EUROPEAN LANGUAGE GRID

D6.3 Call 2 results and description of selected pilot projects

Jana Hamrlová (CUNI), Lukáš Kačena (CUNI), Jan Hajič (CUNI)

Public 30-04-2021

About this document

Project	ELG – European Language Grid
Grant agreement no.	825627 – Horizon 2020, ICT 2018-2020 – Innovation Action
Coordinator	Dr. Georg Rehm (DFKI)
Start date, duration	01-01-2019, 42 months (GA amendment version: AMD-825627-7)
Deliverable number	D6.3
Deliverable title	Call 2 results and description of selected pilot projects
Туре	Report
Number of pages	38
Status and version	Final – Version 1.0
Dissemination level	Public
Date of delivery	Contractual: 30-04-2021 – Actual: 30-04-2021
WP number and title	WP6: Grid Community – Piloting the European Language Grid
Task number and title	Task 6.2: Organisation and publication of calls
Authors	Jana Hamrlová (CUNI), Lukáš Kačena (CUNI), Jan Hajič (CUNI)
Reviewers	Katrin Marheinecke (DFKI), Penny Labropoulou (ILSP)
Consortium	Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Germany
	Institute for Language and Speech Processing (ILSP), Greece
	University of Sheffield (USFD), United Kingdom
	Charles University (CUNI), Czech Republic
	Evaluations and Language Resources Distribution Agency (ELDA), France
	Tilde SIA (TILDE), Latvia
	Hensoldt Analytics (HENS), Austria
	Expert System Iberia SL (EXPSYS), Spain
	University of Edinburgh (UEDIN), United Kingdom
EC project officers	Philippe Gelin
For copies of reports and other ELG-related information, please contact:	DFKI GmbH European Language Grid (ELG) Alt-Moabit 91c D-10559 Berlin Germany Dr. Georg Rehm, DFKI GmbH georg.rehm@dfki.de Phone: +49 (0)30 23895-1833 Fax: +49 (0)30 23895-1810 http://european-language-grid.eu

₩₩ELG

Table of Contents

List of Figures		3
List of Tables		3
Abstract		4
1	Introduction	4
2	Summary of Open Call 2	5
2.1	Overview and timeline	5
2.2	Evaluation process	6
2.3	Survey for Proposers to the Open Call 2	7
2.4	Current status of Open Call 2	9
2.5	Changes made between Open Call 1 and Open Call 2	9
2.6	Lessons learned	10
3	Project support	10
3.1	Webinar for project holders	11
3.2	Reporting required	11
4	List of selected projects	11
4.1	Introduction and summary	12
4.2	Selected projects	12
4.2.1	Project 1 – Multilingual Image Corpus 2021	12
4.2.2	Project 2 – CEFR Labelling and Assessment Services	13
4.2.3	Project 3 – Motion-Capture 3D Sign Language Resources	15
4.2.4	Project 4 – Universal Semantic Annotator: A Unified API for Multilingual WSD, SRL and AMR	
	annotations	16
4.2.5	Project 5 – Sign language explanations for terms in a text	17
5	Future steps	18
5.1	Execution of selected pilot projects	18
5.2	Evaluation of the final results	19
6	Conclusions	19
Α.	Annexes	20

List of Figures

Figure 1: Open calls timeline	6
Figure 2: Announcement of the second open call results on the ELG website	8

List of Tables

Table 1: Proposals submitted for the second open call and accepted for evaluation	6
Table 2: Proposals selected for financial support in the second open call	6
Table 3: List of pilot projects selected for financial support in the second open call	11



Abstract

This document describes the selection process and results of the second open call for ELG pilot projects, which should broaden ELG's portfolio of Language Technologies (LT) and demonstrate the usefulness of the ELG not only as a technology platform but also as a European project and initiative. It follows up on deliverable D6.1, where the detailed description of organizing the submission and evaluation process of the first open call is provided and deliverable D6.2 which described the selection process and results of the first open call for ELG pilot projects. The rich experience from the first open call was reflected in the preparation for the second open call.

Like the first open call, we consider the second open call for ELG pilot projects as very successful, with 103 submitted project proposals. In both calls combined, we received 213 project proposals from 156 different institutions (86 SMEs, 70 research organizations) coming from 32 different countries (including nine eligible countries outside the European Union). This was much more than expected, and it clearly shows a great interest in Language Technologies and the European Language Grid initiative.

A meticulous evaluation and selection process resulted in five pilot projects selected for funding. As of now, all of them have successfully started the first execution phase, called the "Experiment". Continuous support for the projects is provided, including a webinar with Pilot Board members, and there will also be continuous guidance by the assigned project coach for each project.

The structure of this document is as follows: After the Introduction, a brief summary of the second open call is provided in Chapter 2 to review the basic information about the call and the evaluation process, together with the description of the selection process.

Chapter 3 provides a description of additional support for the five selected pilot projects during their execution phase. The projects are presented in detail in Chapter 4. Finally, in Chapter 5, future steps regarding the pilot project from both open calls are laid out.

1 Introduction

To demonstrate the use and the advantages of ELG, in providing basic LT for applications, and as a basis for more advanced LT-based modules or components useful to industry, the ELG project set up a mechanism for using close to 30% of the overall project budget for small scale demonstrator projects ("pilots") through two open calls. We have designed and prepared them on the basis of the ICT-29a call specification using the Financial Support to Third Parties (FSTP) scheme, according to Annex K of the ICT Work Programme 2018–2020¹. In sum, we are providing 1,950,000€ to the selected projects as FSTP with an awarded amount of up to 200,000€ per single project. To set up the open calls, we have established a lightweight submission procedure and an open and transparent evaluation process, in which external expert evaluators participate as reviewers.

The main objective of the open calls is to attract SMEs and research organizations to either

(a) contribute tools and services to the core ELG platform (pilot projects of type A) or

 $^{^{1}\} https://ec.europa.eu/research/participants/data/ref/h2020/other/wp/2018-2020/annexes/h2020-wp1820-annex-k-fs3p_en.pdf$

(b) develop applications using Language Technologies available in the ELG platform (pilot projects of type B).

The project results will be included in the ELG platform to allow for wide dissemination and testing and external evaluation by other entities and/or the public. The ELG project and, later on, the long-term initiative, will also provide further access to promotion and dissemination events.

2 Summary of Open Call 2

This chapter provides the overall overview of the second open call, beginning from basic information about the call to the description of its current status, further followed by an analysis of the feedback supplied so far by all involved parties.

2.1 Overview and timeline

The second open call was opened on 1 October 2020 and closed on 30 November 2020 (23:59 CEST) in accordance with the open calls timeline (see Figure 1). In total, we accepted 103 project proposals for evaluation:

- 60 proposals were submitted by SMEs (28 of type A, 32 of type B); 56 unique SMEs.
- 43 proposals were submitted by research organizations (38 projects of type A, 5 of type B); 42 unique research organizations.

Five applicants (four SMEs and one research organization) submitted two proposals (one type A and one type B). Regarding the type of project, 66 submitted proposals were of type A (contribute resources, services, tools, or data sets to the ELG to increase its coverage), and 37 project proposals were of type B (develop applications using language resources and technologies available in the ELG). 43 applicants who submitted a project proposal in the second open call indicated that they had submitted the same or a similar proposal in the first open call.

We received applications from 28 countries, including eligible countries outside the EU (Iran, Israel, Norway, Serbia, Switzerland, Turkey, United Kingdom).

The total amount of financing requested by the submitted projects was 13,257,919€. The average amount requested per project was 129,000€ which is less than in the first open call (153,000€).



Figure 1: Open calls timeline

Submitted by	Туре А	Туре В	Total
Research organization	38 (36.9%)	5 (4.8%)	43 (41.76%)
SME	28 (27.2%)	32 (31.1%)	60 (58.3%)
Total	66 (64.1%)	37 (35.9%)	103 (100%)

Table 1: Proposals submitted for the second open call and accepted for evaluation

Submitted by	Туре А	Туре В	Total
Research organization	3 (60%)	0 (0%)	3 (60%)
SME	0 (0%)	2 (40%)	2 (40%)
Total	3 (60%)	2 (40%)	5 (100%)

Table 2: Proposals selected for financial support in the second open call

2.2 Evaluation process

The evaluation process was executed as planned (see Figure 1) during December 2020 and January 2021 and finished in early February 2021. Each submitted project proposal was evaluated by three independent external evaluators to ensure an open, transparent, and expert evaluation-based selection process. The whole process was monitored by the Pilot Board (PB), which was set up for the supervision of the pilot projects.

After the proposals had been evaluated by external experts, the project coaches (who are all members of the Pilot Board) prepared summary evaluation reports for the project proposals assigned to them and submitted them via the ELG Open Calls Platform, where they were accessible at the platform's dashboard to all Pilot Board members. Each of the five project coaches summarized 20 project proposals in this way.

