EUROPEAN LANGUAGE GRID

D6.2

Call 1 results and description of selected pilot projects

Authors:	Jana Hamrlová (CUNI), Lukáš Kačena (CUNI), Jan Hajič (CUNI)
Dissemination Level:	Public
Date:	30-08-2020

ᴵᴸᴵ ELG

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, 36 months
Deliverable number	D6.2
Deliverable title	Call 1 results and description of selected pilot projects
Type	Report
Number of pages	54
Status and version	Final – Version 1.0
Dissemination level	Public
Date of delivery	Contractual: 31-08-2020 – Actual: 30-08-2020
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	Julija Melnika (TILDE), Andrejs Vasiljevs (TILDE), Kilian Evang (DFKI)
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
	Sail Labs Technology GmbH (SAIL), Austria
	Expert System Iberia SL (EXPSYS), Spain
	University of Edinburgh (UEDIN), United Kingdom
EC project officers	Philippe Gelin, Alexandru Ceausu
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 © 2020 ELG Consortium

Table of Contents

List of Figures

List of Tables

Abstract

This document describes the selection process and results of the first open call for ELG pilot projects, which will broaden ELG's portfolio of language technologies 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 organising the submission and the evaluation process of the first open call is explained.

The first open call for ELG pilot projects was very successful, with 121 submitted project proposals, of which 110 have passed the formal requirements for evaluation. 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 10 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 planned webinar with the Pilot Board members. Moreover, there will be continuous guidance by the assigned project coach for each project.

The rich experience from the first open call is being reflected in the current preparation for the second open call. This second call will be announced in October 2020.

The structure of this document is as follows: After the Introduction, a brief summary of the first open call is provided in Chapter 2 to review the basic information about the call and evaluation process, together with the selection process and feedback provided so far by evaluators, Pilot Board members, and project proposers. Chapter 3 provides a description of additional support for the 10 selected pilot projects during their execution phase. The projects are presented in Chapter 4. Finally, in Chapter 5, next steps regarding the second open call are laid out, including improvements that are being considered based on feedback from the first open call.

1 Introduction

To demonstrate the usefulness 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 2018–2020. 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 organisations (only SMEs and research organisations were allowed to apply for the first open call) to either

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

(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 the public. The ELG project and, later on, the long-term initiative, will also provide further access to promotion and dissemination activities.

2 Summary of Open Call 1

This chapter provides the overall overview of the first 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 first open call was opened on 01 March 2020 and closed on 30 April 2020 (23:59 CEST) in accordance with the open calls' timeline (Figure 1). In total, we accepted 110 project proposals for evaluation from 103 applicants:

- 62 proposals were submitted by SMEs (36 of type A, 26 of type B); 57 unique SMEs;
- 48 proposals were submitted by research organisations (43 projects of type A, five of type B); 46 unique research organisations.

7 applicants (5 SMEs and 2 research organisations) submitted two proposals (one type A and one type B). Regarding the type of project, 79 submitted proposals were of type A (contribute resources, services, tools, or data sets to the ELG to increase its coverage), and 31 project proposals were of type B (develop applications using language resources and technologies available in the ELG).

We received applications from 29 various countries, including eligible countries outside the EU (Iceland, Israel, Norway, Serbia, South Africa, Switzerland, Turkey, United Kingdom).

The total amount of financing requested by the submitted projects was 16,900,000€. One project requested 283,000€, which is over the limit of 200,000€ per project, and the lowest requested amount was 50,000€. The average amount requested per project was 153,000€ (for details, see D6.1 subsection 3.2.3).
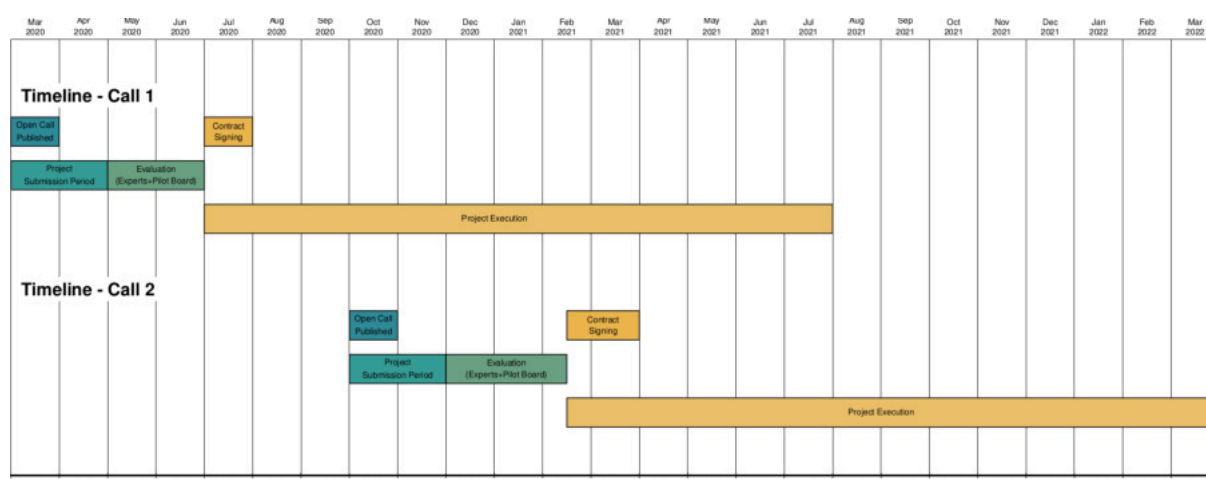

Figure 1: Open calls timeline

2.2   Evaluation process

The evaluation process was executed as planned (see Figure 1) during May and June. 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 were evaluated by external experts, the Pilot Board member responsible for a proposal (project coach) prepared a summary of the three evaluator reports, included their own feedback, and submitted the summary evaluation report to the Pilot Board.

Then, in the final selection meeting, the Pilot Board selected the project proposals which will be funded.

2.3   Current status of Open Call 1

At the time of submitting deliverable D6.1 (May 31, 2020), the first open call was already closed, and each project proposal had been assigned to three independent evaluators. By the end of May, we had received evaluation reports from most of the evaluators, and the date of the final Pilot Board selection meeting was set to June 19, 2020.

In this section, we describe tasks done after the submission of D6.1 and summarise the current state of the first open call.

2.3.1   From D6.1 to D6.2

At the beginning of June, all evaluation reports (including those still missing at the end of May) had been submitted by evaluators through the ELG Open Calls Platform. Then, 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 to all Pilot Board members on the platform's dashboard. Each of the 5 project coaches summarised more than 20 project proposals in this way.

The form used for the summary evaluation reports is attached as Annex 10 to D6.1. In this summary report, the project coach firstly reviewed the three evaluation reports submitted by the external evaluators (items summary of evaluations, project proposal assessment, summary, and recommendation for financing). Then s/he eventually suggested budget adjustments and, if applicable, a change of the total number of points assigned to the proposal in range of at most 30 points (up or down), which is 10% of the maximum 300 points to be received from the three evaluators. Further, in accordance with the Evaluation Criteria (Annex 7 of D6.1), project proposals from SMEs developing applications using language technology already available in ELG (B type projects) received 30 extra bonus points. Finally, the project coach reviewed the Eligibility criteria (uniqueness, relevance for ELG, and project phases) as checked by the evaluators and suggested her/his decision on their fulfilment if the evaluators differed in their opinions.

The project coach 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. This feedback was very helpful and will be discussed in detail in sections 2.5 and 2.6.

After all summary evaluation reports were submitted by the project coaches to the ELG Open Calls Platform, the Pilot Board selection meeting was convened. It took place online on June 19, 2020.

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 of the Pilot Board followed, concluding with a decision of the Pilot Board on:

- Fulfilment of eligibility criteria
- Points awarded by the Pilot Board (+/- 30 points, as suggested by the project coach, possibly modified by the Board)
- Sufficient quality level of a project to approve its suitability for financing
- Potential budget modifications for a given project (reduction, structural change)[1]

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, and 40 project proposals were recommended by the Pilot Board for financial support. The ranked list was cut at the maximum available support (approx. 1,365,000€). Thus, 10 project proposals were selected for funding amounting to a funding sum of 1,363,915€ in total. The highest-ranking non-selected proposal had received 273 points out of the absolute maximum of 360 points (300 by external evaluators + 30 bonus to SMEs submitting B type project + 30 by project coach/Pilot Board).

The two projects immediately below the cut were put on a reserve list, but since all selected projects signed the agreement for financial support, these two reserve projects were eventually informed and not selected for financial support in the first open call.

2.3.2 Status as of 31 July 2020
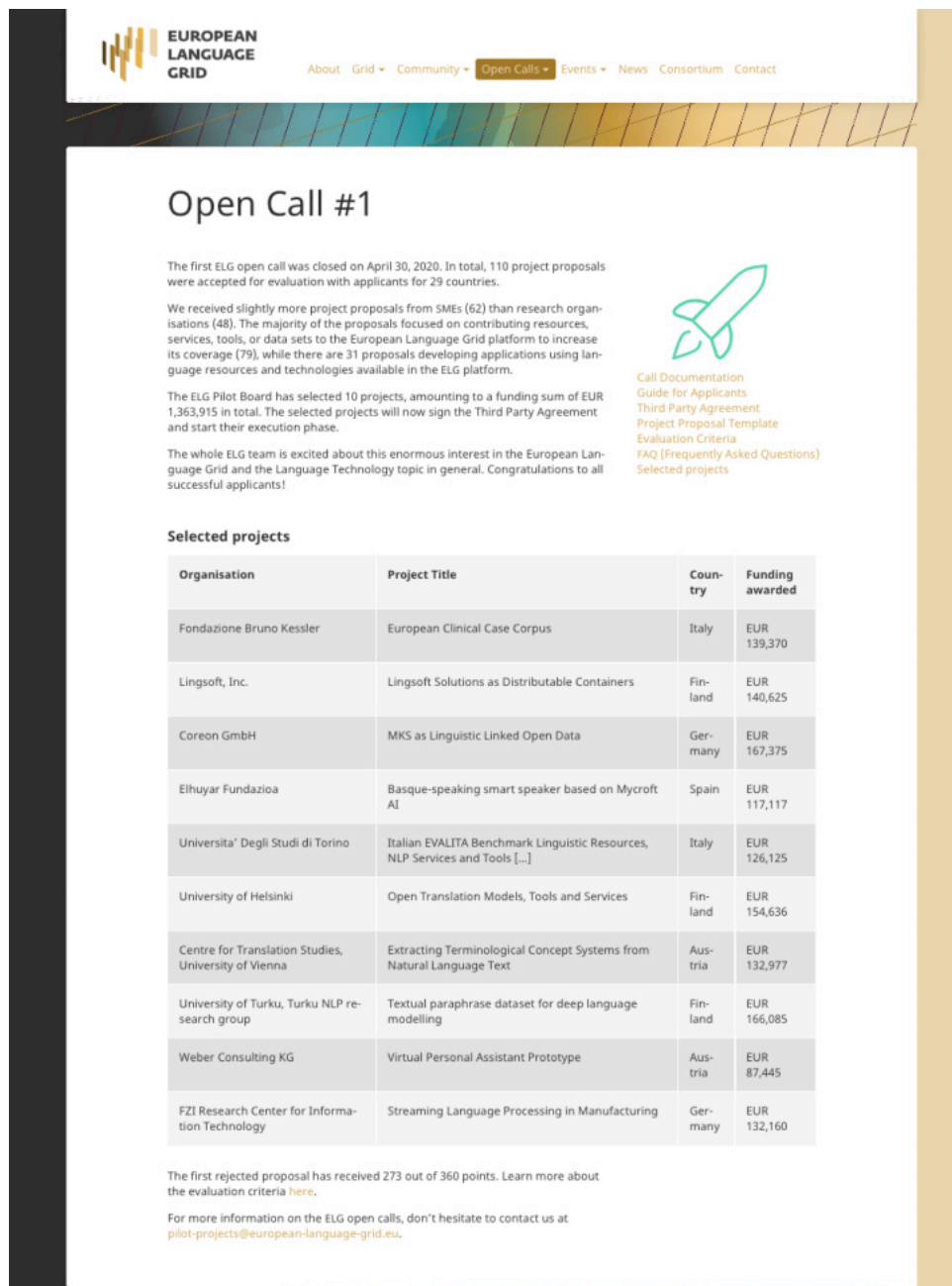
At the end of June, after the Pilot Board selection meeting, the results of the first open call were announced on the ELG website[2] (see Figure 2), including the list of projects selected for funding. The two projects from the reserve list were informed that they might be selected for financial support if any of the selected projects reject the financial support. The remaining projects were informed that they were not selected for financing.

