



EUROPEAN LANGUAGE GRID

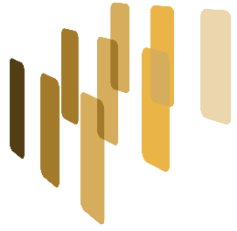
2nd open call for pilot projects

WEBINAR FOR APPLICANTS

pilot-projects@european-language-grid.eu

12-11-2020

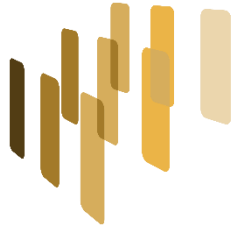
<http://www.european-language-grid.eu>



**EUROPEAN
LANGUAGE
GRID**

Overview of the webinar today

- What is European Language Grid (ELG)
- ELG Grid Platform
- Open Call for Pilot Projects
- Q & A



**EUROPEAN
LANGUAGE
GRID**

What is European Language Grid (ELG)



EUROPEAN LANGUAGE GRID

What is the European Language Grid (ELG)

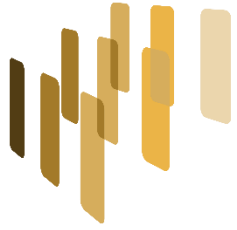
- Cloud-based, scalable technology platform for Language Technologies
- Platform contents:
 - Functional: Running services with APIs
 - Non-functional:
 - Language Resources (data repository, metadata)
 - Software tools, packages
 - Demos, catalogues for additional resources, institutions, other stakeholders
- Marketplace for the European LT business space – allows data, tool, service providers to use the platform for deployment and use
- <https://www.european-language-grid.eu>



EUROPEAN LANGUAGE GRID

Selected ELG Objectives

1. Establish the ELG as the primary platform for LT in Europe, to boost the “Multilingual Digital Single Market”
2. ELG as a scalable computing as well as repository platform for Language Technologies
3. Establish the ELG as the European marketplace (connect demand and supply)
3. Allow both non-commercial as well as commercial usage (easy upload, use, billing services, etc.)
4. Use the platform for easy innovation and prototyping, demonstrate it by the (external) **pilot projects**



**EUROPEAN
LANGUAGE
GRID**

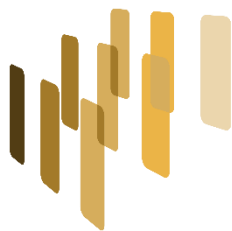
Concept and Methodology

- **Grid Platform**

- Robust, scalable, modern standard infrastructure technologies (constantly developed)
- Technology for deployment of running services, monitoring, billing
- Interactive, modern web interface for all types of users
- Technology for a searchable catalogue (directory) and repository of
 - Functional services, data sets, tools, technologies, models etc.
 - LT companies, research organisations, research projects, service and application types, languages etc.

