EUROPEAN LANGUAGE GRID

D6.4

Evaluation of pilot project execution and results

Authors:	Lukáš Kačena (CUNI), Jana Hamrlová (CUNI), Jan Hajič (CUNI)
Dissemination Level:	Public
Date:	30-04-2022

About this document

Project	ELG – European Language Grid
Grant agreement no.	825627 – Horizon 2020, ICT 2018-2020 – Innovation Action
Coordinator	Prof. Dr. Georg Rehm (DFKI)
Start date, duration	01-01-2019, 42 months (GA amendment version: AMD-825627-7)
Deliverable number	D6.4
Deliverable title	Evaluation of pilot project execution and results
Type	Report
Number of pages	44
Status and version	Final – Version 1.0
Dissemination level	Public
Date of delivery	Contractual: 30-04-2022 – Actual: 30-04-2022
WP number and title	WP6: Grid Community – Piloting the European Language Grid
Task number and title	Task 6.4: Contracting, overseeing and evaluation of completed pilots
Authors	Lukáš Kačena (CUNI), Jana Hamrlová (CUNI), Jan Hajič (CUNI)
Reviewers	Katrin Marheinecke (DFKI), Andrejs Vasiļjevs (TILDE)
Consortium	Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Germany
	Institute for Language and Speech Processing (ILSP), Greece
	University of Sheffield (USFD), United Kingdom
	Charles University (CUNI), Czech Republic
	Evaluations and Language Resources Distribution Agency (ELDA), France
	Tilde SIA (TILDE), Latvia
	Hensoldt Analytics (HENS), Austria
	Expert System Iberia SL (EXPSYS), Spain
	University of Edinburgh (UEDIN), United Kingdom
EC project officers	Philippe Gelin, Miklos Druskoczi
For copies of reports and other ELG-related information, please contact:	DFKI GmbH European Language Grid (ELG) Alt-Moabit 91c D-10559 Berlin Germany Prof. 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 © 2022 ELG Consortium

Table of Contents

List of Figures

List of Tables

List of Abbreviations

API	Application Programming Interface
ELG	European Language Grid
FSTP	Financial Support to Third Parties
ICT	Information and Communication Technologies
LT	Language Technology
NCC	National Competence Centre
PB	Pilot Board
RO	Research organization
SME	Small and medium-sized enterprise

Abstract

We describe the two ELG open calls for pilot projects, the objective of which was to demonstrate the use and the advantages of ELG in providing basic language technologies (LT) for applications and as a basis for more advanced LT-based modules or components useful to industry. Our main goal was to attract SMEs and research organisations to either contribute additional tools or resources to the ELG platform (type A pilot projects) or develop applications using Language Technologies available in the ELG platform (type B pilot projects). We start with the detailed description of the submission and evaluation processes, followed by a presentation of the open call results. Afterwards, we describe the supervision and evaluation of the execution phase of the projects, as well as lessons learned, and feedback gathered from the pilot projects. Overall, we were very satisfied with the setup and with the results of the pilot projects, which demonstrated an enormous interest in ELG and the Language Technology topic in general.

1 Introduction

To demonstrate the advantages of ELG in providing LT for applications and as a basis for more advanced LT-based modules or components that are useful to industry, the ELG project set up a mechanism for using close to 30% of its budget for small-scale demonstrator projects ("pilots") through two open calls. The calls were prepared using the ICT-29a) call specification, making use of the Financial Support to Third Parties (FSTP) scheme according to the ICT Work Programme 2018-2020. In total, we provided 1,950,000€ to the selected projects as FSTP with an awarded amount of up to 200,000€ per project. We established a lightweight submission procedure and a transparent evaluation process, in which external evaluators participated as reviewers.

The main objective of the open calls was to attract SMEs and research organisations to either contribute tools and services to ELG (type A projects) or develop applications using Language Technologies available in the ELG platform (type B projects). The results of the pilot projects are included in the ELG platform for dissemination, testing and external evaluation by other entities or the public.

This deliverable sums up the open calls for pilot projects, its execution, and results. It is a follow up of previous deliverables – D6.1 Pilot Calls Setup, D6.2 Call 1 results and description of selected pilot projects and D6.3 Call 2 results and description of selected pilot projects.

2 Organization of the Open Calls

In this chapter, management structure, timeline, submission, and evaluation processes are described.

2.1 Management Structure and Organization

While agile, simple, and lightweight from the proposers' point of view, the organisation of the two open calls was an internally complex procedure requiring close collaboration of three different teams (Pilot Board, management team, technical team,) with support from a broad panel of external evaluators.

2.1.1 Pilot Board

The Pilot Board (PB) was set up for the supervision of the pilot projects. While the management team took care of the organisation and handling of the open calls and the execution of the pilots, the PB provided a forum so that the ELG project could discuss the progress of the pilots, their feedback and results. The PB was meant to be the main technical and strategic interface between the pilot projects and the ELG project proper, so that ELG could maximise its benefits from supporting the pilots and to make sure that the pilot projects benefit from ELG.

The PB operational procedures were drafted by the management team and approved by the ELG Steering Committee. Afterwards, seven PB members were nominated and approved. The operational procedures defined the main responsibilities of the PB as follows: approval of the open calls and related documentation; pilot project selection process; supervision of pilot project execution, including progress monitoring, evaluation of results and approval of the phased payments.

2.1.2 External Evaluators

An independent panel of experienced external evaluators ensured an open, transparent and expert-evaluation based selection process. The pool of evaluators was created using a separate open call. The evaluators were responsible for evaluating the project proposals and worked remotely using the web interface of the ELG Open Calls Platform. They were selected from the pool, avoiding any conflicts of interest. All evaluators were asked to sign a non-conflict of interest declaration and a confidentiality agreement before being accepted to perform the task.

2.1.3 Management Team

The management team organised the whole open calls process, including managing and directing the technical team. In line with Annex K of the Work Programme and other relevant sections of the Rules for Participation, the management team prepared all prerequisites and procedures: the Open Calls Platform, web content, informational materials, forms, contract templates, presentation and reporting forms and templates, submission procedure, hiring and selection of external evaluators, call management structure, internal auditing, and project results evaluation procedures. In the initial setup phase, the management team tapped the legal and finan-

cial expertise of the Technology Centre of the Czech Academy of Sciences, which is charged by the Czech government to host the National Contact Point (NCP) and other experts related to the preparation, execution and evaluation of EU framework programmes and projects.

2.1.4 Open Calls Platform Technical Team

An essential task was to set up the ELG Open Calls Platform for the proposal submission, evaluation and reporting process. We decided to develop the platform in-house to ensure that it fit our needs.[1] The Open Calls Platform technical team was responsible for developing the platform and for support during each phase of the process.

2.2 Timeline

Figure 1 shows the open calls execution timeline. After the announcement, each call was open for submissions for two months, followed by an evaluation procedure of approx. two months. After signing the contracts with the selected projects, the execution phase started. The expected project duration was 9-12 months. Four projects asked for a short extension of one or two months (which was accepted), mainly due to Covid-19 related delays of dissemination activities.
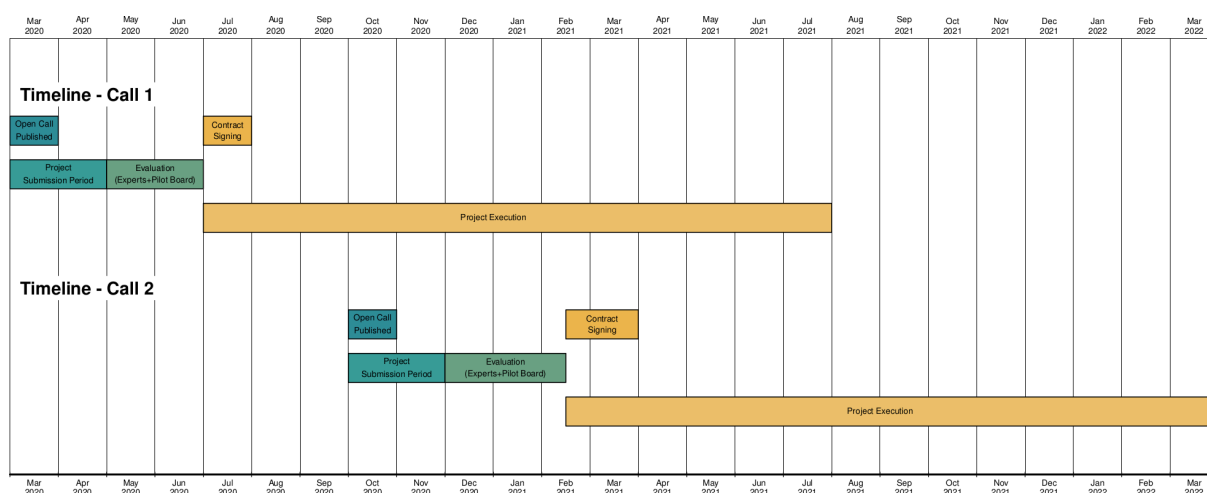

Figure 1: Open calls timeline

2.3 Communication with Stakeholders

Prospective applicants were targeted through various channels, e. g., the open calls website, a survey for stakeholders and other communication and dissemination activities carried out by all ELG consortium members. From early 2019 onwards, the open calls were presented on the ELG website.[2] The content of this landing page was regularly updated, starting from basic information including the timeline and key parameters at the beginning of the project, followed by the call for evaluators[3] and complete information regarding the open calls (until the publication of the selected projects).[4]

We first monitored the interest in the open calls using a survey, which ran from May 2019 until June 2019. A total of 108 respondents participated. The result showed significant interest in the open calls and also a high

---

[1] https://opencalls.european-language-grid.eu
[2] https://www.european-language-grid.eu/open-calls
[3] https://www.european-language-grid.eu/open-calls/call-for-evaluators
[4] https://www.european-language-grid.eu/open-calls/call-for-pilot-projects

demand for more information. Five months before the first call announcement, a second survey was prepared. We disseminated this survey during the first annual ELG conference META-FORUM 2019 in October 2019 in Brussels and collected answers from 47 respondents, 84% of which expressed an interest in taking part in the open calls.

The open calls were promoted through social media (Twitter, LinkedIn), various e-mail distribution lists, internal networks, and collaborators, through the META-FORUM conference and through other means whenever an opportunity arose (e.g., NCC workshops and other conferences).

2.4   Submission Process

As explained in the previous section, in the preparatory period the overall open call procedure was set up, including all related documents and the development of the online platform for the management and evaluation of submissions. After the official announcement of one of our two open calls, applicants could then prepare and submit their project proposals. There was a continuous need for support, mainly answering questions we received by the participants via email.

With regard to the call announcement, we paid special attention to a well-prepared call documentation, which provided all necessary information for applicants, and a user-friendly submission platform. The documentation was prepared as an easy-to-understand document. It contained several annexes: Guide for Applicants, Third Party Agreement, Project Proposal Template and Evaluation Criteria.

In the Guide for Applicants the management team showed, using screenshots, how to submit a project proposal through the platform, i.e., how to create an applicant account, how to log in and manage the account, how to create a new project proposal, fill in the forms and finally submit the proposal. We also maintained a list of (expected) frequently asked questions, for example "Who can apply for a pilot project?", "How much money is allocated for the pilot projects?", and "Does Brexit have any implications on eligibility?".

The Open Calls Platform was developed using the Open-Source Content Management System Drupal with the guiding principle to keep the submission and evaluation process easy and straightforward for the participants and manageable for the call organisers. The platform runs under the ELG domain[5], while physically residing with the technical team to ensure quick reactions to any technical problems.

2.5   Evaluation Process

Preparation and execution of the evaluation process are described in this chapter.