During July, 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 August 6.

At the end of July, 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 individual applicants.

---

[1] This was discussed only for a few projects and reserve projects only.
[2] https://www.european-language-grid.eu/open-calls/open-call-1

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

2.4 Feedback provided so far

In order to receive feedback from everyone involved in the first open call, we conducted several surveys on the open call procedure, including the evaluation and selection process. We started with the feedback from the project proposers. After the evaluation process finished, we conducted a survey among all evaluators involved. The last survey was conducted among the Pilot Board members. Members of the Management Team organising the open call also provided feedback.

2.4.1 By project proposers

Two short surveys were designed for those who submitted a proposal (proposers) and those who uploaded an initial draft but did not submit a final version (non-proposers). The survey consisted of 15 questions, some open

and some multiple choice. The survey topics were clustered into three sections: "motivation", "project proposals", and "your organisation". All information was collected anonymously with the goal to evaluate and improve our call processes. The surveys were conducted in May 2020. Of the proposers, 73 out of 110 (66 %) responded, and of the non-proposers, 6 out of 17 (35 %) responded.

Results of the survey related to the demographics, languages, and domains, as well as motivation and expectations, and they are part of the D6.1 "Pilot Calls Setup" (section 6.2). Here we point out our main conclusions from the proposers' survey that are relevant for the setup of the second open call:

- Almost 70 % of respondents are interested in ELG because of both (functional) services and the datasets.
- Slightly more than two thirds of respondents prefer smaller, agile calls over large, consortium-based calls.
- There is a demand for more detailed documentation (e.g., in the form of a webinar) that allows proposers to better interpret the strategic goals of ELG and get better information on already existing services in ELG.
- More details about the API integration for ELG and about the infrastructure for working with data, applications, and possibly also workflows were requested.
- Some improvements of the ELG Open Calls Platform and its user-friendliness could be made (e.g., limited space).

The main conclusions from the survey among the non-proposers relevant for the setup of the second open call include[3]:

- The vast majority of non-proposers abandoned their draft proposals because they had too little time to finish it.
- Suggestions made by the non-proposers:
  - Inform the proposers if there is another proposal being prepared by the same organisation.
  - Advertise more actively; organisations found out about the call too late.
  - Explain how to involve targeted customers into the proposal.

2.4.2   By evaluators

In June 2020, a survey on the evaluation process was conducted among all involved evaluators. Of all involved evaluators, 40 out of 48 (83 %) responded.

The main conclusions from the survey among the evaluators are[4]:

- Almost two thirds of the evaluators assess the description of the evaluation process in the Call documentation as easy to understand; others see it as comprehensible.
- Half of the evaluators found the structure of the proposal template with regard to the evaluation process to be very well structured, 43% of the evaluators assess the proposals template to be basically well structured with some minor improvements; evaluators assessed the structure of the evaluation report template similarly.

---

[3] The full results of both surveys are attached in Annex 1.
[4] The full results of this survey are attached in Annex 2.

- Both, eligibility and evaluation criteria were assessed as well chosen and easy to understand by a majority of evaluators.
- The ELG Open Calls Platform was assessed as user-friendly and easy to work with by two thirds of respondents, whilst the rest assessed it as acceptable.
- The vast majority of evaluators (93%) assessed the support from the ELG team to be very good.
- Of all evaluators, 45% spent 1–2 hours evaluating per project proposal, and 40% of evaluators spent 2–3 hours.
- Almost 60% of respondents did not consider the remuneration obtained for evaluation to correspond with the expended effort.
- The vast majority (85%) of the evaluators had had some experience with the evaluation of project proposals before the ELG open call and are also planning to take part in the second ELG open call.

2.4.3   By Pilot Board members

A feedback survey among the Pilot Board members is currently being conducted. The main conclusions from the survey results obtained so far and from numerous discussions held with Pilot Board members relevant for the setup of the second open call are:

- Improve the proposal template to get a better indication on the budget composition and requested person months associated with the funding.
- Eligibility and evaluation criteria were sometimes misunderstood and need a better explanation.
- The second open call might want to focus on the gaps identified by ELG after the first open call.
- In the submission phase, it turned out that there was some ambiguity on the part of the proposers regarding the term "platform" (i.e., the ELG Open Calls Platform was mistaken for the ELG grid platform and vice versa).
- Pilot Board members assessed all evaluators in the summary evaluation report done for every single project proposal by rating the performance of the three independent evaluators and indicating evaluators that they would not recommend to be appointed for an evaluation in the second open call.

2.4.4   By Management Team

Overall, the Management Team was satisfied with the first open call and acknowledges the feedback provided by the proposers, evaluators, and Pilot Board members. The Management Team's main conclusions are:

- The call documentation (Annex 3 of D6.1) is well structured and needs only minor improvements.
- Although some proposers and evaluators called for more detailed proposals, the length and depth of the project proposal template meet the requirement of enabling an easy and lightweight procedure.
- In the project proposal template, a better structure of the budget section is needed.
- In some parts of the project proposals template, namely project description, as well as the evaluation reports template, better and more profound explanations, and instructions are needed.
- Due to the high number of submitted project proposals, a larger number of experienced and high-quality evaluators is needed.

2.5   Lessons learned

The following lessons learned were gathered from the feedback provided by proposers, evaluators, Pilot Board members, and the Management Team:

- From the online survey to proposers of the pilot projects we understand that we should explain and describe more profoundly the strategic goals of ELG, goals of the open call and also the ELG infrastructure as such (e.g., what is meant by "API integration into the platform").
- We have established an "online helpdesk"; throughout the project submission period, there were no major issues, all relevant questions were answered within 1–2 working days and also incorporated and published in the FAQ section of the open calls website. However, some of the proposers indicated in the online survey that they would welcome a webinar up front, guiding the prospective proposers through the call documentation and through the process.
- In the ELG project, costs for the open calls platform and for the evaluation of project proposals should have been more carefully planned.
- Explanations and instructions in the call documentation, related annexes, templates, and forms (project proposals template, evaluation and eligibility criteria, evaluation report, etc.) should be fine-tuned (more detailed, profound, clear and easy to understand), along with the open calls platform.
- Budget breakdown needs to be requested in a more detailed structure, with clear requirements for budget justification.
- The group of evaluators should be big enough to secure enough high-quality reviews, even with a large number of submitted project proposals.

3   Project support

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

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

After finishing Phase 1, reporting from 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 Pilot Board member ("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 (provided the project is allowed to progress).

To advertise the selected projects to a wider public, we plan to present the pilot projects at the ELG conference, the (virtual) META-FORUM 2020. There will be a dedicated time slot where all project leaders can present their project idea and explain the project goals. In 2021, when META-FORUM can hopefully be held again as a face-to-face event, pilot projects should show their results and demonstrate the usefulness of ELG. The projects may assign 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 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 organisers.

3.1   Webinar for project holders

The first event where projects can become more familiar with ELG will be an online meeting of the Pilot Board with the pilot projects. It will take place online on 10 September 2020. During the meeting, basic information about the ELG and its technology, as well as guidelines for the project execution, will be presented. Prior to the meeting, the Guide for pilot projects (see Annex 3) will be 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 10 projects selected for financial support in the first open call. All pilot projects are summarised in Table 1.

Organisation	Legal Form	Project Name	Project Type	Country	Funding Awarded
Fondazione Bruno Kessler	RO	European Clinical Case Corpus	A	Italy	139,370€
Lingsoft, Inc.	SME	Lingsoft Solutions as Distributable Containers	A	Finland	140,625€
Coreon GmbH	SME	MKS as Linguistic Linked Open Data	A	Germany	167,375€
Elhuyar Fundazioa	RO	Basque-speaking smart speaker based on Mycroft AI	B	Spain	117,117€
Universita' Degli Studi di Torino	RO	Italian EVALITA Benchmark Linguistic Resources, NLP Services and Tools for the ELG Platform	A	Italy	126,125€
University of Helsinki	RO	Open Translation Models, Tools and Services	A	Finland	154,636€

Centre for Translation Studies, University of Vienna	RO	Extracting Terminological Concept Systems from Natural Language Text	A	Austria	132,977€
University of Turku, Turku NLP research group	RO	Textual paraphrase dataset for deep language modelling	A	Finland	166,085€
Weber Consulting KG	SME	Virtual Personal Assistant Prototype	B	Austria	87,445€
FZI Research Centre for Information Technology	RO	Streaming Language Processing in Manufacturing	A	Germany	132,160€

Table 1: List of pilot projects selected for financial support in the first open call

4.1 Introduction and summary

Although we obtained more project proposals from SMEs than from research organisations, there are 3 SMEs and 7 research organisations among the projects selected. Similarly, although the B type projects from SMEs were preferred, 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.

Four of the eight A type projects aim to enrich the ELG platform with language resources, and six of them plan to provide the ELG platform with various language tools (i.e., two of the projects provide both resources and tools). The two B type projects promise speech applications – a smart speaker and a digital twin based on real-time language translation and analysis. The projects in general often deal with underrepresented languages like Basque, Nordic languages, European minority languages, etc.

Technologically, the projects target a very diverse set of goals and areas. There are projects targeting important interdisciplinary areas (medical informatics, manufacturing), modern technologies relating to language and semantic/world knowledge (Linked Open Data, paraphrasing) and core scalable technologies (distributable containers). Evaluation platforms and advanced, and scalable machine translation still are and will be relevant issues for Language Technologies. Finally, the two speech-oriented applied projects broaden the portfolio of the usual Language Technologies in the desired direction, too.

The awarded budget varies from 87,445€ to 167,375€. All supported organisations are from the EU – three from Finland, two from Austria, Germany and Italy, and one from Spain.

4.2 Selected projects

In this section, the individual supported projects will be introduced. After the Project Name and Project Acronym, the applying Organisation 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 – E3C: European Clinical Case Corpus

Project Name: European Clinical Case Corpus

Project Acronym: E3C

Organisation: Fondazione Bruno Kessler

Fondazione Bruno Kessler is a not-for-profit private research centre. It participates in the E3C proposal through its Natural Language Processing research unit affiliated to the ICT Centre. The NLP group (https://ict.fbk.eu/units/nlp/) has decades of experience in the creation and distribution of linguistic resources.

Currently, more than 30 annotated corpora and lexical resources are being distributed through the website of the group.

Project Abstract: E3C aims to collect and annotate a multilingual corpus of clinical narratives, with the ambition to become a European reference resource. It takes advantage of available corpora of clinical narratives in 5 languages (Italian, English, Spanish, French and Basque), and will collect new data when necessary. The goal is to harmonise current annotations, extend them, and provide baselines for several information extraction tasks.

E3C will contain three nested annotation layers: (i) a manually annotated layer of clinical entities and events, as well as temporal information and event factuality; (ii) a larger semi-automatic annotation layer of clinical entities to be used for training purposes; and (iii) a non-annotated layer of documents to be used for semi-supervised learning.

Total Budget: 139,370€

Type of project: A

Project Schedule: The duration of the project will be 12 months, with two milestones at month 9 (E3C multilingual corpus), and month 11 (ELG integration).

Phase 1: Experiment (Month 1-10)

- Guideline definition: Month 1-3, D1: Annotation guidelines for the 5 languages of the project
- Data collection: Month 1-3, D2: The E3C multilingual corpus
- Data annotation: Month 3-9, D3: The E3C multilingual annotated corpus
- Milestone I: Report and E3C corpus (Month 9)
- Quality assessment: Month 3-9, D4: Quality assessment
- Baselines: Month 9-10, D5: Tasks and baselines for all languages

Phase 2: Integration (Month 10-11)

- Milestone II: Report on ELG integration (Month 11)
- E3C corpus available at ELG (T11)
- Baseline systems available at ELG (T11)

Phase 3: Dissemination (Month 1-12)

- Project web site
- Annotation guidelines published
- Possible organisation of a shared task
- Conference and journal publications

Project Keywords: clinical cases, clinical entities, temporal information, factuality, multilinguality

4.2.2   Project 2 – LSDISCO: Lingsoft Solutions as Distributable Containers

Project Name: Lingsoft Solutions as Distributable Containers

Project Acronym: LSDISCO

Organisation: Lingsoft, Inc.

Lingsoft, Inc. is part of Lingsoft Group, Inc., with a consolidated turnover of about 12.5 million euros in 2019, making it one of the 100 largest language service providers in the world. Lingsoft has offices in Finland and Sweden but operates globally, making available a wide variety of language technology solutions and services designed for the analysis, processing, and utilization of written and spoken language. Lingsoft's core technologies and solutions have been used by tens of millions of users worldwide as part of the Microsoft Office suite of proofing tools.