- **Grid Content**

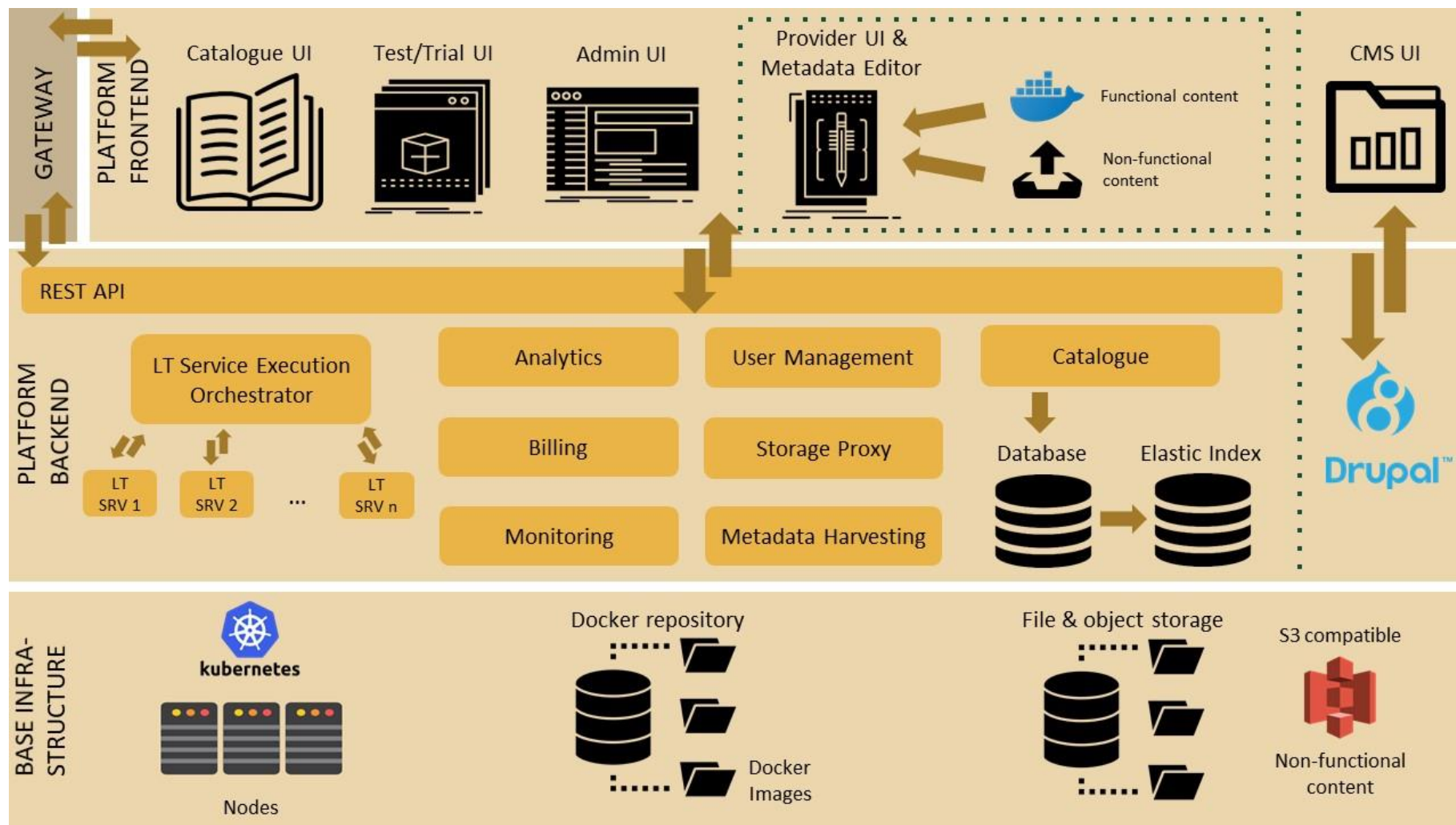
- **Grid Community**



EUROPEAN LANGUAGE GRID

Architecture

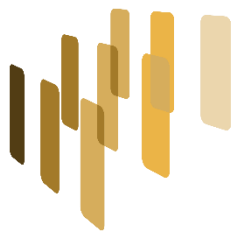
(Current Working Version)





Concept and Methodology

- **Grid Platform**
- **Grid Content**
 - **Content:** services, language resources, data sets, tools
 - **Functional Content:** running services, can be uploaded and integrated into other systems. Stakeholders will be able to upload and make available their own services.
 - **Non-functional Content:** LRs and data sets but also records of companies, research groups etc.
 - Functional content implemented by
 - Containerising an LT service (dockerisation)
 - Easy and efficient for LT providers to create and upload containerised services
- **Grid Community**