2.5.1   Preparation of the Evaluation Process

The most important part of the preparation of the evaluation process was the selection and specification of evaluation criteria that match the objectives to be achieved by the calls. At the same time, the criteria ought to be clear for the external experts evaluating each proposal.

The criteria were defined and described in detail in the call documentation. First, the submitted proposal should fulfil formal requirements (language, submission date, declaration of honour, legal status, eligible coun-

---

[5] https://opencalls.european-language-grid.eu

try, number of submitted proposals per applicant and no conflict of interest) which were checked by the management team before any further evaluation. Then, three independent evaluators checked the binary eligibility criteria: uniqueness, relevance for ELG, and whether the proposal contains all the required phases (experiment, integration, dissemination). These were followed by the graded and ranked evaluation criteria: objective fit, technical approach, business, integration and dissemination plan, budget adequacy, and team.

In order to identify evaluators with experience in language technologies and evaluation, a call for evaluators was published in February 2020. All relevant information (description of tasks, eligibility of candidates, selection criteria, contact email for questions, and a link to the registration form on the Open Call platform) was published on the ELG website as well as on the European Commission Funding and Tender portal. In addition, ELG disseminated the call through various channels. Potential evaluators were asked to fill in a registration form, through which contact information, CV, and professional experience related to evaluation and LT were collected. From about 156 applications, the management team selected 64 evaluators (a total for both project open calls) with relevant expertise in both the subject field(s) and in evaluating projects of at least similar size.

Before assigning projects to evaluators, we sent them an email with instructions regarding the evaluation. We also organised webinars for evaluators in which the evaluation process and evaluation criteria were explained.

All evaluators signed a contract with the ELG project, which included a clause to keep any technical or business information about the evaluated projects in strict confidence, as well as a no-conflict-of-interest declaration.

2.5.2 Execution of the Evaluation Process

Each submitted project proposal was evaluated by three independent external evaluators to ensure an open, transparent and expert evaluation-based selection process. The evaluators were carefully assigned to the proposals by the management team to assure their experience and background balance. We also paid attention to gender (at least one female evaluator per proposal) and country of residence of the evaluator, avoiding at the same time possible personal or nationality-based conflicts of interest. The whole process was monitored by the Pilot Board. Each project proposal was assigned to one of the PB members. These project coaches checked and confirmed or rejected the selection of evaluators with special regard to conflict of interest.

After the evaluation, the project coaches prepared summary reports for each project proposal assigned to them and submitted these to the Open Calls Platform. In these summaries, the project coaches first reviewed the three evaluation reports submitted by the external evaluators. They also suggested potential budget adjustments and changes of the total number of points (the maximum was 300 points, i.e., 100 points from each external evaluator) assigned to the proposal in range of at most 30 points (open call 1) or 45 points (open call 2) up or down, where applicable. According to the evaluation criteria, project proposals by SMEs developing applications using LT that is already available in the ELG (B type projects) received 30 bonus points. Finally, the project coaches reviewed the eligibility criteria (uniqueness, relevance for ELG and project phases) as checked by the evaluators and suggested their decision on their fulfilment if the evaluators differed in their opinions. The project coaches also assessed the performance of the evaluators and the quality of the reports.

After all summary evaluation reports had been submitted by the project coaches, a Pilot Board meeting was convened, in which the final ranking and selection was decided. All proposals were ranked by the total sum of assigned points. The ranked list was cut at the maximum available financial support (1,365,000€ for open call 1 and 585,000€ for open call 2).

3 Results of the Open Calls

There were two open calls, first one opened in March 2020 and second in October 2020. In this chapter, we describe the results of the open calls, list the selected projects, and also mention the changes made between the first and second open call.

3.1 Open Call 1

3.1.1 Overview

The first call was opened on 1 March 2020 and closed on 30 April 2020 in accordance with the timeline (Figure 1). We accepted a total of 110 project proposals for evaluation from 103 applicants.

Submitted by	Type A	Type B	Total
Research organization	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 1: Proposals submitted for the first open call and accepted for evaluation

Seven applicants (five SMEs and two 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 ELG) and 31 proposals were of type B (develop applications using language resources and technologies available in ELG), see Table 1. We received proposals from 29 different 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 was 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€.

At the end of June 2020, the results of the first open call were announced on the ELG website, including the list of projects selected for funding.[6] The two projects from the reserve list were informed that they might be selected for financial support if any of the selected projects rejected the financial support. The remaining projects were informed that they were not selected. In July 2020, contracts with all selected projects were signed, and the first payments were made (half of the awarded financial support), in line with the approved call documentation and procedures. All projects had started their execution phase by August 6. Furthermore, at the end of July 2020, abridged versions of the summary evaluation reports were provided to all applicants through the Open Calls Platform.

3.1.2 Selected Projects

The projects selected in open call 1 are listed in Table 2. All supported organisations are from the EU – three from Finland, two from Austria, Germany and Italy, and one from Spain. The awarded budget varies from 87,445€ to 167,375€.

---

[6] https://www.european-language-grid.eu/open-calls/open-call-1

Organization	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 2: List of pilot projects selected for financial support in the first open call

Although we obtained more proposals from SMEs than from research organisations, there are three SMEs and seven research organisations among the selected projects. Similarly, although B type projects from SMEs were preferred, only two B type projects were accepted for financing which probably reflected the fact that the ELG platform was still being developed at the time of the first open call. Thus, it appeared to make more sense to create missing resources or tools rather than build applications using resources and tools available in ELG.

Four of the eight A type projects aimed to enrich the ELG platform with language resources and six of them planned to provide various language tools (i. e., two of the projects provide both resources and tools). The two B type projects promised speech applications – a smart speaker and a digital twin based on real-time language

translation and analysis. The projects in general often dealt with underrepresented languages such as Basque, the Nordic languages, and European minority languages.

Technologically, the projects targeted a diverse set of goals and areas. There are projects targeting important interdisciplinary areas (medical informatics, manufacturing), modern technologies relating to language and semantic as well as world knowledge (Linked Open Data, paraphrasing) and core scalable technologies (distributable containers). Evaluation platforms as well as advanced and scalable ma chine 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.

3.1.3 Feedback provided and Survey for Proposers

With the goal of evaluating and improving our open call procedure, we conducted several surveys with every-one involved in the first open call. We started with the project proposers. After the evaluation process we also conducted a survey among all evaluators. The last survey was conducted among the Pilot Board members.

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 pro-posals", and "your organisation". The information was collected anonymously.

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. The main conclusions from the proposers' survey that were relevant for the setup of the second open call: Almost 70% of respondents were interested in ELG because of both (functional) services and datasets. Slightly more than two thirds of the respondents preferred smaller, agile calls over large, consortium-based calls.

There was 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 ELG API integration and about the infrastructure for working with data, applications and possibly also workflows were requested. Some improvements of the Open Calls Platform and its user-friendliness were made (e. g., limited space).

3.2 Open Call 2

The second open call was launched in October 2020 and experience from the first open call was reflected.

3.2.1 Changes made between Open Call 1 and Open Call 2

The basic parameters, specified in the ELG Grant Agreement, remained the same for the second open call. Based on the lessons learned from open call 1, we implemented the following changes in the call documenta-tion and the open call procedure:

- We improved the explanation of the strategic goals of ELG and the goals of the open calls. Links to an overview of ELG, its history and context and to an overview of the ELG project and platform were pro-vided in the call documentation.
- We also improved the technical documentation of the ELG infrastructure and provided an easy-to-find list of currently available services – this was done with the launch of ELG Release 1 (June 2020).
- We organized a webinar, which took place during the submission period, on 12 November 2020. We explained the goals of the open call and presented the call documentation. The second part of the webinar was dedicated to questions and a discussion. A recording of the meeting was made available to all applicants.
- The documentation, annexes, templates, and forms along with the Open Calls Platform were further improved. In the proposal template, budget breakdowns were requested in a fixed structure as well as a more detailed budget justification.
- New evaluators were recruited and added to the current group, with the aim to attract more experi-enced evaluators.
- It was decided that the second open call, like the first open call, should have no specific thematic focus.

3.2.2 Overview

The second call was opened on 1 October 2020 and closed on 30 November 2020. We accepted 103 project proposals in total for evaluation.

Submitted by	Type A	Type B	Total
Research organization	38 (36.9%)	5 (4.8%)	43 (41.7%)
SME	28 (27.2%)	32 (31.1%)	60 (58.3%)
Total	66 (64.1%)	37 (35.9%)	103 (100%)

Table 3: Proposals submitted for the second open call and accepted for evaluation

Five applicants (four SMEs and one research organisation) submitted two proposals (one type A and one type B). Regarding the project type, 66 proposals were of type A, and 37 project proposals were of type B. A total of 43 applicants who submitted a proposal in the second open call indicated that they had submitted the same or a similar proposal in the first open call. We received applications from 28 different countries, including eligible countries outside the EU (Iran, Israel, Norway, Serbia, Switzerland, Turkey, United Kingdom). The total amount of financing requested by the submitted projects was 13,257,919€. The average amount requested per project was 129,000€, which is less than in the first open call (153,000€).

In February 2021, the results of the second open call were announced on the ELG website.[7] All applicants were informed about the results. In February and March 2021, contracts with all selected projects were signed, and the first payments were made (half of the awarded financial support), in line with the call documentation and procedures. All projects had started their execution phase by 1 April 2021. Furthermore, in March 2021, abridged versions of the summary evaluation reports were made available to all applicants through the Open Calls Platform.

3.2.3 Selected Projects

The projects selected for financial support in open call 2 are listed in Table 4. The supported organisations are from five EU countries and the awarded budget varies between 85,421€ and 137,227€.

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

Table 4: List of pilot projects selected for financial support in the second open call

---

[7] https://www.european-language-grid.eu/open-calls/open-call-2

Although we obtained more project proposals from SMEs than from research organisations, there are two SMEs and three research organisations among the selected projects. Similarly, only two B type projects were accepted for financing.

Three A type projects aimed at providing tools to enrich the ELG platform. One project contributed multilingual annotated data, tools, and services for image processing whilst the second one aimed at improving the ELG offer of linguistic tools by proposing a unified service powered by state-of-the-art neural models for carrying out annotations on three Natural Language Understanding tasks, i. e., Word Sense Disambiguation, Semantic Role Labelling and Semantic Parsing, in around 100 languages. The third A type project expanded the portfolio of language resources available in ELG by adding a dataset and search tool for Czech sign language. Regarding the B type projects, one of the projects also dealt with sign language. Its goal was to simplify text comprehension for deaf people by linking words and phrases to a sign language encyclopaedia. The other project aimed to develop a set of tools, datasets, and services to enable automatic classification of the reading difficulty of texts on the Common European Framework of Reference.

3.2.4 Survey for Proposers to the Open Call 2

Just like for the first open call, a survey with 15 questions was designed for those who submitted a proposal. The survey had three sections: "motivation", "project proposals", "your organisation". In total, 39 out of 103 proposers (38 %) responded.

Regarding the motivation to submit a proposal, contributing services or resources to ELG to make them available to the ELG community and further development of an existing software or data project were the most frequent reasons reported by the respondents. The main expectations toward ELG were that the platform increases the visibility of the applicant's organisation on the European level and to get access to a large repository of tools and datasets. Also, almost all respondents think that more EU-funded activities dedicated to Language Technology and Language centric AI are needed, preferably in the form of agile calls (with short proposals and quick evaluations, 9-12 months project run-time). Regarding the specialisation of respondents, most frequently they specialised in text analytics, machine translation or speech recognition. Respondents reported more than twenty domains that they specialise in (most frequently health sector), one fourth of all respondents have no particular specialisation.

4   Project Execution

In this chapter, the project management and mid-term review process are described, as well as final review process and project results. Also, key dissemination activities are listed.