Project Abstract: Lingsoft is one of the leading providers of language services and solutions in the Nordic countries. Their services are built on their own language technology tools, either proprietary or based on opensource tools. The project goal is to provide Lingsoft tools and services through ELG by packaging them for spelling/grammar, speech recognition, subtitling, named entity recognition, and machine translation for distribution as Docker containers, and making them available through the ELG grid platform.

The project will result in a set of NLP tools for the Nordic languages being available both for public organisations and companies, allowing companies and public organisations throughout Europe to efficiently incorporate Nordic language support in e.g., subtitled videos or customer service chats.

Total Budget: 140,625€

Type of project: A

Projects Schedule: The project will require approximately 10 calendar months. Tasks may be performed in parallel or overlap in calendar time.

Phase 1: Experiment (Month 1–8)

- Planning and Solution Design, Month 1 [Milestone M1: Design]
- Code refactoring to make the tools packageable as Docker containers, Months 1–6 [M2: Distributable code]
- Performed per "technology family":
    - spelling, grammar, lemmatiser, pos tagger, NER, Months 1 – 2
    - speech and subtitling, Months 2–5
    - machine translation, Months 5–6
- Licensing mechanisms, Months 6–7 [M3: Functional licensing model]
- Docker configuration and packaging, Months 7–8 [M4: Docker containers]

Phase 2: Integration (Month 8–10)

- Testing, Month 8
- Distribution to ELG grid platform, Months 9 - 10 [M6: Project completion]

Phase 3: Dissemination (Month 4–10)

- Dissemination and Marketing Plan, Month 4 [M5: Marketing plan]
- Dissemination activities, Months 4–10

Project Keywords: docker, speech recognition, speech-to-text, machine translation, lemmatiser, named entity recognition

4.2.3 Project 3 – MKS-LLOD: MKS as Linguistic Linked Open Data

Project Name: MKS as Linguistic Linked Open Data

Project Acronym: MKS-LLOD

Organisation: Coreon GmbH

Coreon combines knowledge graphs with terminology to create and deploy Multilingual Knowledge Systems (MKS). It makes search, machine learning, and IoT applications interoperable and bridges today's gap between language and knowledge worlds. Unlike terminology tools – that cater only for language – and unlike taxonomy and ontology tools – that cater for knowledge – Coreon unifies knowledge and language in a highly efficient, collaborative, and visual way.

Project Abstract: Coreon has brought the world of knowledge and language together. Multilingual Knowledge Systems (MKS) enrich language data with structure and enlarge knowledge graphs with terms in many languages. An MKS is an important Semantic Interoperability Asset, a repository that allows two systems to exchange information while ensuring that the precise meaning of exchanged information is understood and preserved between parties. Instead of using the Coreon proprietary API, integration into other systems would be more efficient by using standards known from the LLOD community as well as the Semantic Web.

The project vision is to use ELG as a platform for making such multilingual interoperability assets discoverable and retrievable by complementing Coreon with a SPARQL interface.

Total budget: 167,375€

Type of project: A

Project Schedule:

Phase 1: Experiment (Month 1–7)

RDF-ification and SPARQL endpoint: Technology research, SPARQL endpoint development, data modelling and mapping, beta-ready development.

- Authentication methods: technology research and implementation.
- Beta phase: production-ready/finalization, performance tests.
- Management, data curation.

Phase 2: Integration (Month 8–9)

- Containerization: technology research and development.
- Integration tests: deployment, support, and correction.
- Reference repositories: data curation.

Phase 3: Dissemination (throughout and after the project)

- Advertising in standardization communities.
- Announcement through Coreon as well as ELG channels.
- Continued hosting of selected repositories after the project end.

Project Keywords: LLOD, LOD, Semantic Web, Multilingual Knowledge Graph, SPARQL, Terminology Management, RDF data acquisition, and conversion (for further showcasing).

4.2.4 Project 4 – Basque-speaking smart speaker based on Mycroft AI

Project Name: Basque-speaking smart speaker based on Mycroft AI

Project Acronym: Smart euSpeaker

Organisation: Elhuyar Fundazioa

Elhuyar Fundazioa is an NGO devoted to the promotion of the Basque Language in all areas of everyday life but especially in the area of science and technology. Thus, one of its main activities is the research, development, and commercialization of language and speech technologies. This activity is carried out by an R&D department of about 15 people, composed of researchers, computer scientists, and linguists. This R&D department has been working in NLP and speech technologies since 2001.

Project Abstract: The speech-driven virtual assistant devices known as smart speakers, such as Amazon Echo and Google Home, are being increasingly used. But these commercial products come in just over a handful of languages. Not all official languages of the EU, let alone minority languages, are present in them. The Smart euSpeaker project aims to develop an open-source smart speaker that works in the Basque language (ISO-639 code: eu, hence the name), co-official in the Basque Country. It will be open-source, as it will be based on the also open-source Mycroft voice assistant (https://mycroft.ai/). It will make use of Elhuyar's Basque speech synthesis and recognition technology. The work carried out will be of use for other languages that have speech technologies and are underrepresented in smart speakers.

Total Budget: 117,117€

Type of project: B

Project Schedule

Phase 1: Experiment (Month 1–10)

Task 1: Adapting and improving Elhuyar's ASR and TTS systems: August 2020-March 2021

- Adapting Elhuyar's ASR: August 2020-March 2021
- Improvements and additions to Elhuyar's TTS: August 2020-March 2021

Task 2: Adapting Mycroft to Basque: August 2020-May 2021

- Connecting Mycroft to Elhuyar's ASR and TTS: August-October 2020
- Translation of Mycroft to Basque: November 2020-February 2021
- Addition of local and Basque skills: March-May 2021
- Addition of COVID-19 questions skill: March-May 2021

Phase 2: Integration (Month 10–11)

Task 3: Integration into ELG: May-June 2021

- Integrating Elhuyar's ASR and TTS services into ELG: May-June 2021

- Uploading Basque voice assistant to ELG: June 2021

Phase 3: Dissemination (Month 11–12)

Task 4: Dissemination activities: June-July 2021

Milestones:

After Experiment phase:

- M1: Elhuyar´s TTS and ASR adapted
- M2: Basque voice assistant
- After Integration and Dissemination phase:
- M3: Elhuyar´s TTS & ASR and Basque assistant into ELG
- M4: Development of an MVP

Project Keywords: Smart speaker, Basque, Mycroft AI, ASR, TTS

4.2.5 Project 5 – Italian EVALITA Benchmark Linguistic Resources, NLP Services and Tools for the ELG Platform

Project Name: Italian EVALITA Benchmark Linguistic Resources, NLP Services and Tools for the ELG platform

Project Acronym: EVALITA4ELG

Organisation: UNIVERSITA' DEGLI STUDI DI TORINO

The University of Turin (higher education institution) will participate in the project with members of the Content-Centered Computing group (http://di.unito.it/ccc) from the Department of Computer Science, having a long-term experience in natural language processing (NLP). In the last ten years, the applicant team constantly contributed in the development of language resources and technologies for Italian, covering a wide range of linguistic phenomena, text genres (standard and social media) and NLP services, including long-standing tasks (e.g., dependency parsing) and new challenging tasks such as sentiment analysis and hate speech detection (http://hatespeech.di.unito.it/).

Project Abstract: The Italian language is underrepresented in the ELG platform, currently, including few LT services and corpora - mostly parallel corpora and multilingual dependency treebanks focused on texts featured by standard forms and syntax.

The goal of EVALITA4ELG is to enable ELG users to access the resources and models for the Italian language produced over the years in the context of the EVALITA evaluation campaign. Our aim is to build the catalogue of EVALITA resources and tasks ranging from traditional tasks like POS-tagging and parsing to recent and popular ones such as sentiment analysis and hate speech detection on social media, and integrate them in the ELG platform. The project includes the integration of state-of-the-art LT services into the ELG platform, accessible as web services.

Total Budget: 126,125€

Type of project: A

Project Schedule

Phase 1: Experiment (Month 1–8)

The aim is to create a comprehensive catalogue of resources, tasks, and models

- T.1.1 Mapping of tasks and models, and collection of linguistic resources
- T.1.2 Development of web services for pre-trained AlBERTo models, fine-tuned for sentiment analysis and hate speech detection
- T.1.3 Catalogue of available EVALITA systems and development of services encapsulating them

Phase 2: Integration (Month 8–12)

The aim is to give the wider community access to resources and models through the ELG platform.

- T.2.1 Anonymization of resources
- T.2.2 Definition and harmonization of licenses
- T.2.3 Integration of the software developed in WP1

Phase 3: Dissemination (Month 6–12)

- T.3.1 Production of 3-5 scientific papers
- T.3.2 Organisation of a final event including a tutorial
- T.3.2 Participation in one or two scientific events

Milestones

- N.1 - Release of linguistic resources and models (M6)
- N.2 - LT services integration and test (M10)
- N.3 - Final evaluation, including dissemination results (M12)

Project Keywords: Italian language resources, EVALITA, NLP services, benchmarking, language understanding models

4.2.6   Project 6– OPUS-MT: Open Translation Models, Tools and Services

Project Name: Open Translation Models, Tools and Services

Project Acronym: OPUS-MT

Organisation: University of Helsinki

The language technology research group at the University of Helsinki is an active user of HPC resources and its research strongly depends on large-scale computing. They are heavy users of IT resources provided by CSC (the national infrastructure provider) including their GPU cluster nodes and they have also collected significant experience in cross-border collaborations during the time of the NLPL project (http://wiki.nlpl.eu/) with its focus on building a virtual laboratory for NLP related research on HPC facilities. As a continuation, they are also part of the EOSC-nordic project that puts our NLPL initiative into a wider European perspective.

Project Abstract: OPUS-MT will produce state-of-the-art neural machine translation models that can freely be shared, re-used and integrated in open web services and professional translation workflows. The project will focus on European minority languages and their improved support through multilingual NMT models and transfer learning.

Furthermore, OPUS-MT will deliver easily deployable translation services and tools for quick domain-adaptation and on-demand personalisation. We will emphasise open resources that can freely be distributed and used in research and professional applications. For the latter we want to offer local solutions that are independent of on-line services to avoid security risks with open data transfer.

Total Budget: 154,636€

Project type: A

Project Schedule: The duration of the project is 12 months.

Phase 1: Experiment

The first phase will be devoted to data curation, streamlining the data selection pipeline and training procedures to enable large-scale model development with good language coverage and high translation quality. The second phase will focus on multilingual models and transfer learning for improved support of selected European minority languages. We will start that effort by looking at Sami languages and then move our attention to Celtic languages, Germanic language variants, and dialects, and other cases. At the same time, we will also further develop our CAT tool integration and the general web service implementation.

Phase 2: Integration

The final phase is devoted to the ELG integration and setup. This includes populating the resource catalogue and download services as well as the domain-adaptation and translation service deployment.

Phase 3: Dissemination

After successful integration, we will run broad dissemination activities announcing the services and resources.

Project Keywords: Machine translation, minority languages, multilinguality

4.2.7   Project 7 – Text2TCS: Extracting Terminological Concept Systems from Natural Language Text
Project Name: Extracting Terminological Concept Systems from Natural Language Text

Project Acronym: Text2TCS

Organisation: Centre for Translation Studies, University of Vienna

The CTS has a considerable track record in projects related to language technology and terminology, such as the CEF Automated Translation for the EU Council Presidency, the LTO Language Technology Observatory (HORIZON 2020 project) and CLARIN-AT, where the CTS contributed to the CLARIN European Research Infrastructure Consortium for the Common Language Resource and Technology Network.

Project Abstract: The applicant proposes to bring semantic and terminology technologies together for the benefit of improving automated terminology extraction. Current terminology extraction tools extract isolated terms or concepts but not their interrelations with very few exceptions, e.g., extracting hierarchical relations, such as broader. The proposed Text2TCS application automatically extracts hierarchical and semantic relations from text corpora to create a Terminological Concept System (TCS). It will rely on findings from ontology learn-

ing and machine learning. An automatically created TCS can foster cross-border communication in times of crisis, such as COVID-19, and help to avoid misunderstandings due to terminological inconsistency. Additionally, this type of technology is not yet part of the ELG.