The form used for the summary evaluation reports is attached as Annex 10 to D6.1². In this summary report, the project coaches first reviewed the three evaluation reports submitted by the external evaluators (items summary of evaluations, project proposal assessment, summary, and recommendation for financing). Then

² D6.1 Pilot Calls Setup



they eventually suggested budget adjustments and changes of the total number of points assigned to the proposal in range of at most 45 points (up or down), where applicable. Further, according to the Evaluation Criteria (Annex 7 of D6.1), project proposals by SMEs developing applications using language technology that is already available in the ELG (B type projects) received 30 bonus points. Finally, the project coaches reviewed the Eligibility criteria (uniqueness, relevance for ELG, and project phases) as checked by the evaluators and suggested their decision on their fulfilment if the evaluators differed in their opinions.

The project coaches also assessed the performance and the quality of the reports submitted by the evaluators. This assessment was entered into a separate part of the summary evaluation report. After all summary evaluation reports had been submitted by the project coaches to the ELG Open Calls Platform, the Pilot Board selection meeting was convened. It took place online on 3/4 February 2021. The goals of the meeting were:

- To select the best projects that deserve to be funded.
- To rank the selected projects and decide how many projects will be funded.
- To agree on the next steps.

The Pilot Board went through the projects top-down, starting with the best-rated project. For each project proposal, the project coach briefly introduced the project and presented the main conclusions from its summary evaluation report. A discussion within the Pilot Board followed, concluding with a decision of the Pilot Board on:

- Fulfilment of eligibility criteria.
- Points awarded by the Pilot Board (+/- 45 points, as suggested by the project coach, possibly modified by the Board).
- Is the project good enough to be approved for financing?
- Is any budget modification (reduction, structural change) suggested to the project? (This was only discussed for selected projects only, see below.)

The low-ranked projects were not discussed unless there was a request by a Pilot Board member to do so.

Finally, the proposals were ranked by the total sum of assigned points. The ranked list was cut at the maximum available support (585,000€). Thus, 5 project proposals were selected for funding amounting to a funding sum of 584,396€ in total. The highest-ranking non-selected proposal had received 283 points out of the absolute maximum of 375 points (300 by external evaluators + 30 bonus to SMEs submitting B type project + 45 by project coach/Pilot Board). The lowest-ranking selected proposal had received 284 points, whilst the best project had received 305 points.

2.3 Survey for Proposers to the Open Call 2

Similarly, as for the open call 1, a short survey was designed for those who submitted a proposal (proposers). The survey comprised 15 questions, a mix of open questions and multiple-choice questions. The survey topics were clustered into three sections: "motivation", "project proposals", "your organization". All information was collected anonymously with the goal to evaluate and improve our call processes. Of the proposers, 39 out of 103 (38 %) responded.



Regarding motivation to submit a proposal to the ELG open call, contributing services or resources to the ELG platform to make them available to the emerging ELG community and further development of an existing software or data project were the most frequent reasons reported by the respondents. Main expectations toward ELG were that ELG provides increased visibility for applicant's organization on the European level and to get access to a large repository of functional services and datasets. Also, almost all respondents think that more EU-funded activities dedicated to Language Technology and Language-centric AI are needed, preferably in form of agile calls (with short proposals and quick evaluations, 9–12 months project runtime).

- WI	EUROPEAN LANGUAGE GRID About *	Grid • Community • Open Calls • Eve	nts • News Cons	ortium Contact	
		HHHH	HH.	H	Ħ
	Open Call #	2			
	The second ELG open call for pilot p total, 103 project proposals were ac countries. All proposals were evalu: ELG Pilot Board. The Pilot Board sel EUR 584,396 in total.	rojects was closed on 30 November 2020. In cepted for evaluation with applicants from 28 ted by three independent experts and by th ected five projects amounting for a funding o	a f	2	
	The second open call was also the l project. In both calls combined, we countries amounting for almost 2 n research organisations and five SM tributing resources, services, tools, platform to increase its coverage, w using language resources and tech	ast one funded within the runtime of the ELG supported 15 projects from eight different illion EUR. These projects are executed by te is. There are eleven projects that focus on co or data sets to the European Language Grid hist four projects are developing application nologies available in the ELG platform.	Call Documer Guide for App Third Party A Project Propo S, Evaluation Cr FAQ (Frequer	ntation olicants greement isal Template iteria itly Asked Questions)
	Congratulations to all successful ap projects and all the effort dedicatec ceive a brief feedback on their proje	plicants and a big thank you for all submittec to the second open call. All applicants will re ect proposal in the upcoming weeks.	1		
	Selected projects				
	Organisation	Project Title	Country	Funding awarded	
	Institute for Bulgarian Language "Prof. Lyubomir Andreychin"	Multilingual Image Corpus 2021	Bulgaria	EUR 110,960	
	EDIA BV	CEFR Labelling and Assessment Services	Netherlands	EUR 137,560	
	University of West Bohemia	Motion Capture 3D Sign Language Resources	Czechia	EUR 85,421	
	Sapienza University of Rome	Universal Semantic Annotator: A Unified API for Multilingual WSD, SRL and AMR annotations	Italy	EUR 113,228	
	Sign Time GmbH	Sign language explanations for terms in a text	Austria	EUR 137,227	

The first rejected proposal has received 283 out of 375 points. Learn more about

the evaluation criteria here.

Figure 2: Announcement of the second open call results on the ELG website

Regarding the specialization of respondents, most frequently they specialize in text analytics, machine translation and speech recognition. Respondents reported more than twenty domains that they specialize in (most frequently health sector), one fourth of all respondents have no particular specialization. Full analysis of the survey is part of the Annex 1. **Current status of Open Call 2**

2.4



In February, after the Pilot Board selection meeting, the results of the second open call were announced on the ELG website³ (see Figure 2), including the list of projects selected for funding. Both selected and not selected project leaders were informed about the results.

During February and March 2021, contracts with all the selected projects (see Annex 5 of D6.1 – Third Party Agreement) were signed, and the first payments in the amount of half of the awarded financial support were made, in line with the previously approved call documentation and procedures. All projects started their execution phase by 1 April 2021.

In March, résumés from the summary evaluation report (items summary of evaluations, project proposal assessment, summary, recommendation for funding, and total points assigned) were made accessible through the ELG Open Calls Platform to the individual applicants.

2.5 Changes made between Open Call 1 and Open Call 2

Based on the recommendations discussed in D6.2, the basic parameters given by the ELG grant agreement remained the same for the second open call: We looked for pilot projects that would broaden the ELG's portfolio of Language Technologies. The projects would develop missing services or solutions that support underrepresented languages. At the same time, the pilot projects should demonstrate the usefulness of the European Language Grid as a technology platform. The amount awarded to each project could be up to 200,000€, and the runtime of the projects was expected to be between 9 and 12 months. SMEs and research organisations were eligible to apply, and the projects were to be non-consortial, i.e., only one single organisation or institution was allowed per project.

As suggested in D6.2, we implemented these changes in the call documentation and the open call procedure:

- We improved the explanation of the strategic goals of ELG and the goals of the open calls. Links to an overview of ELG, its history and context and to an overview of the ELG project and the platform were provided in the call documentation.
- We also improved the description of the ELG infrastructure and provided an easy-to-find list of currently available services this was done with the launch of Release 1 of the ELG in the process of which a detailed technical documentation was made available.
- We organized a webinar for proposers. It took place during the submission period, on 12 November 2020. We explained the goals of the open call and briefly presented the call documentation and its annexes. The second part of the webinar was dedicated to questions and discussion. A recording of the meeting and the presentation were made available to all registered project managers.
- The documentation, annexes, templates, and forms along with the Open Calls Platform were fine-tuned.
- Budget breakdowns were requested in a fixed structure as well as a more detailed budget justification.
- New evaluators were recruited and added to the current group of evaluators, with the aim to attract more high-quality and experienced evaluators.
- It was decided that the second open call, like the first open call, should have no specific thematic focus.

//'ELG

³ https://www.european-language-grid.eu/open-calls/open-call-2

2.6 Lessons learned

The following lessons learned were gathered by the Management Team:

- As stated in D6.2, in the ELG project, costs for the open calls platform and for the evaluation of project proposals should have been more carefully planned. Especially, higher remuneration for evaluation reports could have attracted more highly experienced evaluators.
- The quality of evaluation reports submitted by external evaluators was not stable and, in some cases, could have been more profound. This was usually balanced by the project coach or Pilot Board. Lesser quality of some evaluation reports put more burden on the project coach and the Pilot Board, though.
- We aimed at a simple and light-weight procedure which led, in our opinion, to a high number of submitted proposals. At the same time, the simplicity of the project template may have led to an increased number of low-quality proposals that were not adequately described and thought through. Such a high number of project proposals in both calls was demanding – required more person days and increased costs related to external evaluators.