4.1   Project Management and Mid-Term Review

Once the pilot projects were selected and the contracts signed, the continuous support from the ELG consortium started so that the projects could start their execution. The first opportunity where the newly selected pilot projects could become more familiar with ELG were the online meetings with the Pilot Board and other members of the ELG consortium. During these meetings, basic information about ELG and its technology as well as guidelines for project execution were presented.

The project execution (Figure 2) consisted of four phases: Phase 1 – Open Call; Phase 2 – Experiment; Phase 3 – Integration; Phase 4 – Dissemination. After finishing Phase 2, reporting from the applicants was required, and then the Pilot Board decided whether the project is allowed to continue execution (and consequently, whether the next payment, 35% of the awarded financial support, is made). All projects successfully passed this review and were approved to advance to Phase 3 Integration and Phase 4 Dissemination. After finishing Phase 4, a final report was required, and the Pilot Board evaluated the whole project and decided whether the project receives the final payment (15% of the awarded financial support).

As mentioned, each project was supervised by a project coach who was responsible for training the project team, collecting and answering questions during project execution, collecting reports, and guiding the team through the project phases.
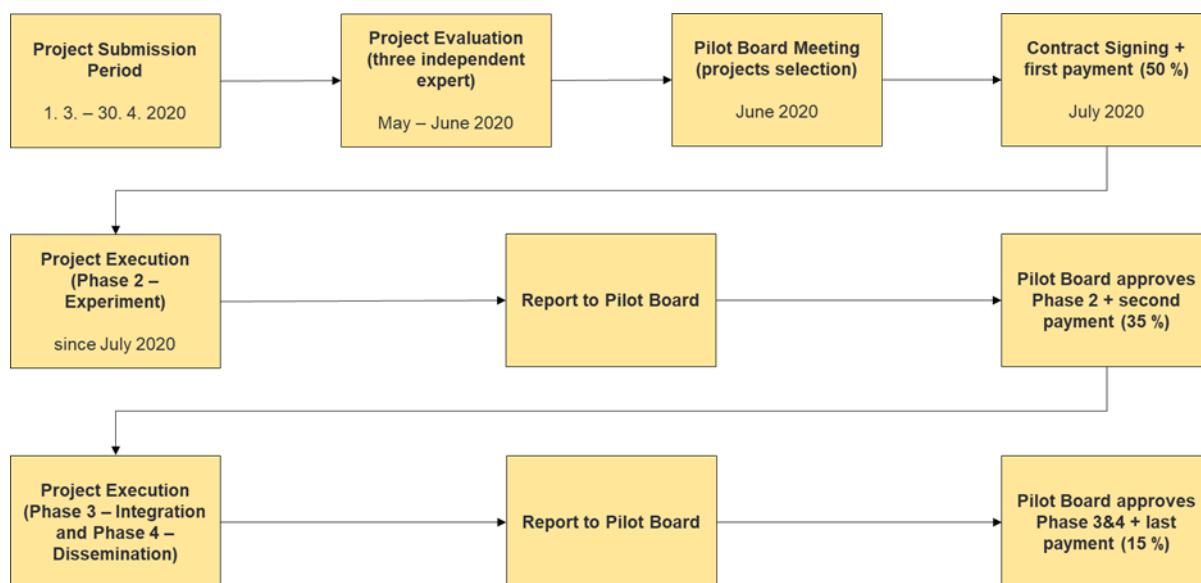

Figure 2: Project execution scheme (dates valid for first open call)

4.2   Final Review Process and Project Results

The final review was based on a very similar procedure to the mid-term reviews:

Step 1: Self-assessment by project team

Step 2: Evaluation by project coach

Step 3: Pilot board approval

As for the mid-term review, the evaluation was done via ELG Open Calls platform. The final step of the final review was a Pilot Board meeting. At this meeting, the project teams were asked to present their project and project results, business plan and future plans. The presentation of the project was based on a video (3-5 minutes long) that was published on the ELG website[8] and during the project expo at META-FORUM 2021 and 2022.

---

[8] https://www.european-language-grid.eu/open-calls/open-call-1

Rather than doing these meetings one by one, we decided to group the pilot projects. The final review meetings took place on-line on 22 September 2021 and 2 November 2021 for pilot projects from the first open call. The final review meeting for the pilot projects selected in the second open call took place on 8 April 2022.

At final the review meetings, the Pilot Board came to the conclusion that all 15 projects achieved their goals and outcomes. More details are described in Table 4, including the assessment of the project execution and results done by the Pilot Board at their final review meeting (with ratings scale ranging from 1 – highly satisfied to 4 – highly dissatisfied). There were 13 pilot projects with which the Pilot Board was '*highly satisfied*' and only 2 projects rated as '*satisfied*'.

Organization	Project Name	Project Goals	Project Outcomes	Goals and Outcomes Achieved	Pilot Board's assessment
Fondazione Bruno Kessler	European Clinical Case Corpus	The primary goal of the project was to build a fully manually annotated paraphrase corpus for Finnish, accompanied by a small-scale Swedish test set. The goals were to (a) build this clearly much-needed resource (b) gather experience and data regarding how such a resource can be built efficiently and what human resources are needed (c) build initial models based on the new resource and produce baseline results.	1) Dataset 2) Derived data and models 3) Annotation software	**Yes**	1 - highly satisfied
Lingsoft, Inc.	Lingsoft Solutions as Distributable Containers	The goal of the project was to provide Lingsoft's existing speech and language technology tools as ELG compatible services. This included four types of main services: speech recognition, machine translation, proofing, and text analysis (lemmatization, morphology analysis and NER).	35 services (one service per language and functionality)	**Yes**	1 - highly satisfied
Coreon GmbH	MKS as Linguistic Linked Open Data	Two major goals were achieved: 1) External solutions can query Coreon's MKS no longer only with the proprietary API but via SPARQL, the established method of the Semantic Web, in an easy way. 2) Understand how to overcome limits of knowledge graph editors, rather complement these data models with linguistic information (terms and all the rich descriptors accompanying terms).	1) Coreon SPARQL endpoint: Eurovoc combi 2) Coreon SPARQL query tool	**Yes**	1 - highly satisfied
Elhuyar Fundazioa	Basque-speaking smart speaker based on Mycroft AI	The main goal of the Smart euSpeaker project was to develop an open-source smart speaker that would work in Basque language, that would be open source, and it would make use of Elhuyar's Basque speech synthesis and recognition technology.	Working Mycroft smart speaker with its basic skills translated to Basque and some new skills too	**Yes**	1 - highly satisfied

Organization	Project Name	Project Goals	Project Outcomes	Goals and Outcomes Achieved	Pilot Board's assessment
Universita' Degli Studi di Torino	Italian EVALITA Benchmark Linguistic Resources, NLP Services and Tools for the ELG Platform	The goal of EVALITA4ELG is to enable the ELG users to access the resources and models for the Italian language produced over the years in the context of the EVALITA evaluation campaign. The 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.	A catalogue of resources and models developed during the various editions of the EVALITA campaign, designed in the form of a knowledge graph that can be inspected through SPARQL queries.	**Yes**	1 - highly satisfied
University of Helsinki	Open Translation Models, Tools and Services	The goals of OPUS-MT project were to develop state-of-the-art neural machine translation models that can freely be shared and integrated in open web services and professional translation workflows. The focus was set on European languages emphasising especially minority languages and multilingual translation models. The goal was to deliver easily deployable translation services in terms of dockerised MT solutions and CAT plugins.	Neural machine translation models	**Yes**	1 - highly satisfied
Centre for Translation Studies, University of Vienna	Extracting Terminological Concept Systems from Natural Language Text	The goal of the pilot project Extracting Terminological Concept Systems from Natural Language Text (Text2TCS) was to develop a language technology that extracts terms grouped by synonymy and relations from natural language texts in multiple languages.	The final Text2TCS LT supports up to 100 languages	**Yes**	1 - highly satisfied
University of Turku, Turku NLP research group	Textual paraphrase dataset for deep language modelling	To build a fully manually annotated paraphrase corpus for Finnish, accompanied by a small-scale Swedish test set. The goals were to (a) build this clearly much-needed resource (b) gather experience and data regarding how such a resource can be built efficiently (c) build initial models based on the new resource and produce baseline results.	1) Textual paraphrase dataset 2) Derived data and models 3) Annotation software	**Yes**	1 - highly satisfied

Organization	Project Name	Project Goals	Project Outcomes	Goals and Outcomes Achieved	Pilot Board's assessment
Weber Consulting KG	Virtual Personal Assistant Prototype	The goal of the project was to build a prototype for a personal virtual assistant (called YouTwindi), which can be installed on a small hardware device or integrated within the ELG container.	Virtual Personal Assistant Prototype (EDDI - Open Source Chatbot Platform)	**Yes**	2 - satisfied
FZI Research Centre for Information Technology	Streaming Language Processing in Manufacturing	The overall goal of the SLAPMAN project is the development and integration of streaming language technology modules to process, analyse and exploit non-structured or semi-structured data and the integration of these modules into the open source IoT toolbox Apache StreamPipes.	1) Various core extensions to the foundational Apache StreamPipes tool 2) StreamPipes extensions with ELG/LT components and the corresponding custom StreamPipes helm chart	**Yes**	1 - highly satisfied
Institute for Bulgarian Language	Multilingual Image Corpus 2021	Main goal was to develop an image dataset with the following main features: (a) a large image collection containing thousands of images and annotations; b) annotation classes belonging to a specially designed Ontology of visual objects; c) automatic object segmentation and classification evaluated by experts; d) the object classes and attaching definitions of concepts translated in at least 20 languages.	Multilingual Image Corpus	**Yes**	1 - highly satisfied
EDIA BV	CEFR Labelling and Assessment Services	The project goals were to develop a set of data collection and labelling tools to facilitate the creation of data sets (corpora). These tools can be used to develop text analysis and classification services to automatically assess a text's reading difficulty against the Common European Framework of Reference (CEFR).	1) Datasets 2) Services	**Yes**	2 - satisfied

Organization	Project Name	Project Goals	Project Outcomes	Goals and Outcomes Achieved	Pilot Board's assessment
University of West Bohemia	Motion-Capture 3D Sign Language Resources	The project aimed to acquire a motion capture dataset of a sign language that covers a wide domain, grammatical context, and multiple signers. The project goal was to deliver: (i) Motion capture dataset of Czech sign language and (ii) Tools enabling search for single glosses, phrases, or small motion sub-units.	Motion capture dataset for Czech sign language Searching tool	**Yes**	1 - highly satisfied
Sapienza University of Rome	Universal Semantic Annotator: A Unified API for Multilingual WSD, SRL and AMR annotations	The objective of the Universal Semantic Annotator (USeA) project was the creation of the first unified API for three primary tasks in Natural Language Understanding (NLU): i) Word Sense Disambiguation (WSD) – the task of assigning the most appropriate sense to a word in context, ii) Semantic Role Labelling (SRL) – the task of extracting the predicate-argument structures within a sentence, and iii) Semantic Parsing (Abstract Meaning Representation, AMR) – the task of representing a text in a structured semantic graph.	Universal Semantic Annotator	**Yes**	1 - highly satisfied
Sign Time GmbH	Sign language explanations for terms in a text	The project SignLookUp serves the goal to make texts easier to comprehend for deaf people. SignLookUp is a technology that links texts to a sign language encyclopaedia. Deaf people can then click on difficult or unknown terms in a text and immediately receive the explanation or description of the word in their sign language which is displayed adjacent to the text.	Deaf Reading Assistant	**Yes**	1 - highly satisfied

Table 5: Pilot project goals and outcomes

4.3   Dissemination Activities

To advertise them to a wider public, the pilot projects were presented at two annual ELG conferences, i.e., META-FORUM 2020 and META-FORUM 2021, in dedicated pilot project sessions in which all projects could present their main approaches and goals. Similarly, the pilot projects will be presented at META-FORUM 2022. In addition, workshops and training events organised by the ELG National Competence Centres (NCCs) were also used as opportunities to present certain pilot projects in the respective countries and regions.