Total Budget: 132,977€

Project type: A

Project Schedule

Phase 1: Experiment (Month 1–8)

07/2020 - 09/2020: data/corpora collection, first TCS extraction experiments and determination of framework

- MS1: determination of data sets, word embeddings and programming framework

10/2020 - 01/2020: systematic investigation of approach, a continuation of a collection of data

- MS2: finalizing experiments, word embedding selection, and preparation of conference publication

01/2021 - 02/2021: final adaptations of application, visualization of TCS, and usability studies and user experiments

- MS3: finalization of usability studies and user experiments
- MS4: final code on GitLab including metadata descriptions

Phase 2: Integration (Month 9–12)

03/2021 - 06/2021: Integration and final usability tests on the platform

- MS5: Docker image generation and testing and integration into ELG

Phase 3: Dissemination (Month 8–12)

02/2021 - 06/2021: Parallel dissemination activities in the mentioned networks and at the proposed conference

- MS6: Conference presentation and dissemination in networks

Project Keywords: Terminology extraction, semantic relations, concept system, ontology learning, machine learning

4.2.8   Project 8 – PARA4DLM: Textual paraphrase dataset for deep language modelling

Project Name: Textual paraphrase dataset for deep language modelling

Project Acronym: PARA4DLM

Organisation: University of Turku, TurkuNLP research group

The TurkuNLP Natural Language Processing group (https://turkunlp.org) focuses on research and education in the fields of language technology and natural language processing. It has a long track record in open resource development - both in terms of datasets and software pipelines. The datasets the group created range from high-quality manually annotated datasets such as the Turku Dependency Treebank, to large scale crawl-based

data such as the 40-language, 90-billion token supporting dataset for the CoNLL'17 Shared Task on Dependency Parsing. The group is also closely involved in the Universal Dependencies project. Methods developed by the group have attained numerous top ranks in various domain shared tasks.

Project Abstract: The applicant proposes to develop and distribute a large dataset of at least 100,000 pairs of statements with equal contextual meaning (paraphrases) and deep learning models trained on this dataset. The crucial property of the dataset is that the selection of the data pairs will be controlled to minimise their lexical and structural similarity, resulting in data better suited for the training and evaluation of deep language models for natural language understanding. The dataset will also directly support research on tasks that require paraphrasing, such as information retrieval, text generation, machine translation, and many others. In this pilot project, we will focus on Finnish as the primary language, with a separate smaller test set for Swedish to facilitate cross-lingual studies.

Total Budget: 166,085€

Project Schedule: The project is scheduled for 12 months.

Phase 1: Experiment (Months 1–10)

Given that we have secured large amounts of initial source data, as well as developed and tested an initial annotation protocol, the experiment phase in the form of dataset creation can be started from day 1 and is expected to continue throughout the project.

Phase 2: Integration (Months 8–11)

The reusable codebase and its documentation will be prepared, the appropriate metadata produced, and the resources fully integrated into ELG.

Phase 3: Dissemination (Months 9–12)

Documentation, use tutorials, and scientific article(s) will be published describing the outcome of the project and assessing the expected gains from replicating this pilot study for other languages. The outcome will be presented at scientific conferences, and an online-accessible workshop presenting the results will be organised.

Project Keywords: natural language understanding, paraphrase, fine-tuning dataset, deep learning, Finnish, Swedish

4.2.9   Project 9 – YouTwinDi: Virtual Personal Assistant Prototype

Project Name: Virtual Personal Assistant Prototype

Project Acronym: YouTwinDi

Organisation: Weber Consulting KG

Weber Consulting KG is an SME focusing on conversational AI and has developed a chatbot framework, which is open source and can be used to create, maintain, and run multiple chatbots. Its experience is related to natural language processing, which is at the core of artificial intelligence and conversation design. Its chatbot framework was developed over the last 12 years.

Project Abstract: YouTwinDi is the next step in human-computer interaction: a digitised world where the digital twin evolves and interacts with other digital twins and makes autonomous decisions in the interest of its human twin. A world where security and digital ethics assure ethical decisions and where IT specialists concur on enhancing the digital landscape with ethical models. A world where language barriers are removed and a continuous match of demand and offer and tailored searches help the human twin to improve his life in all aspects. YouTwinDi will use the most advanced technologies in the domain of real-time language translation and analysis, allowing the user and its digital twin to interact with all the European citizens without being blocked by any language barrier.

Total Budget: 87,445€

Project type: B

Project Schedule: We have divided the project into 5 work packages and each work package into several subtasks.

Phase 1: Experiment

Work package 1: Research & Experiment: SW Research, HW Research, Use Case Definition

Phase 3: Integration

Work package 2: Software Integration: Integration EDDI and ELG, Container setup, Use case Implementation, Testing

Work package 3: Hardware Integration: Preparation of HW, Installation of SW, Integration Tests

Work package 4: Final Prototyping: Finalization of Prototype, Final Tests, Prototype presentation preparation

Phase 3: Dissemination

Work package 5: Dissemination: Set-up of project website, ongoing information updates via website, Social media

WP 5 is related to the dissemination of the project results and will start immediately as the project starts in month 1 and will last until month 12 (end of the project).

Project Keywords: Digital Twin, Virtual Assistant, Chatbot, Open Source, Natural Language Processing, Conversational Artificial Intelligence, Enhanced Dialog Driven Intelligence

4.2.10 Project 10 – SLAPMAN: Streaming Language Processing in Manufacturing
Project Name: Streaming LAnguage Processing in Manufacturing

Project Acronym: SLAPMAN

Organisation: FZI Research Center for Information Technology

FZI Research Center for Information Technology is a non-profit institution for applied research in information technology and technology transfer (independent research organisation). Its task is to provide businesses and public institutions with the latest research findings in information technology. In SLAPMAN, FZI involves the department "Knowledge Management", which has a long experience related to topics such as (Semantic) Data

Management, Stream Processing, and Machine Learning. The focus is on extending and applying the latest research findings in application areas around the Industrial Internet of Things. The involved researchers are experts in distributed stream processing, knowledge engineering, and interactive machine learning on multi-modal data.

Project Abstract: Often underestimated, (semi-) structured textual data sources are an important cornerstone in the manufacturing sector for product and process quality tracking. SLAPMAN develops novel methods for industrial text analytics in the form of scalable, reusable, and potentially stateful microservices, which can be easily orchestrated by domain experts in order to define quality anomaly patterns, e.g., by analyzing machine states and error logs.

The results will be fully available as open-source and integrated into the IIoT toolbox Apache StreamPipes, which was originally created by FZI and is now an incubating project in the Apache Software Foundation. A custom-tailored version of StreamPipes, including the aforementioned modules, will be contributed as a cloud-native application to the ELG.

Total Budget: 132,160€

Project type: A

Project Schedule: The work is organised into 4 work packages:

Phase 1: Experiment (Month 1–10)

- WP1: Requirements (M1)
    - Workshop with ELG community
    - Identification of existing and required LT technologies
    - Technical architecture & integration roadmap
- WP2: Experiment (M2 – M10)
    - Development of adapters for textual data sources
    - Development of streaming LT analytics microservices
    - Integration of existing LT analytics microservices
    - Domain-specific learning module for industrial text analytics tasks
    - Validation of learning modules and streaming algorithms in an experimental setting

Phase 2: Integration (Month 7–12)

- WP3: Integration (M7 – M12)
    - Integration of LT modules into Apache StreamPipes
    - Preparation of custom StreamPipes version for ELG
    - Integration of Apache StreamPipes into ELG
    - Helm Charts for Kubernetes deployment & deployment to ELG platform

Phase 3: Dissemination

- WP4: Dissemination (M1 – M12)
    - Scientific dissemination, developer community and end-users

Milestones:

- M1 Requirements defined
- M7 First version of components available
- M12 System tested and deployed

Project Keywords: Industrial Text Analytics, Natural Language Processing, Streaming Data, Self-Service, Analytics, Open Source

5   Future steps: Open Call 2

5.1   Timing

The second ELG open call will be opened on 1 October 2020 — one month later than initially planned. The main reason for this postponement is to get more time to collect feedback on the first open call, perform an analysis of the selected pilot projects and reflect it in the update of the call documentation for the second open call.

The submission period will last for two months until the end of November, and the proposal submission deadline is 30 November 2020 (see the timeline in Figure 1). The total funding sum available for this call is 585,000€. We expect to select 3 to 5 project proposals that will start their execution phase in February 2021.

5.2   Proposed changes

The basic parameters given by the ELG grant agreement will remain the same for the second open call: We will look for pilot projects that broaden the ELG's portfolio of language technologies. The projects will 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 can be up to 200,000 EUR, and the runtime of the projects is expected to be between 9 and 12 months. SMEs and research organisations are eligible to apply, and the projects will be non-consortial, i.e., only one single organisation or institution is allowed per project.

However, we plan these minor changes in the call documentation and the open call procedure:

- to improve the explanation of ELG's strategic goals and goals of the open call
- to improve the description of ELG infrastructure and easy to find a list of currently available services
- to organize a webinar for proposers during the submission period
- to fine-tune the call documentation, annexes, templates, and forms along with ELG Open Calls Platform
- to request a budget breakdown in a fixed structure and a detailed budget justification and explanation
- to recruit new evaluators, with the aim to attract more high-quality and experienced evaluators
- possibly to focus the second open call stronger thematically (e.g., on missing data sets, technologies or services). The discussion on that decision is ongoing.

6   Conclusions

The project proposals submitted in the first open call were evaluated, 10 pilot projects were selected for financial support, and these projects have successfully started their execution phase.

In Table 2, the type of submitted projects and applying organisations is shown. For projects selected for financing, a similar overview is provided in Table 3. Although we obtained 56% of proposals from SMEs and 44% from research organisations, among the supported projects are only 3 SMEs and 7 research organisations. This is probably because research organisations have more experience with applying for financial support, and thus, they formulate their plans more concisely. Also, although B type projects from SMEs were preferred, only two B type projects have been accepted for financing, but the ratio among submitted projects is similar (72% of A type projects and 28% of B type projects).

Most of the pilot projects support underrepresented languages. The selected organisations are from five different EU countries. The requested budget varies from 87,445€ to 167,375€, with an average value of 136,392€.

The announcement of the second open call is planned on 1 October 2020, and selected pilot projects should start in February 2021. The basic conditions of the second open call remain the same as for the first open call; some minor changes are detailed in section 5.2 of this document.

Submitted by	Type A	Type B	Total
Research organisation	43 (39.1%)	5 (4.5%)	48 (43.6%)
SME	36 (32.7%)	26 (23.6%)	62 (56.4%)
Total	79 (71.8%)	31 (28.1%)	110 (100%)

Table 2: Proposals submitted for the first open call and accepted for evaluation

Submitted by	Type A	Type B	Total
Research organisation	6 (60%)	1 (10%)	7 (70%)
SME	2 (20%)	1 (10%)	3 (30%)
Total	8 (80%)	2 (20%)	10 (100%)

Table 3: Proposals selected for financial support in the first open call

A.   Annexes

	Document	Description	No. of Pages
1	Analysis of the Survey for Proposers to the Open Call 1	Results of two short surveys designed for those who submitted a proposal (proposers) to the first open call and those who uploaded an initial draft but did not submit a final version (non-proposers).	16
2	Results of the Survey on Evaluation Process of the First ELG Open Call	Results of survey on evaluation process conducted among all involved evaluators.	7
3	Guide for pilot projects	The document which will be 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.	4

Annex 1: Analysis of the Survey for Proposers to the Open Call 1

European Language Grid
Open Call 1 – Analysis of the Survey for Proposers

European Language Grid – Analysis of the Survey

for Proposers to the Open Call 1

Two short surveys were designed for those who submitted a proposal (proposers) and those who uploaded an initial draft but did not submit a final version (non-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 organisation". All information was collected anonymously with the goal to evaluate and improve our call processes.

Of the proposers, 73 out of 110 (66 %) responded and of the non-proposers 6 out of 17 (35 %) responded.

Data from Non-Proposers: https://docs.google.com/spreadsheets/d/1sIqCild5HM684sdCv4KnWn840jgykL-sicpW_uqLXJ88/edit#gid=2101180733

Data from Proposers: https://docs.google.com/spreadsheets/d/1aa-c_IIoodYSxNMpyVxtwMoPFqt4jjRh72j4pBDGbvs/edit#gid=1950530074

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.

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	54	74.0
Further development of an existing software or data project	48	65.8
Necessity of financial funding in order to realise a specific project idea	35	47.9
Get more visibility for my organisation and/or products and/or services through the ELG	31	42.5
General interest in the further development and progress of the ELG platform and initiative	27	37.0
Branch out into new markets by means of the ELG platform and initiative	20	27.4
Make use of the ELG cloud platform to distribute my services (my organisation does not run its own cloud platform)	15	20.5
General interest in EU-funded innovation actions	12	16.4
Gather experience with EU-funded open call FSTP projects (FSTP, Financial Support for Third Parties)	5	6.8
Other	5	6.8

Question 2: Are you interested in ELG primarily because of the (functional) services or because of the datasets?