3 Project support

Once the five pilot projects were selected and the contract signed, the continuous support from the ELG consortium started so that the projects can successfully proceed to the project execution. The project execution itself consists of three phases (cf. call documentation in Annex 3 of D6.1):

- Phase 1: Experiment
- Phase 2: Integration
- Phase 3: Dissemination

After finishing Phase 1, reporting from the applicants will be required, and the Pilot Board will decide whether the project will be allowed to continue execution (and consequently, whether the next payment is made). After finishing Phase 3, a final report will be required, and the Pilot Board will evaluate the whole project and decide whether the project receives the final payment.

Each selected project is supervised by one member of the Pilot Board (project coach). The project coach will be responsible for training the project team, collecting and answering questions from the team during project execution, collecting reports, and guiding the project team through all project phases, especially through the Integration and Dissemination phases (if the project is allowed to progress to them).

To advertise the selected projects to a wider public, we plan to present the pilot projects at the ELG conference, the (virtual) META-FORUM 2021 in autumn. There will be a dedicated time slot where all project leaders can present their project idea and explain the goals they are aiming for. In 2022, when the META-FORUM can hopefully be held again as a face-to-face event, the pilot projects should show their results and demonstrate the usefulness of ELG. The projects may assign a part of their dissemination budget to these events.

In addition, there are partners in all European countries who support ELG and act as a bridge between the ELG project and the local players in the field of Language Technology. In many cases, dissemination and training events provided by these National Competence Centres (NCCs) can also be a good opportunity to present pilot

III ELG

projects in specific regions. The actual implementation and design of the dissemination measures will be determined and planned in each individual case with the project coach and the event organizers.

3.1 Webinar for project holders

The first opportunity where the newly selected pilot projects could become familiar with ELG was an online meeting with the Pilot Board on 23 March 2021. During the meeting, basic information about the ELG and its technology, as well as guidelines for project execution, were presented. Prior to the meeting, the Guide for pilot projects (see Annex 2) was sent to all project leaders.

3.2 Reporting required

During the project execution, two reports will be required from the pilot projects. They will be provided with a report template, and the reports will be submitted via the ELG Open Calls platform.

The first evaluation of every project will be performed after Phase 1 ("Experiment"). The project coach will assess the progress of the project and propose to the Pilot Board to approve the second payment (which amounts to 35% of the total requested financial support), or to terminate the project.

The final evaluation will be performed after Phases 2 ("Integration") and 3 ("Dissemination"). The project coach will assess the fulfilment of the project's obligations in these two phases and prepare a short report (to be made public). Along with the report, a recommendation will be made by the project coach to the Pilot Board to approve (or not) the final (third) payment to the project (in the amount of 15% of the total requested budget).

After the project finishes, the project team is required to present their results, business plans, secured venture capital for further development, and future plans. The Pilot Board will assess the finished projects and evaluate the immediate results. It will also formulate recommendations for sustainability and future operation of ELG based on the experience of and with the pilot projects.

4 List of selected projects

This chapter provides a detailed description of the five projects selected for financial support in the second open call. All pilot projects are summarized in Table 3.

Organization	Legal Form	Project Name	Project Type	Coun try	Funding Awarded
Institute for Bul- garian Language	RO	Multilingual Image Corpus 2021	А	BG	110,960€
EDIA BV	SME	CEFR Labelling and Assessment Services	В	NL	137,560€
University of West Bohemia	RO	Motion-Capture 3D Sign Language Resources	А	CZ	85,421€
Sapienza Univer- sity of Rome	RO	Universal Semantic Annotator: A Unified API for Multilingual WSD, SRL and AMR annotations	А	IT	113,228€
Sign Time GmbH	SME	Sign language explanations for terms in a text	В	AT	137,227€

Table 3: List of pilot projects selected for financial support in the second open call



4.1 Introduction and summary

Although we obtained more project proposals from SMEs than from research organizations, there are 2 SMEs and 3 research organizations among the projects selected. Similarly, only two B type projects were accepted for financing which probably reflects the fact that the ELG platform is still being developed and thus it is more logical to create missing resources or tools than build applications using resources and tools already available on the platform.

Three A type projects aim at providing tools to enrich the ELG platform. One project will contribute multilingual annotated data, tools and services for image processing whilst the second one aims at improving the ELG offer of linguistic tools by proposing a unified service powered by state-of-the-art neural models for carrying out annotations on three tasks of Natural Language Understanding, i.e., Word Sense Disambiguation, Semantic Role Labelling and Semantic Parsing, in around 100 languages. The third A type project will expand the portfolio of language resources available in ELG by adding a dataset and a search tool that are for the Czech sign language.

Regarding the B type projects, one of the projects also deals with sign language. Its goal is to make texts easier to comprehend for deaf people by linking texts to a sign language encyclopaedia. The other project aims to develop a set of tools, datasets, and services to enable automatic classification of the reading difficulty of texts on the Common European Framework of Reference.

The awarded budget varies from 85,421€ to 137,227€. All supported organizations are from the EU and from five various countries – Austria, Bulgaria, Czechia, Italy and the Netherlands.

4.2 Selected projects

In this section, the individual supported projects will be introduced. After the Project Name and Project Acronym, the applying Organization is introduced, followed by Project Abstract, the Total Budget and Type of project. Finally, the Project Schedule is described and divided into the three execution phases. Project Keywords are mentioned at the end of each project section.

4.2.1 Project 1 – Multilingual Image Corpus 2021

Project Name: Multilingual Image Corpus 2021

Project Acronym: MIC21

Organization: Institute for Bulgarian Language "Prof. Lyubomir Andreychin"

The Institute for Bulgarian Language Prof. Lyubomir Andreychin (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.

Project Abstract:

The aims of the project Multilingual Image Corpus 2021 are to contribute multilingual annotated data, tools and services for image processing to the ELG platform. In particular, the project will deliver:

1. Large data set containing images and their descriptions in (at least) 20 European languages and providing free access to the data given the strict observation of copyright.



2. Tools for automatic identification, classification and annotation of objects in images and classification and annotation of images.

3. Services for automatic identification, classification and annotation of objects in images and classification and annotation of images.

Total Budget: 110,960€

Type of project: A

Project Schedule:

The duration of the project will be 12 months.

Phase 1: Experiment (Month 1–8)

- T1.1 Development of MIC21 (M1–M8) Compilation of large data set of images; Automatic annotation and manual evaluation of MIC21MS 1
- MIC21 developed (M8)
- T1.2 Processing images (M1–M8) Development tools for automatic identification, classification and annotation of (objects in) images; Testing Fine-tuning
- MS 2 Tools for processing images available (M8)

Phase 2: Integration (Month 9–12)

- T2.1 Building the integrated services for automatic image classification and annotation; RESTful API; Web application; Docker containerisation
- MS 3 The integrated services built (M10)
- T2.2 Integration with ELG platform (M11–M12) Delivering annotated data sets Integration containerised services with ELG platform
- MS 4 MIC21 is integrated with ELG platform (M12)

Phase 3: Dissemination (Month 1–12)

- T3.1 Dissemination and visibility activities
- T3.2 Identifying potential users (two workshops) MS 5 workshops with stakeholders organized (M7–12)
- MS 6 Business and sustainability plan (M12)

Project Keywords: multilingual image corpus, image classification, image annotation

4.2.2 Project 2 – CEFR Labelling and Assessment Services

Project Name: CEFR Labelling and Assessment Services

Project Acronym: CLASS

Organization: EDIA BV

EDIA is on a mission to make educational content suited for personalised learning. Content needs to become increasingly modular and subsequently aligned with learning objectives to improve accessibility and discoverability of educational content so that each student can use the content most suitable to their needs.



Towards this goal, EDIA BV applies machine learning and natural language processing to automatically extract metadata from authentic (e.g., news articles) and proprietary (e.g., books) content.

Project Abstract

This project aims to develop a set of tools, datasets, and services to enable automatic classification of the reading difficulty of texts on the Common European Framework of Reference (CEFR). Being able to accurately and consistently check the readability level of texts is important to authors and teachers. It will allow them to create and discover content at the right level for learners with many different backgrounds and abilities.

As part of the project scope, multiple languages will be supported. Furthermore, the project will deliver an infrastructure that paves the way for adding additional languages in the future.

Total Budget: 137,560€

Type of project: B

Projects Schedule:

The project will require 12 calendar months. Tasks may be performed in parallel or overlap in calendar time.