In addition to that, all pilot projects had their own dissemination activities. Examples and key dissemination activities are listed in Table 6.

Organization	Project Name	Key Dissemination Activities
Fondazione Bruno Kessler	European Clinical Case Corpus	• E3C has been presented at CLiC-It 2020, SEPLN 2021, and at META-FORUM 2020 (2020/12/1-3). • A meeting with experts for collecting feedback was held (2021/01/25) and a website created (e3c.fbk.eu).
Lingsoft, Inc.	Lingsoft Solutions as Distributable Containers	• Presentation at the META-FORUM 2020 • ELG arranged ELRC workshop in Finland • An invited talk for the Icelandic Language Technology Consortium containing both NLP developers/researchers and NLP user organisations (e.g. media or public service bodies)
Coreon GmbH	MKS as Linguistic Linked Open Data	• ISO TC37/SC3/WG3 – Draft Technical Recommendation: "RDF representation for TBX Core" • Two longer posts on https://blog.coreon.com, several mentions on LinkedIn, Twitter • Presentations and papers on: DTT April 2021; LTI value cast April 2021; LDK, September 2021; Tekom November 2021
Elhuyar Fundazioa	Basque-speaking smart speaker based on Mycroft AI	• Press release halfway of the project (It had a great impact: many media wrote articles about it or contacted Elhuyar Fundazioa for interviews • A website to explain the project and to communicate via a blog of the advances done in the development (https://mycroft.eus/)
Universita' Degli Studi di Torino	Italian EVALITA Benchmark Linguistic Resources, NLP Services and Tools for the ELG Platform	• Final event (23-24 September 2021) with hybrid participation, including an overview of the project and the results obtained, a tutorial about integrating systems and resources on ELG, and a round table/decomposed panel with 14 invited speakers chosen among the most active organizers of tasks of EVALITA. The tutorial is available on the event web page for reaching a wider audience. • A project paper has been published on the IJCOL journal (open access)

Organization	Project Name	Key Dissemination Activities
University of Helsinki	Open Translation Models, Tools and Services	• Presentation & booth at META-Forum (2020 and 2021) • Invited talk at KITES symposium 2021 • Invited talk at the Nordic Meeting of the Language Council (2021) • Huawei NLP workshop - Towards Reliable Cross Lingual Machine Translation and Users' Intent Understanding • Webinar on accessibility organised by the Goethe Institute (2021) • Plenary talk at the XVIII Symposium on Translation and Interpreting Studies (2021) • Article in the IEEE Spectrum • Article on OPUS-CAT in The Cronical • Various publications and tweets on OPUS-MT, OPUS-CAT and the Tatoeba-MT challenge (WMT, EACL, NoDaLiDa)
Centre for Translation Studies, University of Vienna	Extracting Terminological Concept Systems from Natural Language Text	• List of publications: https://text2tcs.univie.ac.at/en/publications/ • CogALex-VI Shared Task paper describes our winning system in this shared task • LDK 2021 paper has received the Best Student Paper Award • ACL Findings 2021 was also presented in the ACL Workshop on Multiword Expressions • DL and LLD workshop presentation and panel discussion with project leader • Nine invited presentations on Text2TCS • Further dissemination: TermCoord interview given by project leader on terminology

Organization	Project Name	Key Dissemination Activities
University of Turku, Turku NLP research group	Textual paraphrase dataset for deep language modelling	• Publications announcing or utilizing the data • Kanerva et al. Finnish Paraphrase Corpus. 2021. Proceedings of NoDaLiDa 2021 • Chang et al. 2021. Quantitative Evaluation of Alternative Translations in a Corpus of Highly Dissimilar Finnish Paraphrases • Proceedings of the NoDaLiDa'21 Workshop on Modelling Translation • Chang et al. 2021. Deep learning for sentence clustering in essay grading support. • Proceedings of the 14th International Conference on Educational Data Mining (EDM 2021) • The annotation manual has been uploaded to arxiv
Weber Consulting KG	Virtual Personal Assistant Prototype	• Audience reached with project website https://youtwindi.com/
FZI Research Centre for Information Technology	Streaming Language Processing in Manufacturing	• Existing channels from FZI and The Apache Software Foundation were used. • The results were demonstrated at an Industrial IoT workshop ("KIgoesFactory" in June 2021) and at the virtual FZI "Coffee Call" in June. • Already during the experiment phase, various invited talks at meetups, conferences and company events were held.
Institute for Bulgarian Language	Multilingual Image Corpus 2021	• First Bulgarian dissemination event in the context of the European Language Grid (ELG): • Workshop 1: European Language Grid Workshop • Workshop 2: Multilingual Image Corpus Workshop
EDIA BV	CEFR Labelling and Assessment Services	• Fonto editor demo video • LinkedIn / blog post • IATEFL conference in Belfast in May 2022

Organization	Project Name	Key Dissemination Activities
University of West Bohemia	Motion-Capture 3D Sign Language Resources	• Presentation on European Language Grid (online workshop) - 10th Regional ELG Workshop: Czech Republic, Slovakia • Presentation on Project expo at the virtual META-FORUM 2021, on November 15-17 • Publication - Chapter in ELG book (submitted). • Publication - LREC / SLTAT 2022 workshop - Seventh International Workshop on Sign Language Translation and Avatar Technology - in preparation • Publication - SPECOM 2022 - in preparation
Sapienza University of Rome	Universal Semantic Annotator: A Unified API for Multilingual WSD, SRL and AMR annotations	• Three demo papers presented at EMNLP 2021: • AMuSE-WSD: An All-in-one Multilingual System for Easy Word Sense Disambiguation. • InVeRo-XL: Making Cross-Lingual Semantic Role Labeling Accessible with Intelligible Verbs and Roles. • SPRING Goes Online: End-to-End AMR Parsing and Generation. • Paper in review at LREC 2022. • USeA will also feature in A tutorial for AACL-2022 and a workshop on neurosymbolic AI.
Sign Time GmbH	Sign language explanations for terms in a text	• Social media community informed • Potential customers: 149 high potential customers were identified • Press article

Table 6: Examples and key dissemination activities of pilot projects

5 Lessons Learnt

5.1 Lessons Learnt from the Open Calls

Summary of the main lessons learnt from the execution of the open calls, as gathered through the different surveys:

- We aimed at a simple and light-weight procedure, which led to a high number of submitted proposals. At the same time, the simplicity of the proposal template may have led to a higher number of low-quality proposals that were not adequately described or thought through. In both calls, this rather high number of proposals required more person days and increased the costs related to the external evaluators.
- The quality of evaluation reports submitted by external evaluators was not entirely stable and, in some cases, the reports could have been more profound. This was usually balanced by the project coach or Pilot Board.
- It was a good decision to develop the Open Calls Platform internally. Among others, it provided us with more flexibility, control over deadlines and quick and reliable support from the technical team.
- In the ELG project budget, the costs for the Open Calls Platform and for the proposal evaluation didn't correspond to the initial budget and more resources had to be allocated.

5.2 Recommendations and feedback from the Pilot Projects to ELG

As part of the final review process (see 4.2), the pilot projects were asked to draft a self-assessment report. This report included questions related to feedback on ELG. This feedback was gathered during the autumn of 2021 and spring of 2022; many of the suggestions and issues mentioned by the pilot projects have been solved and acted upon in the meantime. Also, there were numerous positive statements regarding the ELG, and the helpfulness and support received from the ELG team.

Here, we summarize the answers to our questions. Detailed answers and feedback on ELG from the pilot projects can be found in Annex 1.

Question: Based on your experience with the integration of your results to ELG, what should be improved in this regard on the ELG side?

Some projects had no issues or suggestions regarding the integration of their results to ELG. Many projects appreciated the straightforwardness of the integration, the technical documentation and the support provided by the ELG team. Also, there were some useful hints and recommendations to ELG (full list of recommendations is in Annex 1, here we provide just a brief selection):

- The ELG platform is very detailed and extensively documented. This also means that it is very time consuming for new users to familiarize themselves with it. Therefore, it might be beneficial to create a quick guide or special guides for different use cases to shorten the learning curve.
- It would be useful to have a bit more code samples (templates).
- It would perhaps be better if the service provider could themselves test whether their services were ELG compatible/functional before ELG review and approval.
- Initially, there were some challenges regarding the output format(s) of our tool. I would suggest higher flexibility in terms of output formats. Also, sample templates for individualized user interfaces to LTs on

ELG could be very useful for other providers and potentially a visualization option for formats where this is readily available.

- The ELG infrastructure could not provide us with anything to identify a user so that we could implement a limit in the number of calls to the service a user could do in a period of time, nor could the ELG infrastructure implement the limits on their part. We think this is something that should be addressed and fixed.

Question: To encourage providers into sharing through the ELG platform, we aim at making the dataset/service provision procedure an experience as light and user-friendly as possible. Given your recent experience, your feedback on the matter will be highly welcome for future improvements. Could you tell us whether the procedure was easy to follow during your contribution (and add why if this was not the case) in terms of content, metadata, steps to follow or any other matter?

Some projects had no comments or suggestions regarding the dataset/service provision procedure. Many projects appreciated guidelines and found the process to be easy and simple enough. Again, there were some useful recommendations to ELG. Selected relevant comments and recommendations from the pilot projects include:

- It may be a good idea to have a few obligatory fields for the description of the resources to avoid short or noninformative descriptions.
- The documentation about integrating a local service into the ELG platform could use more examples.
- The metadata form filling and review should be improved. It's a lot of fields, a lot of half-baked value restrictions (e.g. not allowing language variants such as Finland Swedish), fields that expand with hidden fields after starting to fill them in (e.g. a sudden mandatory requirement to provide a webpage instead of the pdf I had assumed) and difficult to get an overview before it's too late. Due to our wide range of services and language coverage, we filled in these forms almost 50 times. It seems to us that this process would be worthwhile improving.
- More output formats and a method to provide sample inputs for users to easily try out the tools would be nice.
- Reading through the instructions, a little summary or a guideline may be helpful. Sometimes I was soon very deep in the details for metadata structures already. How about a chatbot interaction in a query-answer style so that the provider simply answers interview-like style and fills this way the meta-data key value pairs. After that a little summary what is missing and what needs to be done / provided next (such as longer chunks of text content, uploads of images etc.).
- The instructions on how to prepare an ELG-compatible service to be hosted on the platform could be best transmitted in the form of a "getting started" video tutorial.
- The authorization of the publication of resources should be maintained in the future.

Question: Could you also pinpoint any aspect that could be improved?

Various suggestions and aspects were mentioned in answers from the pilot projects. It included recommendations regarding the ELG website, metadata and resources descriptions, users' analytics etc. Selected comments are included here, the full list in Annex 1:

- If the ELG is not only to serve as a repository of information for language assets but is also to be an institution that promotes these assets, a better presentation on the website would be necessary. This should also include a shop where these assets can be bought and paid for - like in an app store.

- It would be helpful to provide a bit more information on how/why certain metadata is required or used. Another suggestion would be to allow users to edit the metadata in the same place where it is displayed to end-users. We would like to see some more elaborate analytics under My Grid to gain insights into how our APIs are used. For example: how many users, how many call over time (is usage increasing/decreasing over time). Also, a section for support, discussions or customer reviews feature would be useful to engage with customers.

- Improvements in the future may be in several directions: in terms of functionalities (technology is evolving rapidly), in terms of storage, and in terms of resource description.