European Language Grid
Open Call 1 – Analysis of the Survey for Proposers

Are you interested in ELG primarily because of the (functional) services or be-cause of the datasets?	Answers	Answers (%)
Both services and datasets	51	69.9
(Functional) services	16	21.9
Datasets	5	6.8
No opinion	1	1.4

Question 3: What are your main expectations towards the ELG platform and initiative? Please choose up to three options.

What are your main expectations towards the ELG platform and initiative? Please choose up to three options.	Answers	Answers (%)
That I get access to a large repository of functional services and datasets	40	54.8
That ELG strengthens the LT and language-centric AI community in Europe as a whole	38	52.1
That ELG provides increased visibility for my organisation on the European level	36	49.3
That ELG facilitates future collaborations with other developers by establishing a common platform	30	41.1
That ELG becomes a part of a larger LT/AI platform eco-system in Europe	30	41.1
That I have an additional channel for the exploitation of my organisation's research results	29	39.7
That ELG improves interoperability of LT by establishing a common API	25	34.2
That ELG serves as an information hub and matchmaker for buyers and suppliers of LT	21	28.8
That I have an additional sales channel for my organisation's commercial services or datasets	20	27.4
Other	1	1.4
No answer	1	1.4

Question 4: What should be the focus of the ELG platform and initiative in the next 3-5 years?

This question generated 61 responses. One answer cannot be counted as valid.

What should be the focus of the ELG platform and initiative in the next 3-5 years?	Answers	Answers (%)
Increase the variety and range of services and datasets for a more diverse LT landscape, covering the whole of Europe (also minority languages); establish a recognised one-stop platform	18	30.0
Create visibility and credibility	6	10.0
Easy access; ELG should focus on creating an easy-to-join and easy-to-use infrastructure, so that everybody with relevant resources or needs could join	6	10.0
Focus on language-centric AI	6	10.0
Standardization of Language Technologies data representations and functionalities/; provide standardized unified API for multi-modal language transformation services to be easily used and cost-effective	5	8.3

ELG 2/16

Interoperability with non ELG technologies (be open, not exclusive); connect all European LT initiative	5	8.3
Widen the audience by vertical industry experts who can contribute to services and take advantage of technologies; attract more LT providers	4	6.7
Include not only Europe; cross-lingual focus important	4	6.7
Focus on NLU	3	5.0
Community building	3	5.0
NLP tools to help people to empower themselves with better education; to build a more emphatic human-robot interaction	2	3.3
Speech resources and tools integrated with image and video processing	2	3.3
Larger hardware support for running Language Technologies; sustainability of the LT running in the cloud	1	1.7
Act as data and service hub and information platform with a focus on unified APIs (TAPICC initiatives)	1	1.7
ELG should foster the exchange of base-technologies and services upon which teams can then build their custom-tailored NLP solutions	1	1.7
IoT	1	1.7
Be self-sustaining	1	1.7
To provide LT services including accessibility services like sign language transla-tions	1	1.7
Increase funding	1	1.7

1.2 Project Proposals

Question 5: 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 organisation, department or team participate in?

Number of agile project proposals	Answers	Answers (%)
2 agile project proposals	16	21.9
1 agile project proposals	14	19.2
0 agile project proposals	12	16.4
3 agile project proposals	8	11.0
5 agile project proposals	8	11.0
6 agile project proposals	4	5.5
10 agile project proposals	4	5.5
4 agile project proposals	3	4.1
50 agile project proposals	2	2.7
20 agile project proposals	1	1.4
30 agile project proposals	1	1.4

Question 6: In a typical year, how many consortia-based project proposals (i.e., typical EU Horizon 2020 project proposals) does your organisation, department or team participate in?

Number of consortia-based project proposals	Answers	Answers (%)

⊪ᆥᆥ ELG
European Language Grid
Open Call 1 – Analysis of the Survey for Proposers

1 consortia-based project proposal	22	30.14
2 consortia-based project proposal	14	19.18
0 consortia-based project proposal	13	17.81
3 consortia-based project proposal	10	13.70
10 consortia-based project proposal	5	6.85
4 consortia-based project proposal	2	2.74
5 consortia-based project proposal	2	2.74
20 consortia-based project proposal	2	2.74
8 consortia-based project proposal	1	1.37
16 consortia-based project proposal	1	1.37
50 consortia-based project proposal	1	1.37

Question 7: The first ELG open call was very popular. Do you think more EU-funded activities dedicated to Language Technology and Language-centric AI are needed?

The first ELG open call was very popular. Do you think more EU-funded activities dedicated to Language Tech-nology and Language-centric AI are needed?	Answers	Answers (%)
Yes	67	91.8
No opinion	6	8.2
No	0	0.0

Question 8: Does your organisation have a preference for smaller, more agile calls like the ELG open call or for complex, consortia-based projects?

Does your organisation 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)	49	67.1
No preference	16	21.9
Consortia-based projects (long proposals, complex evaluation, 24-36 months runtime)	8	11.0

Question 9: Do you have feedback for us how we can, based on your own experience, improve the second ELG Open Call, which will be launched in September 2020?

This question generated 45 responses. One answer cannot be counted as valid.

Do you have feedback for us how we can, based on your own experience, im-prove the second ELG Open Call, which will be launched in September 2020?	Answers	Answers (%)
Strategic goals of the ELG were hard to interpret (both on the website and ma-terial); needs to be specific what kind of work is expected to be performed in ELG	8	17.8
No further recommendations at this point	8	17.8

ELG   4/16

European Language Grid
Open Call 1 – Analysis of the Survey for Proposers

ılıılı ELG

Better availability and information of already existing services (e.g. through as webinar); clear documentation and more details about the API integrations for ELG and how the infrastructure working with data, applications and workflows	6	13.3
Improvement of webform (limited space, too restrictive, could be more user-friendly)	6	13.3
Increase budget to fund more projects	3	6.7
More details in guidelines related to specification of budget, eligible cost, etc. (sample proposal could be useful)	2	4.4
Provide good feedback for the first round for all proposals, so that everybody better understands the requirements for a successful proposal	2	4.4
Definition of thematic priorities since it was difficult to assess from the outside whether the general idea of a proposal might be attractive for the ELG at all	1	2.2
Request submission of a 8-10 slide presentation of the proposal to measure how relevant a proposal is	1	2.2
Consider accessibility topic for LT	1	2.2
Focus on advanced LT technologies targeting the future of (and less more traditional LT)	1	2.2
Take language types not represented in the EU into consideration to give a more comprehensive picture about the field	1	2.2
Be more specific on the number of projects to be funded and on the expected budget	1	2.2
Match research team topics with ELG roadmap	1	2.2
Ensure connecting ELG to EU based, growing platforms supporting European languages already and promoting ELG community at the same point in time	1	2.2
Implement an automatic email confirmation after final proposal submission	1	2.2
Website is not very user-friendly, figuring it out is very time costly. It gives the impression of lack of transparency and attempting to hide something.	1	2.2
The objectives (consume data or services from ELG or contribute to) were mutually exclusive, although the best type of project does both.	1	2.2

1.3   Your Organisation

Question 10: Did you submit the proposal on behalf of a SME or a research institution?

Did you submit the proposal on behalf of a SME or a research institution?	Answers	Answers (%)
Prefer not to say	1	1.4
Research institution	32	43.8
SME	40	54.8

Question 11: What types of Language Technology do you specialise in, for example, Machine Translation, Text Analytics, Speech Recognition etc.? Please provide a comma-separated list of technologies.

This question generated 70 responses. One answer cannot be counted as valid.

What types of Language Technology do you specialise in, for example, Machine Translation, Text Analytics, Speech Recognition etc.? Please provide a comma-separated list of technologies.	Answers	Answers (%)

European Language Grid
Open Call 1 – Analysis of the Survey for Proposers

Text Analysis	34.0	48.6
Machine Translation	17.0	24.3
Speech Recognition	13.0	18.6
Natural Language Processing	10.0	14.3
TTS	6.0	8.6
ASR	4.0	5.7
Conversational AI/Chatbots	4.0	5.7
Data Sets	3.0	4.3
Dialog Management/Modeling/Systems	3.0	4.3
NER	3.0	4.3
NLG	3.0	4.3
Information Extraction	2.0	2.9
Machine Learning	2.0	2.9
NLU	2.0	2.9
Speech Synthesis	2.0	2.9
Summarization	2.0	2.9
Text Preprocessing	2.0	2.9
UI/User-front end	2.0	2.9
Academic Tools	1.0	1.4
API Development	1.0	1.4
Artificial Intelligence	1.0	1.4
CAT tools	1.0	1.4
Classification	1.0	1.4
Clustering	1.0	1.4
Computational Terminology	1.0	1.4
Computer Vision	1.0	1.4
Corpus Linguistics	1.0	1.4
Cross-lingual resources	1.0	1.4
Digitisation	1.0	1.4
Keyword Extraction	1.0	1.4
Knowledge Graphs	1.0	1.4
Language Identification	1.0	1.4
Lexical Resources	1.0	1.4
Localisation	1.0	1.4
Low-resource learning	1.0	1.4
Misinformation Processing	1.0	1.4
Multilingual Data	1.0	1.4
Parsing	1.0	1.4
Phone Call Automation	1.0	1.4
Research Data Management	1.0	1.4
Resource Creation	1.0	1.4
Root Language Elaboration	1.0	1.4

ELG   6/16

Semantic Analysis	1.0	1.4
Semantic Annotation	1.0	1.4
Semantic Labeling	1.0	1.4
Sentiment Analysis	1.0	1.4
Sign Language Translation	1.0	1.4
Simplification	1.0	1.4
Speech Analysis	1.0	1.4
Speech Generation	1.0	1.4
Speech Perception	1.0	1.4
Speech Technology	1.0	1.4
Subtitling	1.0	1.4
Supervised Learning	1.0	1.4
Terminology Extraction	1.0	1.4
Text Classification	1.0	1.4
Text Mining	1.0	1.4
Topic Analysis	1.0	1.4
Translation Management	1.0	1.4
Usability	1.0	1.4
WSD	1.0	1.4

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

This question generated 54 responses. One answer cannot be counted as valid.

Do you specialise in certain domains, for example, energy, health, mobility? Please provide a comma-separated list of domains.	Answers	Answers (%)
No certain domains	18	33.3
Health	8	14.8
Legal	4	7.4
Media	4	7.4
Mobility	4	7.4
Energy	3	5.6
Linguistics	3	5.6
News	3	5.6
Automotive	2	3.7
Broadcast	2	3.7
Digital Humanities	2	3.7
Education	2	3.7
Public Services	2	3.7
Forensic	1	1.9
Academic papers	1	1.9
Audiovisual Media	1	1.9

ⅱⅼ ELG

European Language Grid
Open Call 1 – Analysis of the Survey for Proposers

Banking	1	1.9
Cognitive Robotics	1	1.9
Cultural Heritage Preservation	1	1.9
e-Governance	1	1.9
Entertainment	1	1.9
Fake News Detection	1	1.9
Finance	1	1.9
Government	1	1.9
ICT	1	1.9
Industrial automation- COBOTS	1	1.9
Industry	1	1.9
Informatics	1	1.9
Journalism	1	1.9
Language Technologies	1	1.9
Manufacturing	1	1.9
Mining	1	1.9
Paleography	1	1.9
Pharmaceutical	1	1.9
Politics	1	1.9
Retail	1	1.9
Scientific Texts	1	1.9
Smart Cities	1	1.9
Smart Enterprises	1	1.9
Smart Home	1	1.9
Sustainability	1	1.9
Telecommunication	1	1.9
Top Management	1	1.9
Tourism	1	1.9
Training	1	1.9

Question 13: Please specify up to five languages your organisation 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 71 responses. Six answers cannot be counted as valid.

