connection with the grant agreement(s) are registered on a <b>daily/weekly/monthly</b> basis <i>[delete as appropriate]</i> using a <b>paper/computer-based system</b> <i>[delete as appropriate]</i>;</p> <p>XXV. For persons exclusively assigned to one Horizon 2020 activity the</p>	<p><b>Procedure</b></p> <ul style="list-style-type: none"> <li>✓ The Auditor reviewed the brief description, all relevant manuals and/or internal guidance describing the methodology used to record time.</li> </ul> <p>The Auditor reviewed the time records of the random sample of 10 full-time equivalents referred to under Section C: Personnel costs, and verified in particular:</p>

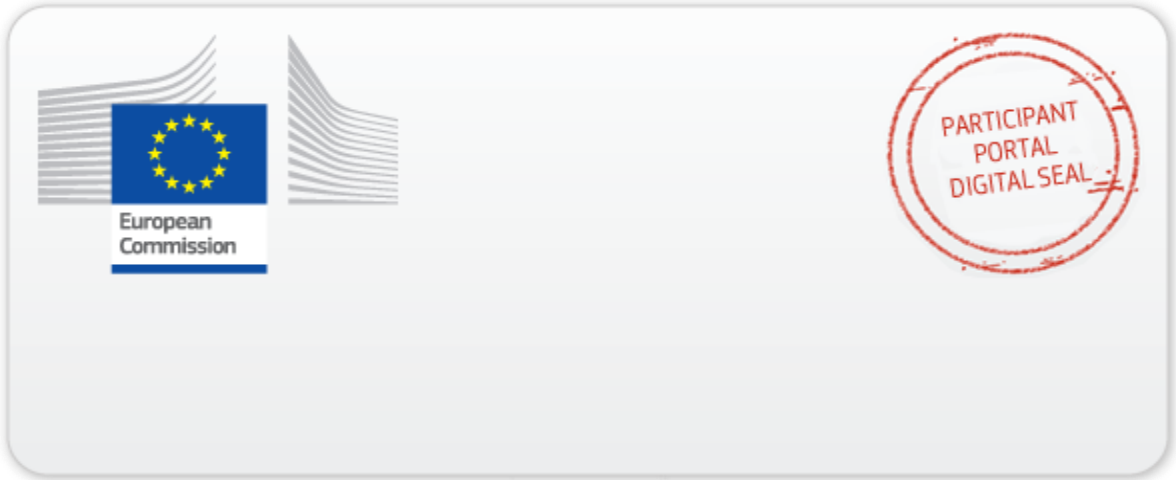
<i>Please explain any discrepancies in the body of the Report.</i>	
<b>Statements to be made by Beneficiary</b>	<b>Procedures to be carried out and Findings to be confirmed by the Auditor</b>
<p>Beneficiary has either signed a declaration to that effect or has put arrangements in place to record their working time;</p> <p>XXVI. Records of time worked have been signed by the person concerned (on paper or electronically) and approved by the action manager or line manager at least monthly;</p> <p>XXVII. Measures are in place to prevent staff from:</p> <ul style="list-style-type: none"> <li>i. recording the same hours twice,</li> <li>ii. recording working hours during absence periods (e.g. holidays, sick leave),</li> <li>iii. recording more than the number of productive hours per year used to calculate the hourly rates, and</li> <li>iv. recording hours worked outside the action period.</li> </ul> <p>XXVIII. No working time was recorded outside the action period;</p> <p>XXIX. No more hours were claimed than the productive hours used to calculate the hourly personnel rates.</p> <p><i>[Please provide a brief description of the <u>time recording system</u> in place together with the measures applied to ensure its reliability to the Auditor and annex it to the present certificate<sup>1</sup>].</i></p> <p><i>[If certain statement(s) of section “F. Time recording” cannot be endorsed by the Beneficiary they should be listed here below and reported as exception by the</i></p>	<ul style="list-style-type: none"> <li>✓ that time records were available for all persons with not exclusive assignment to the action;</li> <li>✓ that time records were available for persons working exclusively for a Horizon 2020 action, or, alternatively, that a declaration signed by the Beneficiary was available for them certifying that they were working exclusively for a Horizon 2020 action;</li> <li>✓ that time records were signed and approved in due time and that all minimum requirements were fulfilled;</li> <li>✓ that the persons worked for the action in the periods claimed;</li> <li>✓ that no more hours were claimed than the productive hours used to calculate the hourly personnel rates;</li> <li>✓ that internal controls were in place to prevent that time is recorded twice, during absences for holidays or sick leave; that more hours are claimed per person per year for Horizon 2020 actions than the number of productive hours per year used to calculate the hourly rates; that working time is recorded outside the action period;</li> <li>✓ the Auditor cross-checked the information with human-resources records to verify consistency and to ensure that the internal controls have been effective. In addition, the Auditor has verified that no more hours were charged to Horizon 2020 actions per person per year than the number of productive hours per year used to calculate the hourly rates, and verified that no time worked outside the action period was charged to the action.</li> </ul> <p><b>Factual finding:</b></p> <p>20. The brief description, manuals and/or internal guidance on time recording provided by the Beneficiary were consistent with management</p>

<sup>1</sup> The description of the time recording system must state among others information on the content of the time records, its coverage (full or action time-recording, for all personnel or only for personnel involved in H2020 actions), its degree of detail (whether there is a reference to the particular tasks accomplished), its form, periodicity of the time registration and authorisation (paper or a computer-based system; on a daily, weekly or monthly basis; signed and countersigned by whom), controls applied to prevent double-charging of time or ensure consistency with HR-records such as absences and travels as well as its information flow up to its use for the preparation of the Financial Statements.

<i>Please explain any discrepancies in the body of the Report.</i>	
<b>Statements to be made by Beneficiary</b>	<b>Procedures to be carried out and Findings to be confirmed by the Auditor</b>
<p><i>Auditor:</i> - ...]</p>	<p>reports/records and other documents reviewed and were generally applied by the Beneficiary to produce the financial statements.</p> <p>21. For the random sample time was recorded or, in the case of employees working exclusively for the action, either a signed declaration or time records were available;</p> <p>22. For the random sample the time records were signed by the employee and the action manager/line manager, at least monthly.</p> <p>23. Working time claimed for the action occurred in the periods claimed;</p> <p>24. No more hours were claimed than the number productive hours used to calculate the hourly personnel rates;</p> <p>25. There is proof that the Beneficiary has checked that working time has not been claimed twice, that it is consistent with absence records and the number of productive hours per year, and that no working time has been claimed outside the action period.</p> <p>26. Working time claimed is consistent with that on record at the human-resources department.</p>

*[official name of the [Beneficiary] [Linked Third Party]]*  
*[name and title of authorised representative]*  
*[dd Month yyyy]*  
 <Signature of the [Beneficiary] [Linked Third Party]>

*[official name of the Auditor]*  
*[name and title of authorised representative]*  
*[dd Month yyyy]*  
 <Signature of the Auditor>



This document is digitally sealed. The digital sealing mechanism uniquely binds the document to the modules of the Participant Portal of the European Commission, to the transaction for which it was generated and ensures its integrity and authenticity.

Any attempt to modify the content will lead to a breach of the electronic seal, which can be verified at any time by clicking on the digital seal validation symbol.