Services and Tools

	Language Group			Totals
	A	B	C	
Basic linguistic processing	146	30	63	239
Entity recognition and linking	102	7	11	120
Classification (sentiment, opinion, topic, language etc.)	40	4	18	62
Machine Translation	105	4	14	123
Other text processing (including parsing, summarisation etc.)	35	7	13	55
Speech tools (ASR, gender detection, synthesis etc.)	26	5	14	45
Totals	454	57	133	644

Services and tools to be ingested by the ELG partners

A – Official EU languages: Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovenian, Spanish, and Swedish

B – EU candidate languages and Free Trade Partners: Albanian, Basque, Catalan, Galician, Icelandic, Norwegian, Scottish Gaelic, Welsh, Serbian, Turkish, Ukrainian

C – Languages spoken by EU immigrants; languages of political and trade partners: Afrikaans, Arabic, Berber, Cebuano, Chinese, Hebrew, Hindi/Urdu, Indonesian, Japanese, Korean, Kurdish, Latin, Malay, Pashto, Persian (Farsi), Russian, Tamil, Vietnamese



Datasets and Resources

	Open Access	Language Group A			Language Group B			Language Group C			Totals
		Corpora	Lexicons	Models	Corpora	Lexicons	Models	Corpora	Lexicons	Models	
META-SHARE	yes	617	447	16	55	54	0	84	51	1	1325
	no	582	550	1	44	65	0	198	94	0	1534
ELRC-SHARE	yes	317	114	0	3	1	0	0	0	0	435
	no	74	16	0	2	1	0	0	0	0	93
ELDA	no	563	1012	0	35	18	0	250	54	0	1932
ELG	mixed	74	108	43	0	0	12	4	1	21	263
Totals		2227	2247	60	139	139	12	536	200	22	5582

Datasets and resources to be provided by the ELG partners

A – Official EU languages: Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovenian, Spanish, and Swedish

B – EU candidate languages and Free Trade Partners: Albanian, Basque, Catalan, Galician, Icelandic, Norwegian, Scottish Gaelic, Welsh, Serbian, Turkish, Ukrainian

C – Languages spoken by EU immigrants; languages of political and trade partners: Afrikaans, Arabic, Berber, Cebuano, Chinese, Hebrew, Hindi/Urdu, Indonesian, Japanese, Korean, Kurdish, Latin, Malay, Pashto, Persian (Farsi), Russian, Tamil, Vietnamese