Language(s)	Mentions	Mentions (%)
Group A (Official EU Languages): English, German, French, Spanish, Italian, Polish, Portuguese, Dutch, Bulgarian, Greek, Romanian, Czech, Estonian, Slovak, Croatian, Finnish, Danish, Hungarian, Irish, Lithuanian, Latvian, Maltese, Slovenian, Swedish	203	
English	47	72.3
German	28	43.1
French	27	41.5

ELG

8/16

European Language Grid
Open Call 1 – Analysis of the Survey for Proposers

Spanish	22	33.8
Italian	14	21.5
Polish	7	10.8
Portuguese	6	9.2
Dutch	6	9.2
Bulgarian	5	7.7
Greek	5	7.7
Romanian	5	7.7
Czech	4	6.2
Estonian	4	6.2
Slovak	4	6.2
Croatian	3	4.6
Finnish	3	4.6
Danish	2	3.1
Hungarian	2	3.1
Irish	2	3.1
Lithuanian	2	3.1
Latvian	2	3
Maltese	1	1.5
Slovenian	1	1.5
Swedish	1	1.5
Group B (Other EU languages; languages from EU candidate countries and Free Trade Partners): Basque, Catalan, Serbian, Norwegian, Galician, German sign language, Occitaine, Turkish	17	
Basque	4	6.2
Catalan	3	4.6
Serbian	3	4.6
Norwegian	3	4.5
Galician	1	1.5
German sign language	1	1.5
Occitaine	1	1.5
Turkish	1	1.5
Group C (Language spoken by EU immigrants; languages of important trade and political partners): Russian, Arabic, Chinese, Swahili, Sanskrit, Sign languages, Ukrainian, Japanese, American Sign language, Latin, Afrikaans, Belarus, Bosnian, Finno-Ugric endangered languages, Global languages, Hebrew, isiXhosa, isiZulu, Korean, Pali, Sesotho sa Leboa, Setswana, Songhay, Vedic, Vietnamese, Yiddish	40	
Russian	6	9.2
Arabic	3	4.6
Chinese	3	4.6
Swahili	2	3.1
Sanskrit	2	3.1
Sign languages	2	3.1

ELG

9/16

ılıₗ'ELG

European Language Grid
Open Call 1 – Analysis of the Survey for Proposers

Ukrainian	2	3.1
Japanese	2	3
American Sign language	1	1.5
Latin	1	1.5
Afrikaans	1	1.5
Belarus	1	1.5
Bosnian	1	1.5
Finno-Ugric endangered languages	1	1.5
Global languages	1	1.5
Hebrew	1	1.5
isiXhosa	1	1.5
isiZulu	1	1.5
Korean	1	1.5
Pali	1	1.5
Sesotho sa Leboa	1	1.5
Setswana	1	1.5
Songhay	1	1.5
Vedic	1	1.5
Vietnamese	1	1.5
Yiddish	1	1.5

Question 14: Please specify the country in which your organisation is based.

This question generated 70 responses.

Please specify the country in which your organisation is based.	Answers	Answers (%)
Germany	10	14.3
Spain	8	11.4
Austria	4	5.7
France	4	5.7
Greece	4	5.7
Italy	4	5.7
Belgium	3	4.3
Bulgaria	3	4.3
Portugal	3	4.3
Romania	3	4.3
Czech Republic	2	2.9
Estonia	2	2.9
Ireland	2	2.9
Latvia	2	2.9
Slovakia	2	2.9
Slovenia	2	2.9

ELG   10/16

Switzerland	2	2.9
UK	2	2.9
Denmark	1	1.4
Finland	1	1.4
Hungary	1	1.4
Norway	1	1.4
Poland	1	1.4
Serbia	1	1.4
South Africa	1	1.4
the Netherlands	1	1.4

Question 15: Do you have any other comments or suggestions you would like to share?

This question generated 27 responses.

Do you have any other comments or suggestions you would like to share?	Answers	Answers (%)
No comment	14	51.9
Thank you for the good work!	7	25.9
NLU one of the most important research topic, it needs lots of attention	1	3.7
LT is a very broad field. It might be helpful to limit the range of applications down further (e.g. Text Analytics, Text Generation, Machine Translation ... etc.) to permit better focusing in proposal selection.	1	3.7
There is a lack of projects proposals like the ELG ; need Open Calls and coverage for more languages	2	7.4
Important to consider people with linguistic disabilities	1	3.7
Important to improve and standardize datasets	1	3.7
Please employ UX professionals, using a Design Thinking approach, to redesign your website and other user assets to make them truly useful and usable	1	3.7

2 Survey Analysis – Non-Proposers

In this chapter, the overall overview of Open Call 1 is provided 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 Motivation

Question 1: Why did you abandon your draft proposal?

Why did you abandon your draft proposal?	Answers	Answers (%)
Too little time to prepare the final proposal	5	83.3
Draft replaced later by a newer version, which was, in fact, submitted	1	16.7
Submission system too complicated (technical issues)	1	16.7
Submission procedure too complicated (process issues)	1	16.7

Available budget too small	1	16.7
Other	1	16.7

Question 2: Are you interested in ELG primarily because of the (functional) services or because of the da-tasets?

Are you interested in ELG primarily because of the (functional) services or be-cause of the datasets?	Answers	Answers (%)
Both services and datasets	4	66.7
(Functional) services	1	16.7
No opinion	1	16.7
Datasets	0	0.0

Question 3: What are your main expectations towards the ELG platform and initiative? Please choose up to three options.

What are your main expectations towards the ELG platform and initiative? Please choose up to three options.	Answers	Answers (%)
That I get access to a large repository of functional services and datasets	5	83.3
That ELG strengthens the LT and language-centric AI community in Europe as a whole	1	16.7
That ELG provides increased visibility for my organisation on the European level	3	50.0
That ELG facilitates future collaborations with other developers by establishing a common platform	2	33.3
That ELG becomes a part of a larger LT/AI platform eco-system in Europe	1	16.7
That I have an additional channel for the exploitation of my organisation's re-search results	3	50.0
That ELG improves interoperability of LT by establishing a common API	1	16.7
That ELG serves as an information hub and matchmaker for buyers and suppli-ers of LT	0	0.0
That I have an additional sales channel for my organisation's commercial ser-vices or datasets	0	0.0
Other	0	0.0
No answer	0	0.0

Question 4: What should be the focus of the ELG platform and initiative in the next 3-5 years?

This question generated 5 responses. One answer cannot be counted as valid.

What should be the focus of the ELG platform and initiative in the next 3-5 years?	Answers	Answers (%)
Develop ELG further with regard to available services and visibility in Europe	1	25
More visibility for smaller languages	1	25
App for collaborative learning	1	25

⫼⫼ᴵ ELG

European Language Grid
Open Call 1 – Analysis of the Survey for Proposers

Interoperability between the available resources	1	25

2.2   Project Proposals

Question 5: 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 organisation, department or team participate in?

Number of agile project proposals	Answers	Answers (%)
2 agile project proposals	3	50.0
4 agile project proposals	1	16.7
10 agile project proposals	1	16.7
14 agile project proposals	1	16.7

Question 6: In a typical year, how many consortia-based project proposals (i.e., typical EU Horizon 2020 project proposals) does your organisation, department or team participate in?

Number of consortia-based project proposals	Answers	Answers (%)
0 consortia-based proposal	1	16.67
2 consortia-based proposal	3	50.00
3 consortia-based proposal	1	16.67
15 consortia-based proposal	1	16.67

Question 7: The first ELG open call was very popular. Do you think more EU-funded activities dedicated to Language Technology and Language-centric AI are needed?

The first ELG open call was very popular. Do you think more EU-funded activities dedicated to Language Technology and Language-centric AI are needed?	Answers	Answers (%)
Yes	5	83.3
No opinion	1	16.7
No	0	0.0

Question 8: Does your organisation have a preference for smaller, more agile calls like the ELG open call or for complex, consortia-based projects?

Does your organisation 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)	5	83.3
No preference	1	16.7
Consortia-based projects (long proposals, complex evaluation, 24-36 months runtime)	0	0.0

Question 9: Do you have feedback for us how we can, based on your own experience, improve the second ELG Open Call, which will be launched in September 2020?

This question generated 5 responses. One answer cannot be counted as valid.

Do you have feedback for us how we can, based on your own experience, improve the second ELG Open Call, which will be launched in September 2020?	Answers	Answers (%)
No feedback at this point	1	25
During the preparation phase inform the proposers if there is another proposal being prepared by the same organisation.	1	25
Advertise more actively, organisation found out about the call too late.	1	25
Explain how to involve the targeted customers into the proposal. What is expected from the business plan as impact and ambition?	1	25

2.3 Your Organisation

Question 10: Did you submit the proposal on behalf of a SME or a research institution?

Did you submit the proposal on behalf of a SME or a research institution?	Answers	Answers (%)
Research institution	4	66.7
SME	2	33.3
Prefer not to say	0	0.0

Question 11: What types of Language Technology do you specialise in, for example, Machine Translation, Text Analytics, Speech Recognition etc.? Please provide a comma-separated list of technologies.

This question generated 6 responses.

What types of Language Technology do Text Analytics, Speech Recognition etc.? Please provide a comma-separated list of technologies.	Answers	Answers (%)
Text Analytics	2	33.3
Cat tools	1	16.7
Cloud services	1	16.7
Collaborative language learning	1	16.7
Dictionaries	1	16.7
Linguistic research and systematisation of data	1	16.7
Natural language interfaces	1	16.7
OCR	1	16.7
Speech recognition	1	16.7
Text annotation	1	16.7

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

This question generated 6 responses.

⫼ ELG

European Language Grid
Open Call 1 – Analysis of the Survey for Proposers

Do you specialise in certain domains, for example, energy, health, mobility? Please provide a comma-separated list of domains.	Answers	Answers (%)
Mobility	2	33.3
Security	2	33.3
Archives	1	16.7
Business information delivery	1	16.7
Education	1	16.7
General domain	1	16.7
Health	1	16.7
Language learning	1	16.7
Linguistics	1	16.7
Military	1	16.7
Open data	1	16.7
Water management	1	16.7

Question 13: Please specify up to five languages your organisation 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 6 responses.

Please specify up to five languages your organisation is primarily interested in. Please state the name of the languages in full, separated with a comma, e.g.: Maltese, Hungarian, French.	Answers	Answers (%)
English	5	83.3
French	4	66.7
German	4	66.7
Bulgarian	2	33.3
Arabic	1	16.7
Basque	1	16.7
Chinese	1	16.7
Estonian	1	16.7
Hebrew	1	16.7
Hungarian	1	16.7
Mandarin	1	16.7
Russian	1	16.7
Slavic languages	1	16.7
Spanish	1	16.7
Swahili	1	16.7
Ukrainian	1	16.7

Language(s)	Answers	Answers (%)
Group A (Official EU Languages): English, French, German, Bulgarian, Estonian, Hungarian, Spanish	18	

ELG   15/16

European Language Grid
Open Call 1 – Analysis of the Survey for Proposers

English	5	83.3
French	4	66.7
German	4	66.7
Bulgarian	2	33.3
Estonian	1	16.7
Hungarian	1	16.7
Spanish	1	16.7
Group B (Other EU languages; languages from EU candidate countries and Free Trade Partners): Basque	1	
Basque	1	16.7
Group C (Language spoken by EU immigrants; languages of important trade and political partners): Arabic, Swahili, Ukrainian, Chinese, Hebrew, Mandarin, Russian, Slavic languages	8	
Arabic	1	16.7
Swahili	1	16.7
Ukrainian	1	16.7
Chinese	1	16.7
Hebrew	1	16.7
Mandarin	1	16.7
Russian	1	16.7
Slavic languages	1	16.7

Question 14: Please specify the country in which your organisation is based.

Please specify the country in which your organisation is based.	Answers	Answers (%)
Bulgaria	2	33.3
Poland	1	16.7
Ukraine	1	16.7
Germany	1	16.7
Italy	1	16.7

Question 15: Do you have any other comments or suggestions you would like to share?

This question generated 3 responses. One answer cannot be counted as valid.

Do you have any other comments or suggestions you would like to share?	Answers	Answers (%)
I would suggest you create a survey/ask for feedback regarding the proposal template, since some parts were ambiguous. In other words, it was not clear what should be described where. There were parts which were requesting description of same aspects.	1	50
Glad to be part of the ELG community	1	50

Annex 2: Results of the Survey on Evaluation Process of the First ELG Open Call

◾◾ ELG

European Language Grid
Results of the Survey on Evaluation Process

European Language Grid – Results of the Survey on Evaluation Process
## of the First ELG Open Call

The survey was prepared to get feedback from evaluators on the evaluation process of the first ELG open call. It was disseminated among all 48 evaluators after they have finished their evaluations of submitted project proposals. Within one week we have obtained responses from 40 respondents.