Phase 1: Experiment (Months 1–12)

- MS1.1 Unlabelled dataset (M3)
- MS1.2 Data labelling tool (M3)
- MS1.3 Labelled dataset (M6)
- MS1.4 CEFR Classification models (M7)
- MS1.5 CEFR Word List (M8)
- MS1.6 CEFR Service (M9)
- MS1.7 Authoring tool (M12)

Phase 2: Integration (Months 4–12)

- MS2.1 Integrated data labelling tool (M4)
- MS2.2 Integrated CEFR service (M10)
- MS2.3 Integrated authoring tool (M12)

Phase 3: Dissemination (Months 1–12)

- MS3.1 Project update I (M4)
- MS3.2 Project update II (M10)
- MS3.3 Project update III Final (M12)
- MS3.4 IATEFL Conference report (EN) (M4)
- MS3.5 Conference report FBM (DE) (M8)
- MS3.6 Conference report (NL) (M10)
- MS3.7 Conference report (ES) (M10)

Project Keywords: Common European Framework of Reference (CEFR), Readability, Content Authoring, Language Learning

₩ELG

4.2.3 Project 3 – Motion-Capture 3D Sign Language Resources

Project Name: Motion-Capture 3D Sign Language Resources

Project Acronym: MC-TRISLAN

Organization: University of West Bohemia

University of West Bohemia (UWB) is a university founded in 1991. UWB has 9 faculties, 62 departments, 130 study programs, 11,000 students, modern university campus, 4 research centres.

The team applied for the grant are academics and researchers of the Faculty of Applied Sciences (FAS) and its European Centre of Excellence: New Technologies for Information Society (NTIS). NTIS has 6 research programs, among which the program "Development of cybernetic control systems, identification, intelligent decision-making and communication system" is closely connected with the Department of Cybernetics (DC). All members of the team are members of DC's section of Artificial Intelligence (AI) with the main focus on speech and Language Technologies and have many years of experience in sign language processing.

Project Abstract

The project expands the portfolio of language resources available in ELG by adding a dataset and a search tool that are for Czech sign language. The project solves the current problem which is a critical lack of quality data necessary for research and subsequent deployment of machine learning techniques in this area. As an output, the project currently provides the largest collection of sign language acquired by state-of-the-art 3D human body recording technology and enables the successful deployment of communication technologies in the future, especially machine translation and sign language synthesis.

Total budget: 85,421€

Type of project: A

Project Schedule:

Phase 1: Experiment (Months 1–11)

- Preparation phase (recording set-up, subject and domain selection) (M1)
- Data acquisition, data processing (M2–M7)
- Data post-processing (M5–9)
- Dataset validation (M9–10)
- Search algorithm (SA) development (M3–9)
- Dataset annotation (M7–11)

Phase 2: Integration (Months 11–12)

- Upload/integration of SA (M11–12)
- Dataset publishing and upload to ELG platform (M12)

Phase 3: Dissemination (Months 3-12)

• Website of the applicant's workplace (M3–4)

₩ELG

- Paper preparation for SPECOM 2021 (M5-6)
- Paper preparation for LREC 2022 (M11–12)

Project Keywords: Sign Language; Czech Language; Data Resource; Motion capture

4.2.4 Project 4 – Universal Semantic Annotator: A Unified API for Multilingual WSD, SRL and AMR annotations

Project Name: Universal Semantic Annotator: A Unified API for Multilingual WSD, SRL and AMR annotations

Project Acronym: USeA

Organization: Sapienza University of Rome

The Sapienza University of Rome is the highest-ranked academic institution in Italy (CWUR) with several excellence departments. Among them, the Department of Computer Science ranked 1st in the area of Mathematics and Computer Science and won a 6.5M€ Excellence Grant. Sapienza NLP (http://nlp.uniroma1.it) is a longstanding research laboratory in the Computer Science Department, headed by Prof. Roberto Navigli, who has two decades of experience in the fields of computational lexical semantics and multilinguality and is the recipient of two ERC awards. The group also manages several projects, i.e., ELEXIS for a European electronic lexicography infrastructure and other national and international projects (totalling more than 3.5M€).

Project Abstract

While already providing several resources, ELG still lacks a comprehensive service performing semantic annotations across different languages. With this project, we aim at improving the ELG offer of linguistic tools by proposing a unified service powered by state-of-the-art neural models for carrying out annotations on three tasks of Natural Language Understanding, i.e., Word Sense Disambiguation, Semantic Role Labelling and Semantic Parsing, in around 100 languages. Our service will reengineer and integrate the most advanced works from our group and make them available to the community via an easy-to-use, unified service that will be accessible through REST APIs, web browsers and programmatically. The ELG tool kit will therefore be enriched with a unique, novel and free-to-use service.

Total Budget: 113,228€

Type of project: A

Project Schedule

Phase 1: Experiment (Months 1–11)

- We hire two engineers and design the API (M1-2)
- MS 1: The API interface is implemented
- We reengineer the code for EWISER, ARES, InVeRo and XL-AMR and seal them into separate Docker containers (M3–5)
- MS 2: Models have been reengineered
- We train models on different datasets and measure their performances at inference time in terms of quality, speed and memory consumption on standard benchmarks and by means of stress tests (M6–7)
- MS 3: A model for each task is trained and can serve queries



- We develop a single-entry point exposing REST APIs to query the models and to put them in the pipeline. We also create a web interface to access it via browser (M8–11)
- MS 4: Full system and Web interface are accessible

Phase 2: Integration (Month 12)

- We integrate our systems into ELG (M12)
- MS 5: Our systems are accessible via the ELG website

Phase 3: Dissemination (Months 8–12)

• A workshop on NLU and a demo paper to a top-notch NLP conference. We will also organize a hackathon (M8–12)

Project Keywords: Natural Language Understanding, REST API, Word Sense Disambiguation, Semantic Role Labelling, Semantic Parsing, Multilinguality

4.2.5 Project 5 – Sign language explanations for terms in a text

Project Name: Sign language explanations for terms in a text

Project Acronym: SignLookUp

Organization: Sign Time GmbH

Sign Time developed a semi-automatic system to translate text into sign language. A 3D-animated avatar is used. The development was funded by H2020 SME Instrument funding (Phase 1 and 2) and by national R&D funding from Austria. For a first impression of our work see our showreel at <u>https://vimeo.com/335880733</u>

Sign Time produces videos in animated sign language for different customers in Europe with a strong focus on the German market. Besides the daily business the company is also working on associated research questions, for instance regarding methods for automatic translations for this very special quality of languages etc. The research field not only includes ICT technical but also linguistic questions. The production of content in sign language includes teaching materials and cultural content and especially citizen information. Most of the translation videos produced are available online.

Project Abstract

The project SignLookUp serves the goal to make texts easier to comprehend for deaf people. Deaf people have a difficult access to texts, learning a written language is a challenge with a hearing impairment. Therefore about 75% of deaf people are functional illiterates. SignLookUp is a technology that links texts to a sign language encyclopaedia. Deaf people can then click on difficult or unknown terms in a text and immediately receive the explanation or description of the word in their sign language which is displayed adjacent to the text. SignLookUp starts with two sign languages but is built to be easily expanded. The product will be licensed for companies and free for deaf people. This technology helps to make textual information on the internet accessible for deaf people.

Total Budget: 137,227€

Type of project: B

յ∦[∥]ELG

Project Schedule

Phase 1: Experiment (Months 1–10)

- Final IT-architecture definition, environment set-up and detailed project planning (M1–2)
- Agile development and test of SignLookUp functions (M3–7)
- Identification of basic terms, explanations, translation into sign language, animation (M4–10)
- Selection of test-customer and test-user (M9–10)
- Agile development and test of SignLookUp additional & supporting functions and bug fixing (M8–10)

Phase 2: Integration (Months 10–11)

- Make application available on ELG (M9–10)
- Final Test (M11)

Phase 3: Dissemination (Months 11–12)

- Contract development (M11)
- Preparation of campaigns (press, social media), proposals for awards and prizes, internet advertising (M11–12)

Milestones

- Detailed project plan available: M1
- First version of sign-lexicon (up to 300 terms per sign language) and program ready: M7
- Test-Version & sign-lexicon ready: M10
- Market-ready lookApp and published on ELG: M11
- Social media and press campaign starting M11

Project Keywords: Sign language, accessibility, deaf, sustainability, avatar, animation, lookup, translation, lexicon, encyclopaedia, SignLookUp

5 Future steps

5.1 Execution of selected pilot projects

All pilot projects will be executed in accordance with the approved project proposal and the Third Party Agreement. Pilot projects are offered guidance and supervision throughout the lifecycle of the project once the contract is in place and the project has started. Control days will be organized (remotely) when projects will have to report progress, any problems, state of completion, etc.

The Pilot Board is set up for the supervision of the pilot projects. It provides a forum so that the ELG project can discuss the progress of the pilots, their intermediate feedback and the results.