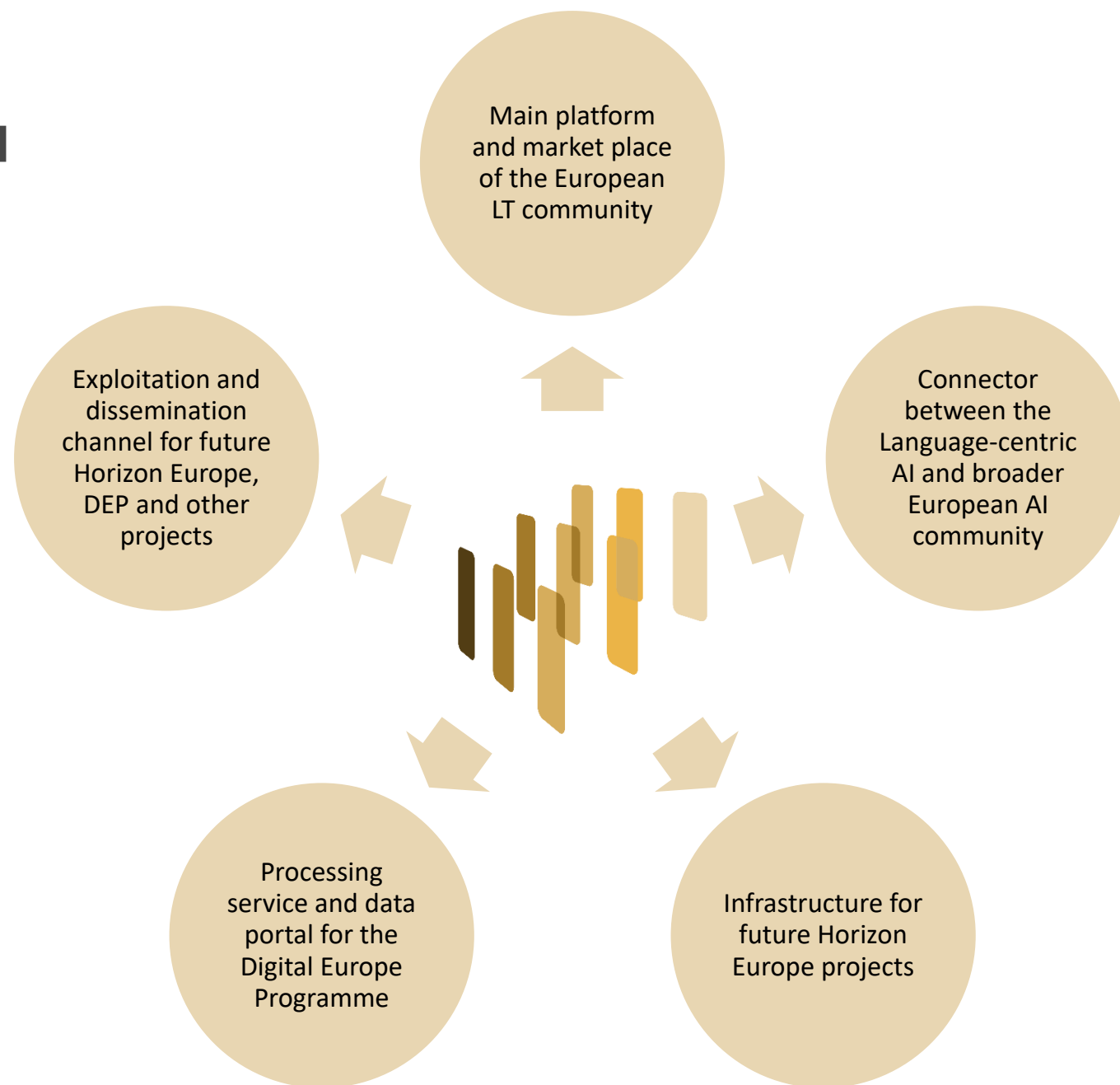
EUROPEAN LANGUAGE GRID

Concept and Methodology

- **Grid Platform**
- **Grid Content**
- **Grid Community**
 - Stakeholders: LT providers and LT consumers (buyer companies), including research centres, universities, administrations, NGOs etc.
 - Plus: 15-20 pilot projects, the six ICT-29(b) RIA projects, META-NET, LT Innovate and other networks and initiatives
 - 32 NCCs as a strong international network that act as national bridges to identify content, interest companies in making available their services
 - European LT Council as an international, pan-European body, in which LT-related matters can be discussed and coordinated.
 - Conferences, training events, presentations, social media, blog posts etc.



EUROPEAN LANGUAGE GRID



Looking forward...



EUROPEAN LANGUAGE GRID

More information on ELG

- Read the overview of ELG, history and context [here](#).
- Find out more about the European Language Grid in the document [Overview of the ELG project and platform](#).



**EUROPEAN
LANGUAGE
GRID**

ELG Grid Platform

ELG platform

- It is web-based: <https://live.european-language-grid.eu>
- It is a catalogue for resources:
 - Each resource is documented with the required metadata
 - “Published” resources are visible by anyone; i.e. log-in is not required
 - Anyone can become a **provider** and contribute a resource
 - A validation process is required for publication