Survey Results

1. **How did you find out that ELG has published call for evaluators:**
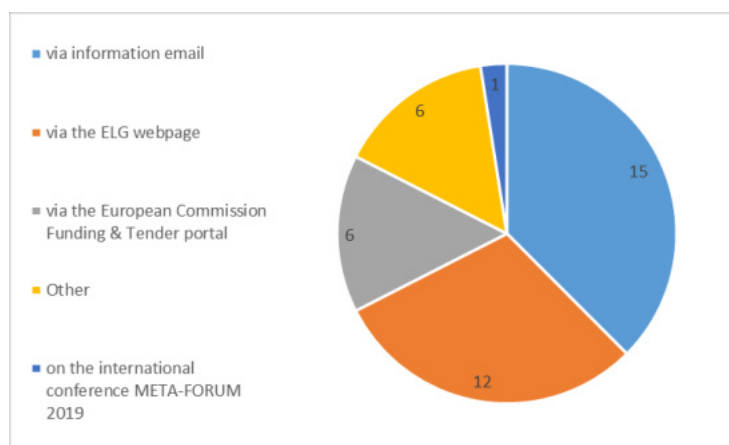
2. **Why are you interested in the evaluation of project proposals for the first ELG open call?**

ELG

1/7
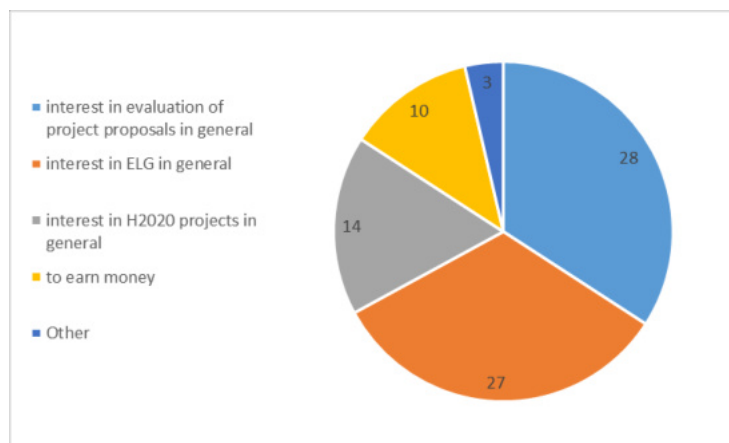
European Language Grid
Results of the Survey on Evaluation Process

3. **How do you assess the description of the evaluation process in the Call documentation (namely chapter 5 - Eligibility Criteria and chapter 7 - Summary of the evaluation process) and Annex Evaluation Criteria:**
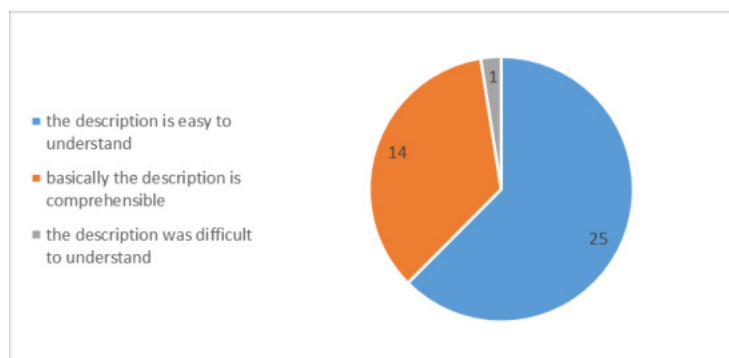
4. **How do you assess webinar for evaluators which took place on May 19, 2020?**
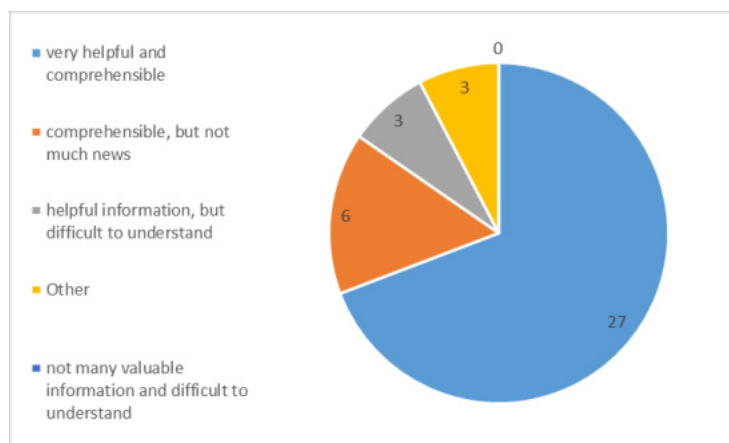
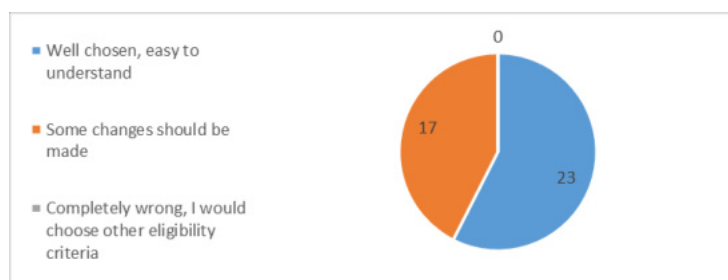5. **How do you assess the structure of the proposal template with regard to the evaluation process?**
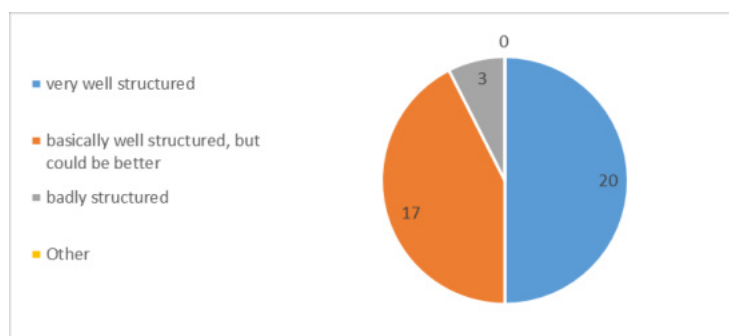
6. How do you assess structure of the evaluation form?

- very well structured, contains all information
- basically well structured, but could be improved
- Other
- badly structured

1 0
18 21
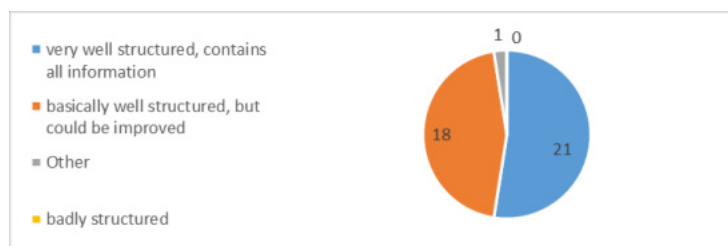
7. How do you find the eligibility criteria (uniqueness, relevenace for ELG, project phases)?

- Well chosen, easy to understand
- Some changes should be made
- Completely wrong, I would choose other eligibility criteria

0
17 23

8. How do you find the evaluation criteria (objective fit, technical approach, BID plan, budget adequacy, team)?

- Well chosen, easy to understand
- Some changes should be made
- Completely wrong, I would choose other evaluation criteria

0
14 26
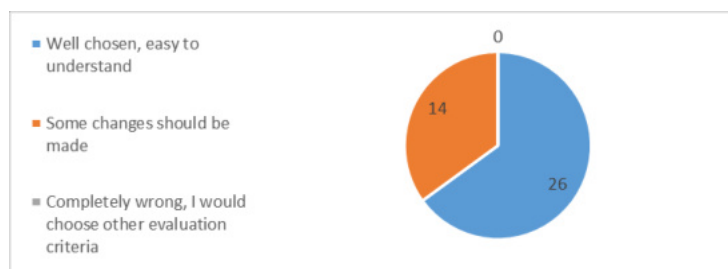
9. What is your experience navigating the ELG Grid with respect to the evaluation of project proposals?

- the grid content is very well structured and it is easy to find relevant information
- the grid content is well structured, but the documentation should be improved
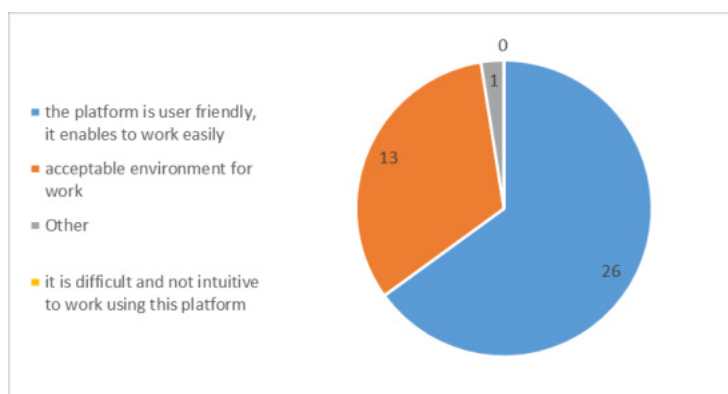- the orientation in the grid content is difficult
- Other

0
1
17 22

⣿⣿ ELG
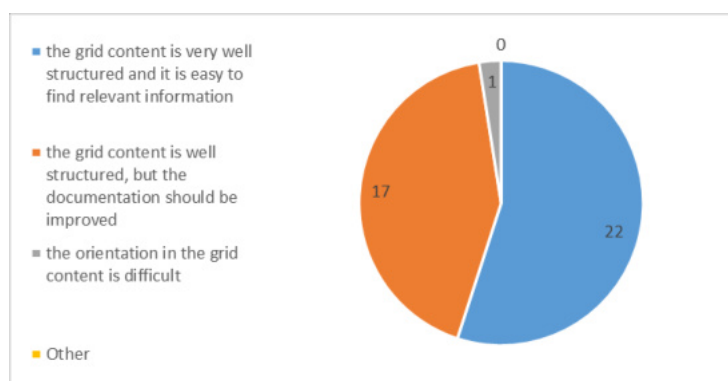
European Language Grid
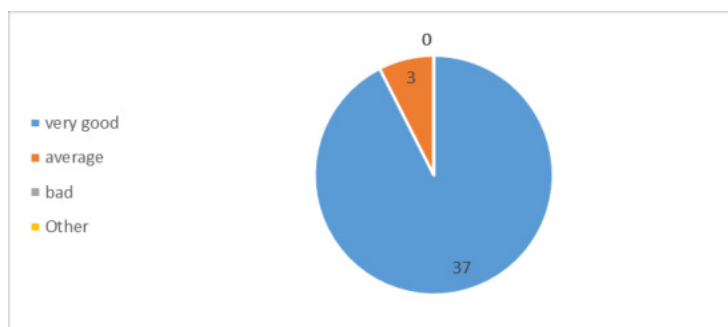Results of the Survey on Evaluation Process

10. How much time did you spend evaluating 1 project proposal (on average)?
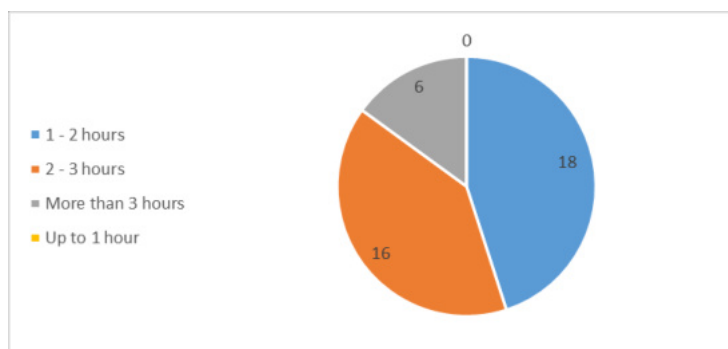
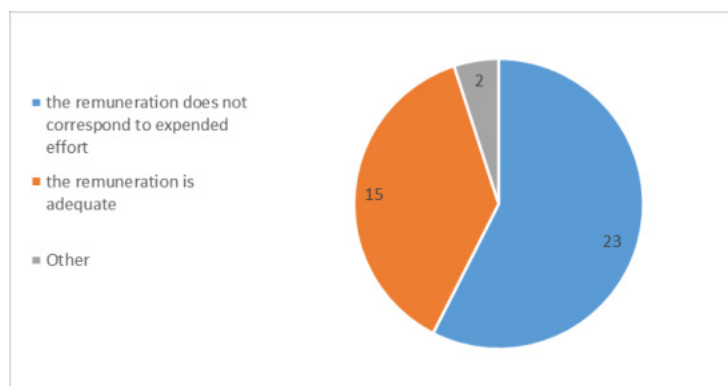**11. How do you assess ELG Open calls platform (registration process, evaluation submission,…)?**

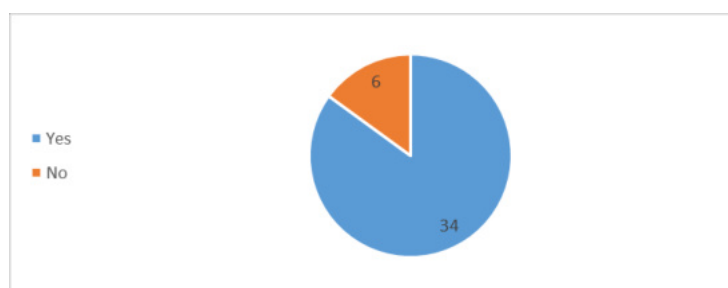**12. How would you assess support from ELG team (information emails, answering questions)?**

13. Are you satisfied with the remuneration obtained for evaluation?

- the remuneration does not correspond to expended effort
- the remuneration is adequate
- Other

14. Have you had experience with evaluation of project proposals before?

- Yes
- No

15. Are you willing to take part in evaluation of project proposals in second ELG Open Call for Pilot Projects?

- Yes
- Maybe
- No
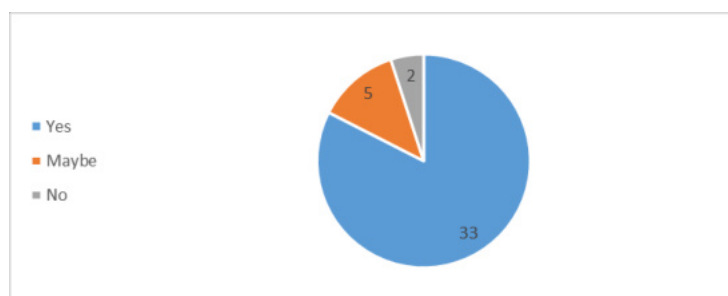
16. Is there anything you want to share with us or comment upon?