- Mainly development documentation on the website. Also, having more examples in the documentation, in general, would ease the understanding and implementation of some of the steps.

- The most important parts to improve are: 1) The slowness of the ELG demo. 2) Metadata form filling is too cumbersome. 3) Commercial service support is still work-in-progress. 4) The review of the metadata should probably be limited to some mandatory items, and the rest should be up to the discretion of the service provider.

- I think that maybe ELG could be a hosting solution for many institutions that do not have the hardware or economic means to put online tools that they have developed and that might be demanding hardware-wise.

- The search functionality does not work very well. It is strictly token-based matching (no proper handling of multi-word units) and does not seem to have a good relevance ranking. If this can be done in a more programmatical way that can easily scale to many submissions, then this would help to motivate more submissions. There should also be some way of stress testing services at least in case that ELG considers providing robust multi-user services that can be applied in production ready solutions.

- Following the example of other resource infrastructures (e.g., ELRA), it would be optimal to have a unique identifier for each resource (e.g., International Standard Language Resource Number or a DOI).

- Our suggestions for improvement mainly concern the system for submitting documents. We think that the submission system (proposal submission and review submission) should allow the upload of formatted text (e.g., bold, italics) and tables (to easily provide statistics and performance indicators).

Question: What do you expect from ELG in the future? What are your recommendations for ELG on how to operate in the future?

Very interesting feedback and views of the pilot projects on the future (expected) role of ELG were gathered. From the pilot project's view, ELG should be a true marketplace for languages and services that ensures visibility of the datasets, tools, and services for both scientific and commercial purposes. To achieve this, ELG should invest more into its own visibility. Communication, networking, matchmaking, and community events are also among expected activities for ELG. Some of the pilot projects suggest that funding (like the ELG open calls) should be secured and supported by European Commission.

- We expect the ELG platform to become a central point of contact for potential customers. It would be beneficial to promote communication and match-making function of ELG. An advantage would be if

there were more events offering the opportunity to meet partners and customers and to explore potential. Also, the initiation of further funded projects would boost this goal.

- ELG should also invest in more marketing activities to improve the visibility of ELG and its possibilities.

- We expect the ELG to provide us with better visibility of our dataset and better availability to the wider public for both scientific and commercial purposes. We expect ELG will provide us with access to a wide range of current and future language resources for our further research.

- We hope that the ELG will focus in the future on engaging more with application developers seeking to leverage language services in their products. We would like to see ELG develop into a true marketplace for languages resources and services.

- A long-term operational model for ELG is a necessary condition which will guarantee the development of language resources and technologies across Europe.

- We would like to see more events for current, past, and future ELG projects. For example, we would like to see more workshops --- or events in general, which may foster collaborations within and outside the ELG partners.

- I think ELG could play an important role in facilitating collaborations across Europe and helping the groups for instance to set up Horizon application consortia. ELG could thus play the role of mediating more funding for language technology and research infrastructure work and acting as an expert organization in support of those that aim for getting more language resource funding.

- ELG is an extremely useful initiative and a great platform to strengthen LR and LT development and sharing in Europe. It could be useful to provide some tutorials at summer schools and the like to really extoll the virtues of the platform and make it better known.

- ELG should continue its work of compiling and being the showcase of language and speech technology resources and services. The initial launching has been done, in my opinion, successfully and already hosting many resources and services. But it is also important, and probably more difficult, to keep up the pace and the work, and obtaining funding for it. But no doubt you are on the good way. I think one of the key issues will be to get all the agents in the field to regard being in ELG as an interesting or even necessary thing, so that they all put their services and resources there. The more there is, the easier it will get to convince new ones. Also, I think the idea of the pilot projects is very interesting and helpful. Their twofold aim (obtaining new services for ELG and helping new developments/tools/resources) is valuable both for ELG and the chosen projects. It was an invaluable help for us. ELG should go on with the pilot projects program.

- Maybe the EU can promote ELG as a transparent alternative to commercial providers that run on advertisements or exploitation of personal data compromising privacy concerns. I would expect that the EU will try to push ELG as its main infrastructure for NLP-related resources and services.

- Our recommendation for ELG for the future would be to have more presence on the market and thus a higher rate of possible adoption.

- We expect that ELG will be a long-lasting platform from which resources will be easily accessible for a significant period. For this, we hope that we will get all the necessary information to curate our resources in ELG properly. Moreover, we expect that the ELG staff will continue giving support for the integration of new resources in the platform, as such integration might be complex in the beginning.

We also hope that there will be more open calls in the future to fund the development of new resources for the ELG platform. In addition, we expect that uploading our resources to ELG will allow us to start enriching research discussions with people who get to know our work.

Question: What are the benefits and the impact of ELG for you, your company / institution, and results of your project?

ELG has been a great benefit for all pilot projects. Their experience (and expectations for the future) is that ELG increased their visibility, strengthens market presence, and helps to ensure greater impact on the scientific and commercial community. ELG also provides a possibility to address a larger international market and find new customers or partners for research and development. Some of the interesting comments are listed here, full list in Annex 1:

- The ELG platform will increase our presence and visibility. This is true for potential customers, but also for possible cooperation partners interested in future collaboration in the field of accessibility for the deaf. Furthermore, the ELG enables us not only to research new technologies, but also to develop them as a product in the same breath. We have already been able to win two customers. In addition, we have already found at least two possible qualified cooperation partners who have expressed their interest in joint further development. This means an expansion of the possibilities for us through the ELG to address a larger international market and thus create a certain breakthrough for the Deaf community.
- We expect the publication of our dataset to have a greater impact on the scientific and commercial community if published through ELG. We hope that ELG provides better opportunities to establish new partnerships in science. We also hope that our dataset can be useful for any commercial SL synthesis development.
- The ELG has made it possible for us to extend the language coverage of our services. The ELG pilot project has also enabled us to connect with leading researchers from various academic institutions involved in text readability.
- The project has increased the national and international visibility of the DCL team and its mother organization, the Institute for Bulgarian Language at the Bulgarian Academy of Sciences, through enabling partnerships and collaboration with tech companies, the academia and society through project-related activities (the ELG dissemination workshop, the workshop's media coverage).
- ELG was a crucial enabler for us. Seeing how hard it is to acquire funding for building corpora from scratch, I am nearly certain that if we did not get this ELG funding, this paraphrase corpus would not exist right now, nor would any of the work we built upon it. To my best knowledge, this is the largest manually built paraphrase corpus for any language currently available, stressing the impact ELG had on the creation of what I believe to be a rather unique resource. It is my hope that ELG will be able to facilitate similar projects also in the future, perhaps not directly in terms of funding, but at least indirectly, through various supporting activities, networking, and helping to pave the way for language technology infrastructure work on EU level.
- The ELG project has allowed us to upgrade our service infrastructure for easier distribution via ELG as well as through other channels. ELG now provides a simple way for us to demonstrate our services to

new users. This is a neat outcome of the project for us even if full commercial support is not yet available. Furthermore, we believe that we will continue to utilize other providers' ELG resources and services for our benefit, especially the open-source data resources.

- The main benefit of ELG for our institution is that it allows us to publish our (future) language resources and language services, just like the one created within this pilot project. Furthermore, it provides an easy-to-use interface to our and other language technologies and since we are a translation studies department this is highly valuable. This allows us to easily integrate ELG and our own Text2TCS service into teaching activities and non-technical research projects. For me personally and for more technically skilled colleagues, the way that ELG allows one to reuse many LTs and LRs in our own implementations is highly valuable and I fully and actively support the notion to actively foster a European LR/LT landscape.

- Apart from the obvious and greatly beneficial effect of the ELG funding on the development of the project, the most important effect of being an ELG pilot project is the visibility this has given to the project. Having our service in the ELG grid allows our company and services to reach much greater audiences. Our products are well known, and we have many users and customers in the Basque Country but being in ELG makes us undoubtedly more known internationally.

- ELG now represents an important and strongly organized repository of data and systems. We expect and encourage further adoption by the numerous research institutions and companies working on Italian language technologies. The standardization and ease of access to models and systems for Italian will foster cross- and multi-task learning /multi- and cross-lingual learning experiments.

6   Conclusions

The objective of the open calls was to demonstrate the use and the advantages of ELG in providing basic LT for applications and as a basis for more advanced LT-based modules or components useful to industry. Our main goal was to attract SMEs and research organisations to either contribute additional tools or resources to the ELG platform or develop applications using Language Technologies available in the ELG platform. All these objectives and goals were successfully achieved.

The results of the two open calls demonstrate an enormous interest in the European Language Grid and the Language Technology topic in general. The interest also indicates that the setup, including documentation, proposal template, platform etc., was easy to follow. In total, we received 213 project proposals from 156 different institutions (86 SMEs, 70 research organisations) in 32 different countries (including nine eligible countries outside the European Union); 15 projects were selected for funding, ten in the first open call and five in the second. The total amount requested was approx. 30 mil. €, while the available funding amounted to only 1.95 mil. € (an oversubscription of more than 15 times).

Overall, we were very satisfied with the open calls setup and with the results of the pilot projects. While the results improved the ELG offering in terms of data, tools and services, and the applications developed using the ELG provided mutual benefit to the developers and ELG, we consider the overwhelming interest in the open calls an extremely important, albeit non-technical result: it demonstrates that Language Technologies are of tremendous interest to both researchers and commercial companies. It also shows that the open calls setup, as

designed and implemented, was very attractive and can be considered as a model in similar undertakings in the future.

During the final review process, the pilot projects provided us with valuable feedback on the ELG platform, its functionalities, internal processes (incl. integration), impact on users and benefits for their institutions, as well as expectations for the future of ELG. For example, very interesting feedback and views of the pilot projects on the future (expected) role of ELG were gathered. From the pilot project's view, ELG should be a true marketplace for languages and services that ensures visibility of the datasets, tools, and services for both scientific and commercial purposes. To achieve this, ELG should invest more into its own visibility. Communication, networking, matchmaking, and community events are also among the expected activities for ELG. Some of the pilot projects suggest that funding (like the ELG open calls) should be secured and supported by the European Commission.

ELG has been a great benefit for all pilot projects. Their experience (and expectations for the future) is that ELG increased their visibility, strengthens the market presence, and helps to ensure a greater impact on the scientific and commercial community. ELG also provides a possibility to address a larger international market and find new customers or partners for research and development.

A. Annex

Detailed feedback from pilot projects

As part of the final review process, the pilot projects were asked to draft the self-assessment report. This report included questions related to feedback on ELG. This feedback was gathered during fall 2021 and spring 2022; many of the suggestions and issues mentioned by the pilot projects have been solved and acted upon in the meantime. Here we summarize relevant answers from pilot projects (answers by pilot projects were slightly edited and sometimes shortened, irrelevant parts of text were skipped).

Question: Based on your experience with integration of your results to ELG, what should be improved in this regard on the ELG side?

- Overall, the integration of items into the ELG platform works straight forward. The support provided by ELG specialists is quick and precise and helps to carry out the integration. Sometimes there have been small delays, but then they could always be resolved. **The ELG platform is very detailed and extensively documented. This also means that it is very time consuming for new users to familiarize themselves with it**. Therefore, it might be beneficial **to create a quick guide** or special guides for different use cases to shorten the learning curve. The description of a detailed step-by-step guide would be another way to simplify the process.
- We had **no issues with the integration** of our results since the start of the project. There were no issues or complex problems from our side, so we didn't find any problems on the ELG side that should be addressed.
- It would be really useful to have a bit more **code samples** (templates) available on GitHub for various types of ELG services (e.g. how to implement async services). This would help speed up development time for providers such as us. During later stages of the project an SDK was provided by ELG. We found

the **SDK to be very useful** and it made the integration more easy. We believe that **promoting the ELG Gitlab repository a bit more could be useful**. This may foster community building and provide greater transparency on ELG development for developers that have questions or that may want to contribute to ELG.