The Pilot Board consists of ELG consortium members. It is the main technical and strategic interface between the pilot projects and the ELG project so that the project can maximise its benefit from supporting the pilots and also to make sure that the pilot projects maximally benefit from the European Language Grid.

₩^IELG

Each selected project is supervised by one member of the Pilot Board (Project Coach) appointed by the Pilot Board. The Project Coach is responsible for:

- monitoring the project planning and its progress,
- coordinating the training of the Awardee's project team with respect to the ELG potential,
- collecting and answering questions from the team during the execution of the project with the support of the ELG partners,
- collecting reports and guiding the project team through all activities (phases), especially through the Integration and Dissemination phases
- recommending to the Pilot Board if the project should be allowed to progress to the next phase.

The assessment will be coordinated with the ELG project consortium; the Project Coach will seek especially technical help from the consortium to evaluate the pilot project results, especially in terms of technical testing to determine if the objectives of the (2) Experiment phase have been met. The Project Coach will assess the progress of the project and propose to the Pilot Board to approve the second payment to the Awardee, or to terminate the project after the (2) Experiment phase. The pilot projects will be provided with report template, reports will be submitted via ELG Open Calls platform.

5.2 Evaluation of the final results

The final evaluation of a project will be performed by the Project Coach after the (3) Integration activity and if the project fulfils the (4) Dissemination obligations. The Project Coach will then prepare a short report (to be made public) and recommend to the Pilot Board to approve (or not) the final (third) payment to the project Awardee. The pilot projects will be provided with a template, reports will be submitted via ELG Open Calls platform. After a project has finished, the project team is required to present their results, business plans, secured venture capital for further development and future plans. The Pilot Board will assess the finished projects and evaluate the immediate results. It will also formulate recommendations for sustainability and future operation of the ELG based on the experience of and with the pilot projects.

6 Conclusions

The proposals submitted to the second open call were evaluated, five projects were selected for financial support, and these projects have successfully started their execution phase.

In Table 1: Proposals submitted for the second open call and accepted for evaluationTable 1, the type of submitted projects and applying organizations are shown. For projects selected for financing, a similar overview is provided in Table 2. The selected organizations are from five different EU countries. The requested budget varies from 85,421€ to 137,560€, with an average value of 116,879€.

All pilot projects from both open calls are currently being executed. After phase 2 (Experiment) they will undergo a mid-term review. Then, the final evaluation of the results will be performed. The Pilot Board and other members of the ELG consortium guide and supervise the projects throughout their lifecycle.



A. Annexes

	Document	Description
1	Analysis of the Survey for Proposers to the Open Call 2	Results of a short survey designed for those who submitted a proposal (proposers).
2	Guide for pilot projects	The document which was sent to the pilot projects at the beginning of the execution phase. This document contains (i) a guide and basic information about the ELG pilot projects execution and (ii) a guide to the European Language Grid itself.

European Language Grid – Analysis of the Survey

for Proposers to the Open Call 2

A short survey was designed for those who submitted a proposal (proposers). The survey comprised 15 questions and a mix of open questions and multiple-choice questions. The survey topics were clustered into three sections: "motivation", "project proposals", "your organization". All information was collected anonymously with the goal to evaluate and improve our call processes.

Of the proposers, 39 out of 103 (38 %) responded.

1 Survey Analysis – Proposers

1.1 Motivation

Question 1: What were the most important reasons for you to prepare and submit a proposal to the ELG open call? Please choose up to three.	Answers	Answers (%)
Contribute services or resources to the ELG platform to make them available to the emerging ELG community	23	59.0
Further development of an existing software or data project	23	59.0
Necessity of financial funding in order to realize a specific project idea	21	53.8
Get more visibility for my organization and/or products and/or services through the ELG	19	48.7
Branch out into new markets by means of the ELG platform and initiative	15	38.5
General interest in the further development and progress of the ELG platform and initiative	12	30.8
Make use of the ELG cloud platform to distribute my services (my organization does not run its own cloud platform)	10	25.6
General interest in EU-funded innovation actions	8	20.5
Gather experience with EU-funded open call FSTP projects (FSTP, Financial Support for Third Parties)	4	10.3

Question 2: Did you submit a proposal in the first ELG open call for pilot pro-	Answers	Answers (%)
jects?		

Yes	16	41.0
No	21	53.8
More than one	2	5.1

vices or because of the datasets?		
Both services and datasets	23	59.0
(Functional) services	9	23.1
Datasets	6	15.4
No opinion	1	2.6

Question 3: Are you interested in ELG primarily because of the (functional) ser- Answers Answers (%)

Question 4: What are your main expectations towards the ELG platform and initiative? Please choose up to three options.	Answers	Answers (%)
That ELG provides increased visibility for my organization on the European level	23	59.0
That I get access to a large repository of functional services and datasets	19	48.7
That ELG strengthens the LT and language-centric AI community in Europe as a whole	16	41.0
That I have an additional channel for the exploitation of my organization's re- search results	16	41.0
That ELG facilitates future collaborations with other developers by establishing a common platform	15	38.5
That ELG improves interoperability of LT by establishing a common API	13	33.3
That ELG serves as an information hub and matchmaker for buyers and suppliers of LT	13	33.3
That I have an additional sales channel for my organization's commercial ser- vices or datasets	13	33.3
That ELG becomes a part of a larger LT/AI platform eco-system in Europe	10	25.6
That we obtain funding for realizing important NLP ideas	1	2.6

Question 5: What should be the focus of the ELG platform and initiative in the Answers Answers (%) next 3-5 years? (This question generated 27 responses.)

Cover more languages; cover all EU languages; multilingualism; increase the number of European languages for automatic translations	4	14.8
Become a reference hub to share knowledge, dataset among EU players; facili- tating the deployment and availability of LT services all across Europe; making ELG a viable alternative to services like AWS Marketplace for LT; to have a com- mon EU environment for information and exploitation of LT	4	14.8
To launch as soon as possible as much as possible services, because English speaking countries are in huge advantage; provide datasets and running services; expanding the range of datasets and services	3	11.1
Improving transparency; openness; promote its use in citizen science projects	3	11.1
A great contribution would be unified formats for data sets and services that make it easy to combine them (e.g. same format for all NEW datasets); data interoperability	2	7.4
More attention to i) under resourced language technologies ii) Data2Text NLG	1	3.7
Making the initiative closer to users so SMBs can benefit from seamless access to a number of LT and integrate them into their WFs	1	3.7

₩ELG

Specific product portfolios and quality control / evaluation on the data/services offered to build trust	1	3.7
Since it's a new platform, keep listening to your users and be open to their com- ments and let that also guide you for the next 3-5 years. Additionally: ELG could help to establish EU companies as a privacy conscious alternative to the US based speech tech companies who are often further ahead in their R&D.	1	3.7
Fostering language research and collaboration between different research groups	1	3.7
 Fund innovation in NLP. Provide useful data for benchmarking and training NLP models. Improve visibility of small- and medium-sized NLP vendors to the market- place. 	1	3.7
Developing a generic LT/AI eco-system integrating both an academic and an in- dustrial approach and motivation	1	3.7
Serve as support for different companies that want to develop different solu- tions based on language technologies, supporting them with financing and dis- tributing the different technologies created by them.	1	3.7
Creation of functional APIs, Distribution and marketing to funded projects, Build network and community between participants.	1	3.7
Providing direct API calls on a robust self-hosted IT infrastructure, and not providing links to third-party APIs. More Azure/Google Cloud Platform, less Meta-Net	1	3.7
We believe it would be advantageous if the ELG would focus further on bringing together the efforts that have been made in LT and language-centric AI in Europe. While more and more resources are created in research and industry, many areas are still missing out on the positive side-effects of bringing these resources and thereby their creators and audience together. However, this could foster increased collaboration, faster progress of the domain and sustainable use of (research) funds, e.g. by making it possible that data or tools that are usually just used for a limited time for a single purpose by a single agent get reused and exploited more thoroughly by a bigger crowd. Moreover, an ELG that brings stakeholders and existing resources together might also raise awareness in the community in general for the need to standardize/harmonize and integrate tools and data, thereby making a) the resources more interoperable and reusable, and b) their creators more sensitive to issues like copyright, licensing, data and metadata representation and long-term preservation. This is especially valuable for communities where such matters are not yet in the focus of attention (e.g. because resources were not findable by others or perceived as too individual and not mutually benefitting - with the ELG platform similarities and common interests will hopefully become more evident). Finally, the platform and initiative could serve as a tool were stakeholders can experience at first hand the advantages of shared and integrated language resources, thereby slowly raising peoples' attitudes towards data and tool sharing in general.	1	3.7

1.2 Project Proposals

Question 6: In a typical year, how many agile project proposals (i.e., short proposal, quick evaluation, rather short project runtime, e.g., the Financial Support for Third Party setup like the ELG Open Calls) does your organization, department or team participate in?