Respondents answers

One practical problem for the first round was browsing "what you already have" and the APIs and resources. That could be a section for the proposers to prove that this x is where I will integrate or this y is what is the "gap" in ELG resources that we will fill in. Well done and good luck with choosing the right projects!

In the evaluation process, put the eligibility criteria before the evaluation criteria.

Recommendation1: for the logic flow of the evaluation add the eligibility criteria prior to the evaluation as such.
Recommendation2: improve the search engine in the catalogue.

I suggest it might be better if there is an overall score for each proposal, in addition to summing up individual evaluation scores.

It took me quite some proposals to read before interpreting "integration" in a broad sense (technical/business). That, in turn, created some initial confusion as to what the phases should be. Future evaluators should receive an information pack with 1-2 anonymized evaluated proposals. The remuneration is, probably, 25-50% lower than what is should have been.

The project proposals are too short to properly evaluate the selected criteria. Participants didn't have enough space to describe them. For example, the experimental phase was one of the most difficult to assess. Additionally, the evaluators had to put quite a lot of effort into the review process to search the proposals for useful information to evaluate all the criteria.

"Uniqueness" is a very strong word for a hard requirement; I would like to see that a little flexibility in that phrasing.

My only suggestion concerns the limited information offered by proposals for orienting evaluations. Taking into account that some proposals are very ambitious and that the requested funds are significant would be very useful more detail into the proposals or in the attached documents. Several proposals lacks attachments with adequate explanations. Overall the criterion should be that every evaluator should be evaluate the proposals only ion the light of the information offered by the applicant, and this is often very difficult.

My general impression, the evaluation process has been well organised and thoroughly thought. The only required enhancement is clear specification during the call for evaluators in what way they will be contracted. This really matters. Being even formally "an employee" of Charles University for a few hours, can eliminate the expert for a next three years from proposals evaluation with participation of Charles University in other programmes due to conflict of interest (!). The process should be mutually beneficial. Regarding Open Call 1 proposals' quality, I have noticed that not all applicants paid enough attention to the budget structure required by the ELG consortium and this eliminates a few valuable proposals from recommending them to be selected in my case. Some of the applicants unfortunately ignored the necessity to divide costs between phases or did not provide direct costs figure. I would like to suggest the consortium to put budget structure to the eligible criteria section for the applicants. This will focus their attention on more carefully budget description. The goal of the ELG is to become a real LT marketplace, as I have understood, and valuable examples and technologies to be implemented are the key for project's further success. Another noticeable aspect is a dissemination plan. Some of applicants did forget to pay attention to collaboration with the ELG, which is crucial for the consortium. Thus, I recommend to highlight this in the relevant documents and Open Call 2 dissemination activities for the new applicants. The consolidation of Business, Integration and Dissemination Plan in one evaluation criterion does not look relevant. It sounds for me as an attempt of mirroring the three key phases: experiment, integration and dissemination. However, technical aspect or development should replace business in this case. If I am wrong with my assumption, integration phase, which is more technical should be eliminated from this criterion. I suggest to have a clear splitting of criteria between business and technical aspects. Moreover, the innovation can have not only technological nature, which is more relevant for the LT, but a business related too. Thus impact criterion, which is usually used in H2020 programme, together with dissemination activities and business model is more versatile and comprehensive. It would be nice to include it as a part of the evaluation process.

It would be helpful if projects were pre-assessed with regard to existing ELG entries that may be similar. I think this could be done automatically to generate a list of links to check against the proposal. It would make checking the uniqueness criteria much easier.

ılı ELG

European Language Grid
Results of the Survey on Evaluation Process

The different ELG environments are not clearly structured and it is often confusing navigating in them. The ELG platform should be much more publicised and its resources should be better organised and illustrated.

Cross-reading the other proposals at the end would be very interesting.

Very good overall. I think the project should follow same pattern that other programmes: Excellence (Technical), Impact (market economic and business plan), Implementation (project plan and budget). Everything in maximum 6 pages like DIH projects or other project of subgranting.

The level of the proposals is very high, so a finer grid for evaluation might be helpful for making a more informed decision.

It might be helpful to know what is the preferred reject rate :-)

Sometimes, I found it difficult not to give 10 points for one or another evaluation criterium even if globally the project was not very convincing. I wonder if this can be explained by the way these criteria are formulated. In addition, I would like to suggest that the evaluators have to propose a ranking of their projects after finishing all evaluations.

I had a hard time assessing the proposals regarding the usefulness of the ELG as a technology platform. Most projects seemed to exploit the ELG as a shop window, and I felt that more was needed. However, this criterion remained quite obscure to me. The integration part was also very sketchy in most proposals and I wonder whether it was so because the proposers had access to little details regarding the technical aspects of the ELG.

For questions 5 and 6, answered that this could be better, because I felt both the proposer template and the evaluation web-form had some overlap in terms of categories.

Legal note: The mode of remuneration as "employee" was not pre-announced, and may create difficulty for some evaluators, whose contract does not permit them to take on other employment, or requires a difficult/time intensive approval process in order to be granted permission to take on secondary employment.

General Point A list of reference documents would be helpful Evaluation Criteria - Objective fit: the supplied descriptions are essentially of the yes/no type (ie no guidance for evaluation) e.g. "will the project contribute services tools or data sets". - Tech approach: what is the motivation underlying the "innovative" criterion? These projects are *not* supposed to carry out research, right? Evaluator needs further guidance here. - Team: I found it very difficult not to award 10 points to most if not all projects because the teams seemed to have an appropriate distribution of skills. Again - what exactly are you looking for here? Perhaps this should be an eligibility criterion. SUMMARY: The yes/no question "Do you recommend....?" is too coarse. "Do you think this project should be considered for financing" would be better. Also, in this case, it is an assessment, not an evaluation - and as for the accompanying explanation, I opted to summarize strong and weak points. Perhaps this could be mentioned explicitly if you concider it appropriate. Eligibility Criteria - Relevance: this overlaps the Objective Fit criterion above. Either assimilate the two or sharpen the distinction between them Rating Scale: I found the 4 point scoring system is too coarse. EU evaluation allows half points - which gives 8 points and makes it much easier to rank proposals and apply discretion.

-

It was great to be part of ELG evaluators. This was my first opportunity and I am looking forward for more.

ELG

7/7

Annex 3: Guide for pilot projects

ELG – FSTP – Pilot Projects Open Call 1

Guide for pilot projects

This document contains (i) guide and basic information about the ELG pilot projects execution and (ii) guide to European Language Grid as such.

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.

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 this year's ELG conference, the (virtual) META-FORUM 2020 (date to be announced shortly). 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 2021, when the META-FORUM can hopefully be held again as a face-to-face event, pilot projects should show their results and demonstrate the usefulness of ELG. Part of the allocated budget for dissemination can be assigned to this dissemination task within these events.

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 next 18 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 measures 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 projects 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 pilot-projects@european-language-grid.eu).

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
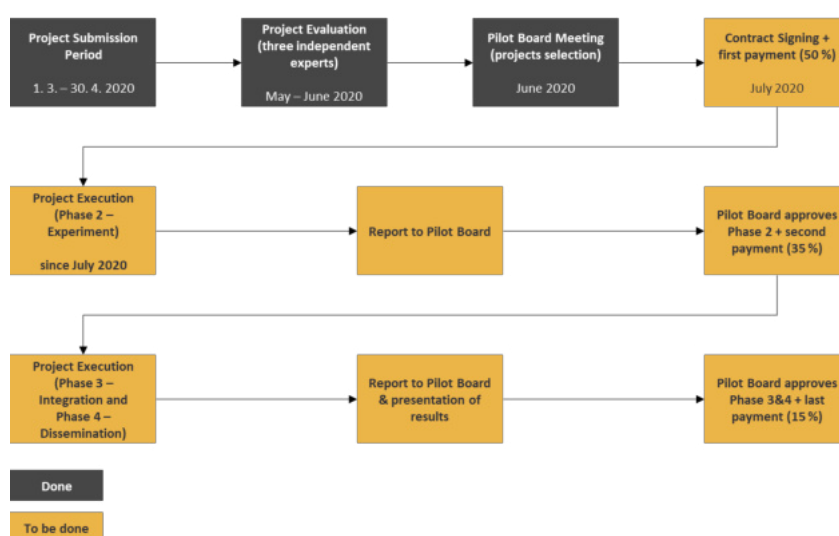
2. Guide to European Language Grid

The European Language Grid (ELG) platform 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** (https://european-language-grid.readthedocs.io/en/latest/) 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 - https://www.european-language-grid.eu/wp-content/uploads/2020/06/1stRegionalELGWorkshop_1_Overview.pdf)

- 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.

- How to test the ELG integrated services:

  https://european-language-grid.readthedocs.io/en/latest/all/TestService.html

  - How to download a data resource:

    https://european-language-grid.readthedocs.io/en/latest/all/Down.html

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 - https://www.european-language-grid.eu/wp-content/uploads/2020/06/1stRegionalELGWorkshop_4_ProvidingServicesTutorial.pdf)

⫲⫲⫲ ELG

European Language Grid
Open Call 1 – Guide for pilot projects

Integrating LT services into ELG

Step 1: Preparing the resource

- Introduction and main technical requirements:
  https://european-language-grid.readthedocs.io/en/latest/all/RegisterFunc.html#
- Technical documentation (specifications for the ELG integrated services APIs):
  https://european-language-grid.readthedocs.io/en/latest/all/LTPublicAPI.html
- Dockerization tips with help for Python and Java-based tools:
  https://european-language-grid.readthedocs.io/en/latest/all/Dockerization.html

Step 2: Describing and providing access to the resource

For release 1, providers were asked to upload XML metadata records compliant with the ELG metadata schema. A metadata editor form will soon be made available.

- Instructions and examples:
  https://european-language-grid.readthedocs.io/en/latest/all/RegisterFunc.html#describe-a-functional-lt-service
- Downloadable examples and ready-to-use templates:
  https://gitlab.com/european-language-grid/platform/ELG-SHARE-schema

Registering data resources (e.g. corpora, lexica, etc.) into ELG

Step 1: Preparing the resource

- Introduction and main technical requirements:
  https://european-language-grid.readthedocs.io/en/latest/all/RegisterNonFunc.html#

Step 2: Describing and providing access to the resource

For release 1, providers were asked to upload XML metadata records compliant with the ELG metadata schema. A metadata editor form will soon be made available.

- Instructions and examples for corpora (datasets):
  https://european-language-grid.readthedocs.io/en/latest/all/RegisterCorpus.html
- Instructions and examples for lexical/conceptual resources (e.g. lexica, ontologies, terminological glossaries, etc.):
  https://european-language-grid.readthedocs.io/en/latest/all/RegisterLexConc.html
- Instructions and examples for models and grammars:
  https://european-language-grid.readthedocs.io/en/latest/all/RegisterLangDesc.html#