- The organization of the platform is very **well designed** and adequate technical assistance is provided. We have not delivered resources as regular users (some of our resources were migrated from other platforms such as ELRC and we delivered the outcomes of our pilot project), so we do not know if regular users can receive **technical support**. It would be good for such support to be available after the end of the ELG project. The **technical documentation** for providers clearly explains how to contribute each type of entity. The documentation helps providers to decide which is the best option for integration of resources.

- We have **no significant suggestion** to share about the integration of our results. Overall, our experience with the integration of our results into ELG is going quite smoothly and according to our original plans.

- The upload process was very smooth and problem-free with the web-based interface. The metadata is very rich and I think it is very good that it allows clear external links to GitHub and the project's website landing page. I am not 100% certain whether we should or should not **upload pre-trained models** to ELG physically, since such models get old relatively fast and perhaps a GitHub+colab -based environment is more natural for such fast-moving targets. Presently, it seems to me that it is not easily possible to upload several related resources at once, since the "Related to LRT" field can only be filled with an already approved resource, so one needs to upload serially, which takes its time.

- The **revision of integrated resources could be faster**, e.g. instead of asking the ELG team to internalize the resource and then making changes, it would be easier for the person acting as 'provider' to modify the resource directly and later submit it for the ELG team control. The metadata field "funded by" should present more options. In the case of our project, it would have been more suitable to have the possibility to separate the two concepts of resource creator and resource provider, which in the case of shared tasks can often be distinct entities.

- Currently, the process of calling existing ELG platform services involves creating corresponding access tokens which expire after a given time. The **decision on the expiry** is generally considered a trade-off between ease of use and security. In contrast to user issued request-response interactions, standing queries such as running processing pipelines within StreamPipes are typically considered long running. As the tool's original focus did not evolve around calling external service calls, it currently does not provide **proper mechanisms to handle token refresh**. Though not a secure alternative, ELG could provide long-lived access tokens which would aid overall development process. Such long-lived access tokens should be easily generated after logging in to the ELG platform in addition to language dependent options such as the Python client.

- It would perhaps be better if the **service provider could themselves test** whether their services were ELG compatible/functional before ELG review and approval? We believe this would have made the integration process more efficient than now.

- Initially, there were some challenges regarding the **output format(s) of our tool**. I would suggest **higher flexibility in terms of output formats**, which in our case could be solved by the fantastic ELG team. Also **sample templates** for individualized user interfaces to LTs on ELG could be very useful for other providers and potentially a visualization option for formats where this is readily available, e.g. RDF.

- Our experience with the integration of our services into ELG was very good. The staff that assisted us was very helpful at all times and answered all our questions and doubts promptly and adequately, so everything went smoothly. Only thing worth mentioning: the ELG infrastructure could not provide us with anything **to identify a user** so that we could implement a **limit in the number of calls** to the service a user could do in a period of time, nor could the ELG infrastructure implement the limits on their part. We think this is something that should be addressed and fixed.

- Streamlining the submission of resources through programmatic interfaces (APIs) would be useful for a more **systematic extension of the services** and the resource catalogue at ELG. There is already work in that direction, which is great, but some manual effort that needs to go through the graphical interface requiring manual work is still needed. We also still experienced some instability of the translation services, which may become problematic if there would be extensive use by online users. The **validation of submissions** is also a bit problematic. This seems to be a manual labour-intensive task that is performed by only a few people who cannot manage to quickly comply with incoming resources. I don't think this will scale if it already seems to be very slow with the few resources and services that come in now basically from the pilot projects.

- Difficult to say from reviewing our project. As we were pioneers together with ELG for featuring SPARQL endpoints, it is probably up to the next organisations and projects to review what we have developed together with ELG. Therefore, the integration was rather handcrafted and not yet based only on the published information on live.european-language-grid.eu. SPARQL is also a bit of a different "animal". ELG may **consider imposing quotas**, i.e. amount of returned records / triples per query, so that a query issued via ELG cannot overload the SPARQL endpoint. But this is common to all SPARQL endpoints.

- The ELG data/tools integration system is very flexible. It can host different types of annotation and annotation tools. At the same time, however, on first use it requires a **significant amount of learning time**. As regards metadata, there are a certain number of items to enter and options to choose from. To assist users in their submissions, a support service is therefore fundamental. The ELG team that helped us in completing our procedure was very helpful and qualified. Concerning tool adaptation, having even more concrete examples to follow (e.g. ad hoc created examples, source code of other submitted tools) would help.

Question: To encourage providers into sharing through the ELG platform, we aim at making the dataset/service provision procedure an experience as light and user-friendly as possible. Given your recent experience, your feedback on the matter will be highly welcome for future improvements. Could you tell us whether the procedure was easy to follow during your contribution (and add why if this was not the case) in terms of content, metadata, steps to follow or any other matter?

- A platform whose task it is to contain a large number of diverse assets, tools and functions, which can then also be easily found and tried out on the platform, is of course not entirely simple. Therefore, it is to be expected that a simple handling of the system is a challenge. Nevertheless, ELG has managed to find a **reasonable user-friendly compromise**. The creation of new items is in principle simple and also understandable. However, it is sometimes not easy to understand which of the inputs have what effect on the display of the items. Also, the display is somewhat simple and does not look appealing (e.g. logo very small) because you cannot customize it.

- The process of creating new items linked to source data seems to be **simple enough**. We think that the high diversity of language resources doesn't allow the process to be simpler. The availability of adding metadata to specify a particular item is sufficient.

- All in all, we found it **fairly easy to share our resources and services** on the ELG. For the most part the documentation was up-to-date and provided useful guidance. The services we provided were services that can be called in an asynchronous manner. At the time of integration, documentation was mostly focused on building synchronous services. Therefore, the code samples provided by ELG were not always useful. With some support from the ELG technical team we managed to create the services in the end. Providers have to provide quite a bit of metadata which can be a bit overwhelming initially. Our submission was also not accepted at our first attempt because we hadn't quite understood the meaning of some metadata fields. Again, this was quickly resolved with support from the ELG technical team.

- The public availability of data enabled by an authorized person ensures that inappropriate resources would not appear on the platform. In case the provider decides to update something, the changes would not become visible until the act of their publishing. The **authorization of the publication** of resources should be maintained in the future. The way the metadata is organized, the steps for delivering the resources and the search functionalities are very well designed. In some cases, the descriptions of the resources are too short, and the user has to visit the page of the resource to decide whether he/she is interested in it or not. It may be a good idea to have a few **obligatory fields for the description** of the resources to avoid short or noninformative descriptions.

- As is often the case, developers wish for more documentation, especially for more examples. The documentation about adding an item to the ELG catalogue is already quite extensive, with several figures to guide users through the steps that need to be taken. The same is also valid for updating an item. However, the **documentation about integrating a local service** into the ELG platform could use **more examples**.

- It was reasonably easy in my opinion, and the web interface is much appreciated. I was having hard times to orient myself in the "**Categories / Intended Application" list**; it was quite long, and it took a while to come up with an exhaustive list of what I thought could be possible uses for a paraphrase corpus. Further research will hopefully demonstrate my choice to be incomplete. Categories / Domain label lacks any help and I found it difficult to come up with any values, not being sure what was supposed to go there. In **text features / size unit**, I found "triple" but would have needed "pair". Unlike all other fields, this one did not accept proposals for new values and making a choice was quite difficult.

- The setting up of meta-data for the project site and also the organization site with uploading the configuration files was cumbersome during the beginning of the project time, last year. This was luckily improved and also the steps to follow are sufficiently described. Besides that, we would have liked to have **one access point** for searching Language Technologies and Language data and resources on the ELG site instead of first entering one catalogue and then trying to find the suitable content there and if necessary going into another catalogue. In addition, the overall online documentation is well-structured and provides a central and easy to use landing page for getting started.

- Generally, online guidelines and human integration support from ELG team have been very good and sufficiently detailed. The **metadata form filling and review** should be improved. It's a lot of fields, a lot of half-baked value restrictions (e.g. not allowing language variants such as Finland Swedish), fields that expand with hidden fields after starting to fill them in (e.g. a sudden mandatory requirement to provide

a webpage instead of the pdf I had assumed) and difficult to get an overview before it's too late. Due to our wide range of services and language coverage, we filled in these forms almost 50 times. It seems to us that this process would be worthwhile improving.

- Given instructions were very useful for providing the LT descriptions in a structured format. More **output formats** and a method to provide **sample inputs** for users to easily try out the tools would be nice. The steps to follow were very clear and the **online manual** was very helpful in this regard. It would be important to ensure that it **stays updated** with changes in ELG and it should include a guide on how to provide an individualized user interface or in our case output format. The support for standard LR and LT formats would be helpful, i.e., TBX, SKOS, RDF, etc.

- Our impression was also that the interface was generally **very well documented**. I really cannot think of any way to further improve and ease the integration process.

- The platform looks great, and functionality is fine. **Metadata creation** is a bit tedious, and the XML schema seems to change from time to time, which can be annoying when those are generated using some routines. However, I have no clear suggestion on how to improve this and to make it easier. The effort of describing the essential properties of the resources will stay and cannot be avoided (even though researchers never like to take care of that part).

- Reading through the instructions, a little summary or a guideline may be helpful. Sometimes I was soon very deep in the details for metadata structures already. How about a **chatbot interaction** in a query-answer style so that the provider simply answers interview-like style and fills this way the meta-data key value pairs. After that a little summary what is missing and what needs to be done / provided next (such as longer chunks of text content, uploads of images etc.). Would also be a showcase for natural language interaction with ELG.

- The procedure was easy to follow. To encourage providers, the **number of obligatory fields** should be kept as low as possible. The procedure to modify an already integrated resource should be faster. From the systems perspective, the procedure was also well described in the documentation, but several details became much clearer only with a direct exchange with a responsible person on ELG side. The instructions on how to prepare an ELG-compatible service to be hosted on the platform could be best transmitted in the form of a **"getting started" video tutorial**. This would help new users that want to get a new service ready for ELG.

- At the time of submission, it was unclear **where to provide launch documentation** about the container we provide. Typically, we would describe how to launch this container with a docker-compose file. It took us quite a while to figure these out. The **steps to follow** were not always clearly described, which can lead to delays, especially when you are not dealing on everyday bases with the ELG platform.

- To accommodate **different types of datasets/tools** and make them easy to understand and access, one must accept some compromise. In this case, the main compromise is the large amount of metadata to provide to share a resource. Adding this metadata requests thorough resource documentation and curation, which also helps keeping track of the quality of the resource. Fortunately, after a first submission has been made, the following ones are faster. This is possible thanks to the platform functionalities of (i) creating a new version of a resource starting from a previous version; (ii) copying the metadata to create a new resource from another resource. As regards the **steps to follow**, their sequence is clear.

Question: Could you also pinpoint any aspect that could be improved?

- If the ELG is not only to serve as a repository of information for language assets but is also to be an institution that promotes these assets, a **better presentation on the website** would be necessary. This should also include a **shop where these assets can be bought and paid for - like in an app store**. This would also make it possible to launch targeted marketing campaigns concerning a specific category of assets. Also, the accessibility aspect of the platform should be considered. We would recommend that the entire platform be able to be translated into sign language.

- It would be helpful to provide a bit more information on **how/why certain metadata is required** or used. Another suggestion would be to allow users to **edit the metadata** in the same place where it is displayed to end-users. That would make it more obvious how the metadata is used. To make the process more lightweight it might be helpful to provide an (default) option where providers only must provide a minimal set of metadata. More elaborate metadata could then be provided under an "advanced" screen. We would like to see some more **elaborate analytics under My Grid** to gain insights into how our APIs are used. For example: how many users, how may call over time (is usage increasing/decreasing over time). Also, a section for support, discussions or customer reviews feature would be useful to engage with customers.