Number of agile project proposals	Answers	Answers (%)
2 agile project proposals	10	25.6
3 agile project proposals	8	20.5
5 agile project proposals	7	17.9
1 agile project proposals	4	10.3
4 agile project proposals	3	7.7
10 agile project proposals	3	7.7
0 agile project proposals	1	2.6
6 agile project proposals	1	2.6
7 agile project proposals	1	2.6
20 agile project proposals	1	2.6

Question 7: In a typical year, how many consortia-based project proposals (i.e., typical EU Horizon 2020 project proposals) does your organization, department or team participate in?

Number of consortia-based project proposals	Answers	Answers (%)
1 consortia-based project proposal	15	38.5
3 consortia-based project proposal	9	23.1
2 consortia-based project proposal	8	20.5
5 consortia-based project proposal	3	7.7
0 consortia-based project proposal	2	5.1
6 consortia-based project proposal	1	2.6
20 consortia-based project proposal	1	2.6

Question 8: Both ELG open calls were very popular. Do you think more EUfunded activities dedicated to Language Technology and Language-centric AI are needed?

Answers	Answers	(%)

Yes	38	97.4
No opinion	1	2.6
No	0	0.0

Question 9: Does your organization have a preference for smaller, more agile calls like the ELG open call or for complex, consortia-based projects?	Answers	Answers (%)
Agile calls (short proposals, quick evaluation, 9-12 months project runtime)	28	71.8
No preference	8	20.5
Consortia-based projects (long proposals, complex evaluation, 24-36 months runtime)	3	7.7

Your Organization 1.3

Question 10: Did you submit the proposal on behalf of a SME or a research in- stitution?	Answers	Answers (%)
Research institution	16	41.0
SME	23	59.0
Question 11: What types of Language Technology do you specialize in, for ex- ample, Machine Translation, Text Analytics, Speech Recognition etc.? Please provide a comma-separated list of technologies. This question generated 37 responses with multiple values.	Answers	Answers (%)
Text Analytics	14	37.8
Machine Translation	12	33.4
Speech Recognition	7	18.9
AI-Solutions	4	10.8
Dialogue Systems	3	8.1
Machine Learning	3	8.1
Natural Language Processing	3	8.1
Corpus Linguistics	2	5.4
Information extraction	2	5.4
Part of speech tagging	2	5.4
Sentiment Analysis	2	5.4
Automated essay scoring	1	2.7
Automatic error correction	1	2.7
Automatic Translation Services	1	2.7
Automation	1	2.7
Conversational agents	1	2.7
Corpus creation and relevant LT	1	2.7
Custom LT development	1	2.7
Data Science	1	2.7
Dependency parsing	1	2.7
Distributional semantics	1	2.7
Emotion detection	1	2.7
Human-robot interaction	1	2.7
Keyword extraction	1	2.7
Lemmatizer	1	2.7
Linguistic data	1	2.7
LT integration	1	2.7
LT processing for non-standard texts	1	2.7
Metaphor detection	1	2.7
Multilingual language generation	1	2.7

₩ELG

Multilingual services computational interlinguas 1 2.7 Natural Language Generation 1 2.7 OCR 1 2.7 1 2.7 Offline speech recognition 1 2.7 **Opinion Mining** 1 2.7 Parsing 1 Semantic Matching 2.7 Sign Language Translation 1 2.7 1 Speech analysis 2.7 Speech Processing 1 2.7 Summarization 1 2.7 Text Classification 1 2.7 1 Text quality prediction 2.7 Text Recognition 1 2.7 Text to speech 1 2.7 Transcription tools 1 2.7

Question 12: Do you specialize in certain domains, for example, energy, health, mobility? Please provide a comma-separated list of domains.

Answers Answers (%)

This question generated 30 responses.		
No certain domains	8	26.7
Health	7	23.3
Finance & banking	3	10.0
Media & News	3	10.0
Tourism	3	10.0
Education	2	6.7
Energy	2	6.7
Insurance	2	6.7
Mobility	2	6.7
Accessibility	1	3.3
Computational Linguistics	1	3.3
Computer-assisted language learning	1	3.3
Computer-mediated communication	1	3.3
Culture & Arts	1	3.3
Design	1	3.3
Heritage preservation	1	3.3
Hospitality	1	3.3
Innovation, open innovation	1	3.3
Job offers	1	3.3
Learner corpora	1	3.3
Legal	1	3.3

European Language Grid Open Call 2 – Analysis of the Survey for Proposers

Marketing	1	3.3
Non-standard language	1	3.3
Patents, IP	1	3.3
Robotics	1	3.3
Smart City and IoT	1	3.3
Software localization	1	3.3
Telecommunication	1	3.3
Underwater	1	3.3
Voice ordering for distribution franchises	1	3.3

Question 13: Please specify up to five languages your organization is primarily interested in. Please state the name of the languages in full, separated with a comma, e.g.: Maltese, Hungarian, French.

This question generated 36 responses.

Language(s)	Mentions	Mentions (%)
Group A (Official EU Languages): Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovenian,		
Spanish, and Swedish	126	
English	29	80.6
French	22	61.1
German	20	55.6
Spanish	17	47.2
Italian	13	36.1
Dutch	4	11.1
Greek	4	11.1
Polish	3	8.3
Romanian	3	8.3
Hungarian	2	5.6
Portuguese	2	5.6
Swedish	2	5.6
Czech	1	2.8
Irish	1	2.8
Maltese	1	2.8
Slovak	1	2.8
Slovenian	1	2.8
Group B (Other EU languages; languages from EU candidate countries and Free Trade Partners): Armenian, Basque, Catalan, Serbian, Swiss German,		
Turkish	11	
Basque	2	5.6
Catalan	2	5.6
Turkish	2	5.6

European Language Grid Open Call 2 – Analysis of the Survey for Proposers

Armenian 2.8 1 Deutsche Gebärdensprache 1 2.8 Österreichische Gebärdensprache 1 2.8 Serbian 1 2.8 1 2.8 Swiss German Group C (Language spoken by EU immigrants; languages of important trade and political partners): Arabic, Chinese, Hindi, Sign languages, Latin, Persian, Quechua, Russian, Songhay, Swahili, Vietnamese, Yiddish 13 Arabic 1 2.8 Chinese 1 2.8 Hindi 1 2.8 Sign languages 1 2.8 Latin 1 2.8 Minority languages from America 1 2.8 Persian 1 2.8 Quechua 1 2.8 Russian 2.8 1 Songhay 1 2.8 Swahili 1 2.8 Vietnamese 1 2.8 Yiddish 1 2.8

Question 14: Please specify the country in which your organization is based. This question generated 38 responses.	Answers	Answers (%)
Germany	6	15.8
Italy	6	15.8
Spain	6	15.8
Switzerland	3	7.9
Czechia	2	5.3
France	2	5.3
Greece	2	5.3
the Netherlands	2	5.3
Austria	1	2.6
Belgium	1	2.6
Iran	1	2.6
Ireland	1	2.6
Luxembourg	1	2.6
Poland	1	2.6
Romania	1	2.6
Serbia	1	2.6
United Kingdom	1	2.6

₩^{II}ELG

₩IELG

Question 15: Do you have any other comments or suggestions you would like to share?	Answers	Answers (%)
This question generated 12 responses.		
No comment	5	41.7
Great initiative. Thank you for the good work!	3	25.0
We're a niche player (solutions in Dutch for the healthcare sector), I think ELG could support us in several ways to allow us to cover more languages spoken within the EU. Most of the language are also quite niche (but worth it!) compared to English.	1	8.3
It would be interesting to delevop an ELG community to share experiences.	1	8.3
Transparency on the ELG scoring must be improved in order for each submission to have full clarity on where it ranked. It was extremely frustrating for us to learn that reviewers rated our first submission worthy of funding but we did not know how far off our proposal was from actually making the top 10 which re- ceived funding. Knowing this would also have enabled us to make a better-in- formed decision regarding whether or not to submit another proposal for the second call.	1	8.3
We as SME are principal sponsors of the UNESCO Chair www.teclin.org.	1	8.3

ELG – FSTP – Pilot Projects Open Call 2 Guide for pilot projects (version: March 2021)

This document contains (i) guide and basic information about the ELG pilot projects execution (ii) guide to European Language Grid as such and (iii) instructions for registering metadata records during the preparation phase.

1. Project execution

All pilot projects will be executed in accordance with the approved project proposal and the Third Party Agreement.

Pilot projects will be offered guidance and supervision throughout the lifecycle of the project once the contract is in place and the project has started. Control days will be organized (possibly remotely) when projects will have to report progress, any problems, state of completion etc.