- Improvements in the future may be in **several directions**: in **terms of functionalities** (technology is evolving rapidly), in terms of **storage**, and in terms of **resources description**. For example, more information for testing and evaluation of the resources, tools and services might be provided, if it is available, and users could be allowed to give feedback for the resources based on their experience. Another simple example, the sort function can be accomplished with more options: newest, most frequently viewed or most frequently downloaded resources. However, most important is for the ELG to continue to function as a coordination centre and a cloud-based platform. It would also be good to keep the opportunity for demonstrating the importance of language technologies through the implementation of relatively small projects.

- Mainly development **documentation on the website**. Although the documentation for integrating a project into the ELG platform is already quite good (well-organized into sections, sometimes providing images and/or visualizations), it would be great to provide more documentation about the integration of systems that are not hosted on the ELG platform but interact with it through a proxy server. Also, having **more examples in the documentation**, in general, would ease the understanding and implementation of some of the steps. It would be great to have a running example, whenever possible throughout the steps.

- The general **overview of the ELG website** is a little confusing as it not quite clear at first sight how to find specific LT components. Maybe, the wording "Technologies" is somewhat misleading and could be better described as "Services" or "LT Components". In addition, while the documentation and description of the LT components is generally sufficient, it lacks some **examples** within the try/out tab to better illustrate how to interact with the component. Currently, a Python SDK is provided to interact with ELG LT services. While this surely addresses most of the use cases, further clients for other languages such as Java could further increase the dissemination.

- The most important parts to improve are: 1) The **slowness of the ELG demo** (and also real integration?). Our ASR is faster than real-time, this is an important aspect which is not working in the ELG demo. All "try out" services we've tried take way too long, so it's not only our services. There's a clear risk here that a user thinks "great service, I wonder if I can get the tool/service from somewhere else

where it's faster". 2) **Metadata form filling** is too cumbersome. 3) **Commercial service support** is still work-in-progress. 4) The **review of the metadata** should probably be limited to some mandatory items, and the rest should be up to the discretion of the service provider, e.g. to describe what their tool does or provide links to documentation, descriptions in non-English, etc.

- It would be great to provide a **template for possible adapted user interfaces** for the "Try out" function of ELG. The extremely helpful ELG integration team provided us with individual guidance on this point and we are very happy with the final solution. Some samples/templates could give other ELG LR and LT provides inspiration for their own services. Also **providing a sample input** to LTs could be useful also for other LTs and the representation of this sample input could be improved (currently it is really just a text description, but it could be a placeholder instead of "Type text to annotate" in the input window or something similar). Furthermore, it might be more attractive to allow people to try out provided services without requesting them to login if at all possible. Trying out LTs might provide people with an additional incentive to register with ELG.

- I think that maybe ELG could actually be a **hosting solution for many institutions** that do not have the hardware or economic means to put online tools that they have developed and that might be demanding hardware-wise (neural networks often need disk space, memory, CPUs, GPUs...) and, therefore, economically too. This was not our case (having our services in ELG gives us and our services international visibility, but we already have them hosted in servers), but it can be interesting for many agents. In order to achieve that, ELG would need to offer powerful hardware, monetization options, etc.

- The **search functionality** does not work very well. It is strictly token-based matching (no proper handling of multi-word units) and does not seem to have a good relevance ranking. If this can be done in a more programmatical way that can easily scale to many submissions then this would help to motivate more submissions. There should also be some way of **stress testing services** at least in case that ELG considers to provide robust multi-user services that can be applied in production ready solutions. Team members also mentioned that some services and connections were a bit unstable and token validity was short. I have not tested this myself but a real stress test might reveal further shortcomings.

- Concerning the ELG hub itself – little to improve from today's point of view. We've also played with a couple of services there, and it is a nice experience. By design, we could not learn from another SPARQL endpoint. For the furture, ELG may also consider having a TryOut UI for **"aggregated" queries**, i.e. querying the one endpoint and "piping" its results into another endpoint. This is whery the URI/LLOD aspect comes into play. This may really be a brilliant showcase for SPARQL, URI/LLOD and of course ELG itself. We could imagine the one endpoint being rather a dataset, the other endpoint being a service - and then bring the results together.

- Following the example of other resource infrastructures (e.g., ELRA), it would be optimal to have a **unique identifier for each resource** (e.g.,International Standard Language Resource Number or a DOI). The first time a dockerized system is run, the **service startup takes some time**, while the service is much faster in the following calls.

- Improvement should be done in the following areas: **Provide more language sets**. Provide an **easy integration process** which functions seamless. In addition provide a convenient user experience with intuitive user interface that also non technical people can work with ELG. Improve ELG platform support by introducing service levels and a ticketing system.

- Our suggestions for improvement mainly concern the **system for submitting documents**. We think that the submission system (proposal submission and review submission) should allow the upload of for-matted text (e.g., bold, italics) and tables (to easily provide statistics and performance indicators). In fact, this would make the submitted project proposals easier to understand and the final reviews more clear and complete (the current choices seem to be inserting data in text lines or adding pointers to external documents). In addition, we reckon that the minimum of 600 characters for each box of the review should be reduced to 300 or 400 hundred.

Question: What do you expect from ELG in the future? What are your recommendations for ELG on how to operate in the future?

- We expect the ELG platform to become a **central point of contact for potential customers**. For exam-ple, if they are looking for accessible communication options for deaf people. The same applies to co-operation partners with a project idea in which sign language plays a role. Therefore, it would be bene-ficial to promote this **communication and match-making function of ELG**. An advantage would be if there were **more events** offering the opportunity to meet partners and customers and to explore po-tential. Also the initiation of further funded projects would boost this goal. For example that could help to expand the assets to additional languages. Another possible idea to make the ELG more attractive for companies and to make the competition more interesting would be to give awards for the best pro-jects.
- ELG should also invest in more marketing activities to **improve the visibility of ELG** and its possibilities. ELG is not widely known yet in Europe although it is a great project and has huge potential.
- We expect the **ELG to provide us with better visibility** of our data-set and **better availability to the wider public** for both scientific and commercial purposes. We also expect to be able to present addi-tional datasets and tools that will be created in the future via ELG and expand the current dataset. It is crucial for developing artificial intelligence-based tools for sign language synthesis to have large and diverse data. On the other hand, we expect ELG will provide us with **access to a wide range of current and future language resources** for our further research. We would also like ELG to give rise to MoCap datasets from others.
- We had hoped to see a bit more **commercial interest** in the services we developed and integrated with the ELG. Therefore, we hope that the ELG will focus in the future on engaging more with application developers seeking to leverage language services in their products. We would like to see **ELG develop into a true marketplace** for languages resources and services. Also, we are hoping to see more oppor-tunities provided to engage with the ELG community. Now there are little opportunities for (online) in-teraction on the ELG platform. As a consequence, we found it quite difficult to see how and by whom our services are used or what needs users of the ELG platform may have. From the newsletters we did see some numbers on the number of resources and services shared on ELG, but less **information about users and usage**. Finally, we are looking forward to **integrating with the billing services** once they be-come available on the ELG. This would really help providers like us commercialize their technologies and help drive the development of new services and support for additional languages.
- A **long-term operational model for ELG** is a necessary condition which will guarantee the development of language resources and technologies across Europe. We are aware of plans for a non-profit legal en-tity taking over operation of the ELG platform after the end of the ELG project. Our opinion, however, is

that the sustainability and further development of the ELG should be guaranteed by EC support. Three main types of platforms for resources can be generally recognized: such that provide options for maintenance and free distribution of resources and tools; such that provide information for resources and tools; and such that distribute resources and tools. The ELG as a platform for resources, tools and services combines the three types and to ensure the sustainability of the distribution in the future, elaboration of the maintenance functionality could be envisaged.

- We would like to see more **events for current, past, and future ELG projects**. For example, we would like to see more **workshops** --- or events in general, which may foster collaborations within and outside the ELG partners. We would also love to see more efforts in advertising ELG projects to other (ELG or non-ELG) partners and members. We believe that we could learn more about how to organize our own projects and improve them by discussing with other people, teams and groups that are currently following our same path or did so in the past (in previous editions or similar calls). In general, any occasion to share our ideas, listen to the experiences of other experts, and discuss.

- I have to say this ELG pilot project was a gift from heaven to us, as we were able to carry out work which would otherwise be next-to-impossible to get funding for. By being able to fund actual resource creation rather than machine learning on already existing resources, ELG has had an enormous impact. The massive number of proposals you received for review only shows how important and much-needed such infrastructure funding is. I understand it is impossible to keep funding projects through ELG and the two pilot calls seem to be the only pilot calls there will be. But I think ELG could play an important role **in facilitating collaborations across Europe** and helping the groups for instance to set up Horizon application consortia. ELG could thus play the role of mediating more funding for language technology and research infrastructure work and acting as an expert organization in support of those that aim for getting more language resource funding. For instance, keeping seminars on relevant open calls which would support language infrastructure development work would potentially have a great impact.

- In the future, it would be great to have **free accessible LT technologies** and text data without the need to login which could be achieved by providing different API-auth-tokens and disabling them by detected misuse. In addition, long-lived access tokens as exemplified using the Python Client could be generated and managed from the browser for ease of use. Furthermore, a rich set of client options allow to increase the likelihood of broad dissemination and acceptance. As such, multiple client options for a variety of languages (Python, Java, etc.) allow to easily interact with ELG LT components to further improve the overall developer experience.

- We hope that ELG continues to develop to fulfill the ambitious vision. This includes **supporting commercial use of ELG**, i.e. SMEs like us getting paid for our provided services from ELG end users. We are also keen on **getting statistics** on who is using our services through ELG to e.g. monitor that our terms of service are followed. We have the impression that this is not yet ready, but in good progress on ELG side. ELG must also clarify which types of services are supported in the near future. For example, our own SaaS for ASR supports (near) real-time ASR but ELG doesn't yet, GPU support is lacking at ELG, and probalby other things too. Are these cases already on the roadmap or not? At the moment, it looks like ELG can be a fantastic forum for the academic and open source community, especially if the docker images themselves can be shared for free, but it does not yet seem to be a mature service platform for commercial language technology providers like us.

- ELG is an extremely useful initiative and a great platform to strengthen LR and LT development and sharing in Europe. It could be useful to **provide some tutorials at summer schools** and the like to really extoll the virtues of the platform and make it better known. Also shared tasks and challenges organized in relation to ELG might help to disseminate the platform. With an increased number of users and developers it might become even more attractive. Other than that storage space is an issue since many deep learning-operated models (like ours) require more than the standard setting, which thankfully could be extended in our case. However, if **innovative solutions** using neural networks should be integrated, it might be reasonable to strengthen the ELG **hardware capacities**. To set itself apart from Huggingface and the like it might be wise to accommodate many LR and LT-output specific formats, such as Translation Memories, RDF, SKOS, TBX and ideally even provide a visualization interface for LRs or at least for formats where visualizations are readily available (e.g. RDF).

- ELG should continue its work of compiling and being the showcase of language and speech technology resources and services. The initial launching has been done, in my opinion, successfully and already hosting many resources and services. But it is also important, and probably more difficult, **to keep up the pace and the work**, and obtaining funding for it. But no doubt you are on the good way. I think one of the key issues will be to **get all the agents in the field to regard being in ELG** as an interesting or even necessary thing, so that they all put their services and resources there. The more there is, the easier it will get to convince new ones. Also, I think the idea of the pilot projects is very interesting and helpful. Their twofold aim (obtaining new services for ELG and helping new developments-tools-resources) is valuable both for ELG and the chosen projects. It was an invaluable help for us. ELG should go on with the **pilot projects program**.