Evaluation of the results of the "(2) Experiment" Phase and access to the "(3) Integration" and "(4) Dissemination" phases

The Pilot Board is set up for the supervision of the pilot projects. It provides a forum so that the ELG project can discuss the progress of the pilots, their intermediate feedback and the results.

The Pilot Board consists of ELG consortium members. It is the main technical and strategic interface between the pilot projects and the ELG project so that the project can maximise its benefit from supporting the pilots and also to make sure that the pilot projects maximally benefit from the European Language Grid.

Each selected project will be supervised by one member of the Pilot Board ("Project Coach") appointed by the Pilot Board.

The Project Coach will be responsible for:

- monitoring the project planning and its progress,
- coordinating the training of the Awardee's project team with respect to the ELG potential,
- collecting and answering questions from the team during the execution of the project with the support of the ELG partners,
- collecting reports and guiding the project team through all activities (phases), especially through the Integration and Dissemination phases
- recommending to the Pilot Board if the project is allowed to progress to the next phase.

The assessment will be coordinated with the ELG project consortium; the Project Coach will seek especially technical help from the consortium to evaluate the pilot project results, especially in terms of technical testing to determine if the objectives of the (2) Experiment phase have been met. The Project Coach will assess the progress of the project and propose to the Pilot Board to approve the second payment to the Awardee, or to terminate the project after the (2) Experiment phase.

The pilot projects will be provided with report template, reports will be submitted via ELG Open Calls platform.

'ELG

Dissemination and presentation of pilot projects and results

To make the selected projects known to a wider public, we plan to present all pilot projects at the ELG conference, the (virtual) META-FORUM 2021. There will be a dedicated time slot where all project leaders can present their project idea and explain the goals they are aiming for. At the META-FORUM 2022 (which will hopefully be held again as a face-to-face event), all pilot projects should show their results and demonstrate the usefulness of ELG. Part of the allocated budget for dissemination can be assigned to the dissemination task within these events.

Moreover, the ELG consortium will produce a book at the end of the ELG project. All pilot projects will be invited to write a short paper on their projects, which will then be published in the resulting ELG book. This contribution can be partly based on the report that the pilot projects write at the end of their project runtime.

For dissemination purposes, we have also created a page on the <u>ELG website</u> with a list of all selected projects. In addition, we would like to include in the <u>ELG catalogue</u> metadata records (landing pages) dedicated to:

- the project and the targeted results
- the organization that runs the project and
- the services and resources that are in the process of being prepared (linked to the project & organization); these will be marked as "under construction" until the final delivery of the project outcomes.

In addition, there are National Competence Centres (NCCs) in all European countries who support ELG and act as a bridge between the ELG project and the local players in the field of Language Technology. These NCCs will over the upcoming months help conduct dissemination and training events. In many cases, these events can also be a good opportunity to present pilot projects in a specific region. The actual implementation and design of the dissemination activities will be determined and planned in each individual case with the project coach and the event organisers.

Evaluation of the final results

The final evaluation of a project will be performed after the (3) Integration activity by the Project Coach and if the project fulfils the (4) Dissemination obligations. The Project Coach will then prepare a short report (to be made public) and recommend to the Pilot Board to approve (or not) the final (third) payment to the project Awardee.

The pilot projects will be provided with report template, reports will be submitted via ELG Open Calls platform.

After a project has finished, the project team is required to present their results, business plans, secured venture capital for further development and future plans. The Pilot Board will assess the finished projects and evaluate the immediate results. It will also formulate recommendations for sustainability and future operation of the ELG based on the experience of and with the pilot projects.

Communication between pilot projects and ELG

Regarding **Third Party Agreement** and **payments**, please contact **ELG Open Calls Management Team** (Charles University, e-mail <u>pilot-projects@european-language-grid.eu</u>).

Regarding the **execution of your project**, co-operation with ELG, questions regarding the Grid etc. – please always **contact your project coach**. If they are not able to answer themselves, they will get the answer from other members of the consortium.

All relevant questions and answers will be put into the FAQ document that will be available to all pilot projects.

Project Calls and Execution – overall workflow



''ELG

2. Guide to European Language Grid

The <u>European Language Grid (ELG) platform</u> offers access to a multitude of assets related to Language Technology (LT), including *commercial* and *non-commercial Language Technologies* for all European languages, *data resources* (such as models, datasets, lexica, terminologies, grammars), as well as information on *LT-related projects, organizations*, and *groups*.

The **ELG User Manual** (<u>https://european-language-grid.readthedocs.io/</u>) includes all the information needed to prepare and register **Language Resources and Technologies (LRTs**). For ease-of-use, we include here links to specific chapters and relevant material (e.g. publications and presentations).

Introductory material

Overview of the ELG project and platform

Overview of the European Language Grid (ELG) project and short description of the platform operations: Rehm, G., Berger, M., Elsholz, E., Hegele, S., Kintzel, F., Marheinecke, K., Piperidis, S., Deligiannis, M., Galanis, D., Gkirtzou, K., Labropoulou, P., Bontcheva, K., Jones, D., Roberts, I., Hajic, J., Hamrlová, J., Kačena, L., Choukri, K., Arranz, V., Vasiļjevs, A., Anvari, O., Lagzdiņš, A., Meļņika, J., Backfried, G., Dikici, E., Janosik, M., Prinz, K., Prinz, C., Stampler, S., Thomas-Aniola, D., Pérez, J. M. G., Silva, A. G., Berrío, C., Germann, U., Renals, S., and Klejch, O. (2020). European Language Grid: An Overview. In Nicoletta Calzolari, et al., editors, Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020), pages 3366-3380, Marseille, France, May. European Language Resources Association (ELRA).

(http://www.lrec-conf.org/proceedings/lrec2020/pdf/2020.lrec-1.413.pdf)

- Overview of ELG, history and context: Rehm, G (2020) Overview of ELG. 1st Regional ELG workshop (slides - <u>https://www.european-language-grid.eu/wp-</u> <u>content/uploads/2020/06/1stRegionalELGWorkshop 1_Overview.pdf</u>)
- Short description of the platform operations, with an emphasis on the metadata schema used in the ELG catalogue: Labropoulou, P., Gkirtzou, K., Gavriilidou, M., Deligiannis, M., Galanis, D., Piperidis, S., Rehm, G., Berger, M., Mapelli, V., Rigault, M., Arranz, V., Choukri, K., Backfried, G., Perez, J. M. G., and Garcia-Silva, A. (2020). Making Metadata Fit for Next Generation Language Technology Platforms: The Metadata Schema of the European Language Grid. In Nicoletta Calzolari, et al., editors, *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, pages 3428-3437, Marseille, France, May. European Language Resources Association (ELRA).

(http://www.lrec-conf.org/proceedings/lrec2020/pdf/2020.lrec-1.420.pdf)

Using the ELG platform (e.g. to test a service)

You can play around through the catalogue with the LRTs that have already been integrated in ELG to see how they work.

- European Language Grid (ELG) January 2021 Screencast Video
 - o <u>https://www.youtube.com/watch?v=LD6QadkkZiM</u>
- How to test the ELG integrated services:
 - o <u>https://european-language-grid.readthedocs.io/en/stable/all/2_Using/TestService.html</u>
- How to download a data resource:
 - o <u>https://european-language-grid.readthedocs.io/en/stable/all/2_Using/Down.html</u>

₩**ELG**

Providing LRTs to ELG

To share a Language Resource or Language Technology through the ELG platform, you must provide

- the resource itself (in the form of integrated service or downloadable resource), which must comply to the ELG technical specifications, and
- a metadata record that describes it.

A short introduction is provided at: Galanis, D. (2020) How to integrate services or data sets into the ELG platform. 1st Regional ELG workshop (slides - <u>https://www.european-language-grid.eu/wp-</u> <u>content/uploads/2020/09/1stRegionalELGWorkshop 4 ProvidingServicesTutorial.pdf</u>). Please note that this introduction doesn't include the option of the ELG interactive editor for the description of records that is presented below.

Step 1: Preparing the resource/service

ELG compatible LT services

- Introduction and main technical requirements: <u>https://european-language-grid.readthedocs.io/en/latest/all/3_Contributing/Service.html</u>
- Technical documentation (specifications for the ELG integrated services APIs): <u>https://european-language-grid.readthedocs.io/en/latest/all/A2_API/LTPublicAPI.html</u>
- Dockerization tips with help for Python and Java-based tools: <u>https://european-language-</u> grid.readthedocs.io/en/latest/all/3 Contributing/Service.html#step-0-dockerize-your-service

Data resources (e.g. corpora, lexica, etc.)

Under data resources we include corpora, language descriptions and lexical/conceptual resources.