- I hope that ELG will continue to provide open services and resources that support the NLP community and end users of language technology. I hope it does not disappear behind a pay-wall and I am also skeptical that this would work as a business model. Maybe the EU can promote **ELG as a transparent alternative to commercial providers** that run on advertisements or exploitation of personal data compromising privacy concerns. I would expect that the EU will try to push ELG as its **main infrastructure for NLP-related resources and services**. In order to really create an alternative to commercial solutions a lot of optimization needs to be done to provide robust and reliable solutions for millions of users and requests.

- Aiming at becoming a **reference for language technology in the EU**, ELG should continue its effort towards easy integration and merging of resources to and from other infrastructures (e.g., CLARIN). This includes harmonization of data formats and metadata. On the service side, there are several routes for improvements we can envision. First, **combining services in a pipeline** could make the language grid more useful in everyday use. For instance, a combination of a custom POS tagger and a custom parser. However, for this to work, all input/output would have to be exposed to the API more formally and according to a common communication protocol. Second, having services on the ELG that **perform evaluation for NLP shared tasks** (also originated from different evaluation campaigns, such as Semeval, Iber-LEF, etc,) could be useful. Hosting the evaluation services, in the dockerized fashion that we are accustomed to in the ELG platform, we could make sure that evaluations of old tasks are reproducible in the long term. Finally, it could be helpful to **ease the sign-up process** and open it to the public.

- Would be great to provide **follow-on financing options** and guidance to develop the prototype into a minimum viable product. If this is not available, many of the results from the ELG funded projects will

simply die or not used any further. Our recommendation for ELG for the future would be to have more **presence on the market** and thus a higher rate of possible adoption. In our network, most people have never heard about ELG. This can be changed by active social media activity and also by involving the private sector more into the ELG work. As an example someone could mentioned here the GAIA-X - a federated data infrastructure for Europe project.

- We expect that ELG will be a **long-lasting platform** from which resources will be easily accessible for a significant period of time. For this, we hope that we will get all the necessary information to curate our resources in ELG properly. Moreover, we expect that the ELG staff will continue **giving support** for the integration of new resources in the platform, as such integration might be complex in the beginning. It would be important that resources and tools developed under the ELG pilot projects receive particular attention, so that their diffusion and, possibly, their further development (e.g., new languages) is supported by ELG. For instance, ELG resources from pilot projects could be suggested to be used in EU projects or initiatives. We also hope that there will be **more open calls** in the future to fund the development of new resources for the ELG platform. In addition, we expect that uploading our resources to ELG will allow us to start **enriching research discussions** with people who get to know our work.

Question: What are the benefits and impact of ELG for you, your company / institution, and results of your project?

- ELG platform will **increase our presence and visibility**. This is true for potential customers, but also for possible cooperation partners interested in future collaboration in the field of accessibility for the deaf. Furthermore, the ELG enables us not only to research new technologies, but also to develop them as a product in the same breath. Since the ELG is in the field of research and development, there is a greater opportunity to exchange ideas, collaborate or revise/develop technologies with other competitors or interested parties in the field of linguistic innovation. This opportunity provides a greater chance to advance the digitization of language and gain new insights. In the course of this, we were able to create a marketable product and the underlying basic version of an encyclopedia in ÖGS (Austrian Sign Language) and DGS (German Sign Language). We have already been able to win two customers who use LookApp on their websites. In addition, we have already found at least two possible qualified cooperation partners who have expressed their interest in joint further development. These are not only true for German or EU language area, but also in the Asian, Arabic and American area. This means an expansion of the possibilities for Sign Time through the ELG to **address a larger international market and thus create a certain breakthrough** for the Deaf community.

- We expect the publication of our dataset to have a **greater impact on the scientific and commercial community** if published through ELG. We hope that ELG provides better opportunities to **establish new partnerships** in science. We also hope that our dataset can be useful for any commercial SL synthesis development. The topics used for dataset (weather forecasts, animal description) creation can be directly applied for the domain without major restrictions, and the publication via ELG is suitable. The dictionary and signs in context are universal and the variation in signers enables a larger selection of avatars (male/female, big/small, old/young, etc.) We hope our dataset will be used for sign language synthesis realized by 3D avatar.

- The ELG has made it possible for us to **extend the language coverage** of our services. Also we have been able to **create an infrastructure** (most notably a data labelling application) to easily add additional

languages in the future. The ELG pilot project has also **enabled us to connect with leading researchers** from various academic institutions involved in text readability. We believe this will help us in the future to validate the credibility and quality of our work One of our project objectives was to extend the market reach of our Services and API's. By integrating into the ELG we believe that we have significantly lowered the barrier of adoption for developers and hope usage numbers will further increase as the ELG becomes more known as a marketplace for language technology.

- The ELG project resonates with DCL's (Department of Computational Linguistics) dedication to developing, **providing and making publicly available high-quality largescale language resources and technologies**. The opportunity to work on the ELG project has contributed greatly to **expanding the team's expertise** in collecting, processing, curating, annotating and modelling largescale multimodal content. The project has increased the national and international visibility of the DCL team and its mother organisation, the Institute for Bulgarian Language at the Bulgarian Academy of Sciences, through **enabling partnerships and collaboration** with tech companies, the academia and society through project-related activities (the ELG dissemination workshop, the workshop's media coverage). Established, young and prospective scientists both from the Institute and from other academic institutions have been involved in the implementation of the project, attracting new valuable members to the DCL team.

- The immediate impact for the Sapienza NLP group and Sapienza University of Rome is **enabling research in semantics** (lexical and sentence-level) through the **development of state-of-the-art models**. A direct consequence is that ELG has funded researchers and engineers in Natural Language Processing, an area that is becoming increasingly pervasive in today's technology and, hence, our life. The demand for NLP experts is also increasing not only in Academia but also in Industry. The possibility of partaking in such a project has increased the interest of some of our researchers, so, in the long term, we envision further collaborations with ELG and/or its partners, given the success of this project.

- ELG was a **crucial enabler** for us. Seeing how hard it is to acquire funding for building corpora from scratch, I am nearly certain that if we did not get this ELG funding, this paraphrase corpus would not exist right now, nor would any of the work we built upon it. To my best knowledge, this is the largest manually built paraphrase corpus for any language currently available, stressing the impact ELG had on the creation of what I believe to be a rather unique resource. Further, one of our project annotators has since become a PhD student and gotten a personal grant from a private foundation, which was undoubtedly helped by her involvement in this project. Further two of our annotators have found follow-up work in different projects at the university and will strive to stay in the academia. It is my hope that ELG will be able to facilitate similar projects also in the future, perhaps not directly in terms of funding, but at least indirectly, through **various supporting activities, networking, and helping to pave the way for language technology infrastructure** work on EU level.

- We were able to **experiment with LT technologies**, especially with the ELG provided LT technologies which we can use in the future. Furthermore, we were able to further develop StreamPipes with new components such as a Python Wrapper which comes in handy for all machine learning code, a file management system and StreamPipes Client API. Besides, we managed to extend the scope of potential applicational use cases for StreamPipes which turns out to be also well-suited in scenarios where processing is outsourced to external services. This accelerates the development of new processors by wrapping existing services as pipeline elements which can be easily leveraged by non-technical domain experts to create LT pipelines in a self-service fashion.

- The ELG project has allowed us to **upgrade our service infrastructure for easier distribution** via ELG as well as through other channels. ELG now provides a simple way for us to demonstrate our services to new users. This is a neat outcome of the project for us even if full commercial support is not yet available. Furthermore, we believe that we will continue to utilize other providers' ELG resources and services for our benefit, especially the open source data resources. From our experience with trying to utilize open source tools from the academic community, the ELG idea of the researchers (and other developers) providing their open source tools as shareable dockers with an exposed API sounds like a great improvement over the current situation.

- The main benefit of ELG for our institution is that it allows us to **publish our (future) language resources and language services**, just like the one created within this pilot project. Furthermore, it provides an easy-to-use interface to our and other language technologies and since we are a translation studies department this is highly valuable. This allows us to easily integrate ELG and our own Text2TCS service into teaching activities and non-technical research projects. Furthermore, ELG financially supported the development of a terminological tool that has long been needed in terminology science and practice, which also positively impacts this field. For me personally and for more technically skilled colleagues, **the way that ELG allows one to reuse a large number of LTs and LRs in our own implementations is highly valuable** and I fully and actively support the notion to actively foster a European LR/LT landscape.

- Apart from the obvious and greatly beneficial effect of the ELG funding on the development of the project, the most important effect of being an ELG pilot project is the **visibility** this has given to the project. Having our service in the ELG grid allows our company and services to **reach much greater audiences**. Our products are well known and we have many users and customers in the Basque Country, but being in ELG makes us undoubtedly more known internationally. Also, because of being a pilot project, ELG has invited us to give talks in METAForum 2020 and SEPLN 2021, which is a great dissemination opportunity. Also, when the prestigious and well known IEEE Spectrum magazine wrote an article about ELG, one of the pilot projects chosen to write about was ours, and it had a large presence in the article, being present even in the title. It was great for our project and our language to be mentioned there, and this would be unthinkable were we not in ELG.

- ELG will be a **good resource** for people who want to reuse models, start NLP research with string baselines or just need language technology in order to facilitate their work (e.g. academic research / education in digital humanities, translation studies, translator training, …). We will try to investigate the possibilities to use ELG resources and services in our daily research and also in our degree programs. Students may benefit from the **resources offered in the ELG catalogue** and we will also promote the ELG tools in neighboring disciplines such as translation studies, general linguistics and computer science/AI research.

- MKS2LLOD was a huge effort for us with significant risk. Without the funding from ELG we had hesitated to start with that development. Through ELG we could start the development and are very happy with the results. This is where European funding is of **great help to SMEs** like us. The benefits are significant. We have a new **unique selling point**. We are recognised as a pioneer in bringing terminology resource onto the next level, into other business processes. Thus terminologists no longer contribute "only" language resources to their organisation. Terminologists become data curators for the semantic web. They facilitate to move from an application-centric IT landscape to a data-centric landscape. Thus,

a terminology resource becomes a significant pillar in the set of resources of future, modern IT landscapes. The SPARQL endpoint makes this tangible and increases the trust into our technology and proves our vision.

- For the Italian NLP community, ELG now represents an **important and strongly organized repository of data and systems**. We expect and encourage further adoption by the numerous research institutions and companies working on Italian language technologies. The standardization and ease of access to models and systems for Italian will **foster cross- and multi-task learning /multi- and cross-lingual learning experiments**. This effect has been observed on a small scale in the past when a number of tasks have been presented at EVALITA following a common annotation scheme. We expect this effect to be amplified by leveraging the technological support of the ELG. Furthermore, EVALITA could serve as an example for similar initiatives from analogous evaluation campaigns on other EU languages (e.g. IberLEF on Spanish and GermEVAL on German). Our own institution, University of Turin, and in particular the research group participating in this project (Content-centered Computing) has a longstanding tradition in the creation of language resources. It is a natural by-product of this project that the future resources and systems by our team will be published in the ELG.

- The ELG open call has allowed us to obtain funding for the development of a new multilingual annotated corpus. This corpus will be a relevant resource for developing future systems in the clinical domain. The fact that the resource is **freely available for research purposes will foster its exploitation** by other groups in NLP. In addition, being able to share our resources on the ELG platform will allow us to i) easily distribute the outcome of our work and ii) have a long-lasting repository (also on GitHub). The commitment of ELG with shareability has also encouraged us to build CC-BY resources and to pay special attention to documentation.