Corpora are structured collections of data selected according to specific criteria in order to represent as comprehensively as possible a research question. The most common cases are:

- text corpora: monolingual, bilingual or multilingual collections of texts in a specific domain, such as corpora of news articles, scientific publications, legal documents, medical records, tweets, etc.
- corpora of audio recordings, e.g., of broadcast news, or lists of sentences recorded by individuals from a specific region with a dialect accent, etc.
- collections of videos, such as interviews with politicians, sign language corpora, etc.
- corpora combining all of the above, such as a multimedia corpus of video lectures, with their audio recordings, transcripts, subtitles and their translations.

Under language descriptions, we comprise:

- models, including Machine Learning models, statistical models, word embeddings, n-gram models,
- computational grammars of a language, language variety or for a specific domain or phenomenon.

The vast majority of these consist of a text part, but videos and images are also foreseen for cases such as sign language grammars.

Examples of lexical/conceptual resources include

- computational lexica, that are used for computational processing, and include morphological, syntactic and semantic information;
- dictionaries in digital format,

₩**ELG**

- ontologies and controlled vocabularies,
- monolingual and multilingual terminological glossaries,
- word lists, gazetteers of place names, proper names, etc.

They typically consist of a text part, but they may also comprise audio and video files, as in the case of multimedia lexica with sound recordings (e.g., pronunciation of a word) and images (e.g. pictures denoting the sense of a word), sign language lexica with videos, etc.

The only technical requirement for data resources currently are their packaging into a .zip archive in order to be uploaded at the ELG storage sytem. We are in the process of making more detailed specifications that we will soon share with you.

Step 2: Describing and providing access to the service or resource

To register the services or resources, you can make use of the ELG interactive editor or create ELG-compliant metadata files in XML format and upload them to the platform. You will find more information on their description requirements:

- For services: <u>https://european-language-</u> grid.readthedocs.io/en/stable/all/3 Contributing/Service.html#describe-the-service
- For corpora (datasets): <u>https://european-language-grid.readthedocs.io/en/stable/all/3_Contributing/Corpus.html#describe-the-corpus</u>
- For lexical/conceptual resources (e.g. lexica, ontologies, terminological glossaries, etc.): <u>https://european-language-grid.readthedocs.io/en/stable/all/3</u> Contributing/LexConc.html#describe-<u>the-lexical-conceptual-resource</u>
- For models:
 <u>https://european-language-grid.readthedocs.io/en/stable/all/3_Contributing/Model.html#describe-the-model</u>
- For grammars:

https://european-languagegrid.readthedocs.io/en/stable/all/3 Contributing/Grammar.html#describe-the-grammar

3. Instructions for the registration of metadata records of pilot projects during the preparation phase

The following instructions are meant to help you provide the required metadata records for:

- the project
- your organization, and
- the resources you will provide to the ELG catalogue.

The following sections include instructions adapted to the requirements of the pilot projects. For your convenience, we have added here what you need for this procedure with links to the relevant sections in the ELG online guide, where needed.

General instructions

Step 1 – Register as a provider

• You can register at https://live.european-language-grid.eu/ using the icon at the top right corner. Once your account is activated, visit your profile page and request to be granted "Provider" permissions.

Step 2 – Create the metadata records

All metadata records must comply to ELG's metadata schema, ELG-SHARE.

You can use the ELG interactive editor or one of the ready-to-use templates or examples in XML format provided at our GitLab repository. In both cases, some requirements are specific to Open Call projects and are described below.

- Sign in at https://live.european-language-grid.eu/ and add the metadata records with your preferred option. To ensure linking between the various items, please register them in the following order:
 - Organization
 - Project
 - Services and resources.
 - Please note that only one person is responsible for creating and editing each record.
- **Projects**: Supply all the information regarding your project as described in your proposal as well as the approved budget and timeline. Add the following metadata elements to link to the ELG call:
 - o fundingSchemeCategory: ELG Open Call
 - relatedCall: ELG Open Call II
- ELG-compliant services: Normally, you would contribute the service (in the format of a docker image according to the ELG technical specifications) together with its description (metadata record); however, the dockerization and integration will be done through the course of your project. For the time being, you can only describe the service in order to be published at the catalogue as "Work in progress". When you describe the service, please tick on the "Under construction" box in the editor or upload form.

When you have finished the dockerization and you are ready to proceed to the next step, you will notify us so that we give you the rights to and update it with the required information.

• Data resources uploaded at ELG: Normally, you would contribute the resource together with its description (metadata record). For the time being, you can only create and upload the metadata record in order to be published at the catalogue as "Work in progress".



When you have finished the formatting and packaging of the resource and you are ready to proceed to the next step, you will notify us so that will notify us so that we give you the rights to edit the metadata record and update it with the required information.

Step 3 – Submit the metadata records for publication

You can edit the records until you are satisfied with the description. You can then proceed to the submit them for publication at your dashboard. Once submitted, you can no longer edit them, but you will not see them yet published at the catalogue. They are only visible to ELG administrators. We will check your contribution and integrate it into the ELG catalogue if everything is in order, and contact you otherwise.

Instructions for metadata records per item type

Organizations

- Examples
 - Companies
 - Expert system: <u>ELG catalogue page</u> & <u>XML file</u>
 - SAIL: <u>ELG catalogue page</u> & <u>XML file</u>
 - Universities & public research organizations
 - Charles University: <u>ELG catalogue page</u> & <u>XML file</u>
- Templates for pilot projects
 - Company: <u>XML file</u>
 - University & research organization: XML file
- Documentation of the schema: <u>Metadata elements for organisations</u>

Projects

- Examples
 - European Language Grid: <u>ELG catalogue page</u> & <u>XML file</u>
 - European Clinical Case Corpus: <u>ELG catalogue page</u> & XML file
 - 0
 - Template for pilot projects: XML file
- Documentation of the schema: <u>Metadata elements for projects</u>

ELG-integrated LT services

- Examples
 - Information Extraction services
 - ANNIE's Named Entity Recognizer: ELG catalogue page & XML file
 - Cogito Discover Semantic Annotator: ELG catalogue page & XML file
 - Machine Translation services
 - UEDIN Machine Translation Service for German to English: <u>ELG catalogue page</u> & <u>XML file</u>
 - Transformer en-cs: Machine Translation Model Trained Using Tensor2tensor: <u>ELG</u> <u>catalogue page</u> & <u>XML file</u>
 - o Text classification services
 - Cogito Discover Text Categorizer: <u>ELG catalogue page</u> & <u>XML file</u>
 - Automatic Speech Recognition services
 - SAIL LABS Speech-to-text for Spanish: <u>ELG catalogue page</u> & <u>XML file</u>
 - o Text-to-Speech Generation services
 - Tilde Text-To-Speech (TTS), Latvian Language, Woman's Voice: <u>ELG catalogue page</u> & <u>XML file</u>
- Templates for under construction services from pilot projects
 - Information Extraction services: <u>XML file</u>

₩ELG

- Machine Translation services: XML file
- Text classification services: XML file
- Automatic Speech Recognition services: XML file
- Text-to-Speech Generation services: XML file
- Documentation of the schema: <u>Minimal metadata for tools/services</u>

Corpora

- Examples
 - o Raw corpora
 - BMI Brochures 2011-2015 (Processed): <u>ELG catalogue page</u> & <u>XML file</u>
 - BuzzFeed-Webis Fake News Corpus 2016: <u>ELG catalogue page</u> & <u>XML file</u>
 - Annotated corpora:
 - INTERA Corpus the English POS annotated part of the English-Slovene SVEZ ACQUIS Corpus: <u>ELG catalogue page</u> & <u>XML file</u>
- Templates for under construction resources from pilot projects
 - o <u>Bilingual raw corpus</u>
 - o <u>Annotated corpus</u>
- Documentation of the schema: Minimal metadata for corpora

Lexical/Conceptual resources

- Examples
 - o Bootstrapped Lexicon of English Verbal Polarity Shifters: ELG catalogue page & XML file
 - English Ontology lexicon: ELG catalogue page & XML file
- Templates for under construction resources from pilot projects
 - o <u>Monolingual computational lexicon</u>
 - o <u>Multilingual terminological lexicon</u>
- Documentation of the schema: Minimal metadata for lexical/conceptual resources

Language descriptions (models & grammars)

- Examples
 - o PANACEA Environment Corpus n-grams EL (Greek): ELG catalogue page & XML file
 - Tourism Italian grammar: <u>ELG catalogue page</u> & <u>XML file</u>
 - English Model (CoNLL-2003) for NameTag: XML file (not published on catalogue)
 - Templates for under construction resources from pilot projects
 - o <u>ML model</u>
 - o <u>N-gram model</u>
 - o <u>Computational grammar</u>
- Documentation of the schema: Minimal metadata for models and grammars