ELG platform – Use catalogue

The screenshot shows the ELG platform catalogue interface. At the top, there is a navigation bar with the ELG logo and the text 'EUROPEAN LANGUAGE GRID RELEASE 1'. The navigation menu includes 'Technologies', 'Resources', 'Community', 'Events', 'Documentation', and 'About ELG'. Below the navigation bar is a search bar with the placeholder text 'Search for services, tools, datasets, organizations...' and a 'Search' button. On the left side, there is a sidebar with a 'Clear all filters' button and several filter categories: 'Language resources & technologies' (with sub-items: Corpus (599), Tool/Service (164), Lexical/Conceptual resource (148), Language description (7)), 'Languages' (with sub-items: English (512), Czech (222), German (142), French (119), Polish (110), Show more), 'Service functions' (with sub-items: Part-of-Speech Tagging (72), Tokenization (67), Dependency parsing (60), Lemmatization (60), Morphological annotation (60), Show more), and 'Licences' (with sub-item: Creative Commons Attribution 4.0 International (141)). The main content area displays search results for 'al wassit' Arabic dictionary. The first result is titled '"Al wassit" Arabic dictionary' and is categorized as a 'Lexical/Conceptual resource'. It has 81 views and 0 downloads. The second result is titled '"Al wassit" LMF Arabic dictionary' and is also categorized as a 'Lexical/Conceptual resource'. It has 10 views and 0 downloads. The third result is titled '2006 CoNLL Shared Task - Ten Languages' and is categorized as a 'Corpus'. It has 192 views and 14 downloads. The search results are displayed in a list format with details for each item, including a description, keyword, language, and licence.

Resource types

- Text search
- Facets

ELG Platform - Use catalogue

Corpus



Technologies Resources Community Events Documentation About ELG



A Small Dataset for English-to-Czech Speech Translation in the Travel Domain

Version: undefined

Corpus

Overview

Download

This small dataset contains 3 speech corpora collected using the Alex Translate telephone service (<https://ufal.mff.cuni.cz/alex#alex-translate>). The "part1" and "part2" corpora contain English speech with transcriptions and Czech translations. These recordings were collected from users of the service. Part 1 contains earlier recordings, filtered to include only clean speech; Part 2 contains later recordings with no filtering applied. The "ctest" corpus contains recordings of artificially created sentences, each ... [Read More](#)

Keyword

speech corpus ASR
machine translation

Corpus subclass

unspecified

Corpus part

AUDIO	Linguality type bilingual
	Multilinguality type unspecified
	Language Czech English

Resource provider



Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL)

Funded by

Vývoj metod pro návrh statistických mluvených dialogových systémů

QT21: Quality Translation 21

Additional information

[Landing page](#)

Log-in is not required

ELG Platform – Use catalogue

Functional service



RELEASE 1

Administration Upload Technologies Resources Community Events Documentation About ELG

Dimitris



GATE: English Named Entity Recognizer annie-named-entity-recognizer

Version: v8.6

ToolService

Overview Download/Run **Test/Try out** Code samples

Identify names of `persons`, `locations`, `organizations`, as well as `money amounts`, `time and date expressions` in English texts automatically.

Keyword

Named Entity Recognition English

Intended application

Named Entity Recognition

Resource provider

GATE Team, University of Sheffield
[Website](#)

Additional information

[Landing page](#)

Input content resource

Language
English

Data format
JSON

Processing resource type
file

more

Function

Function
Named Entity Recognition

Language dependent
true

Output resource

Language
English

Data format
JSON

Processing resource type
file

more

Log-in is required



GATE: English Named Entity Recognizer annie-named-entity-recognizer

Version: v8.6

Overview Download/Run **Test/Try out** Code samples

Type text to annotate

John lives in England.

SUBMIT

Processing

Overview Download/Run **Test/Try out** Code samples

John lives in England.

Location

Name	Value
kind	locName
rule	InLoc1
locType	province
ruleFinal	LocFinal

BACK

Annotations

- Location
- Person

ELG Platform - APIs

- All integrated functional LT services can be called from external clients (e.g. command line) via a common public API (Public LT API)
- E.g. <https://live.european-language-grid.eu/execution/processText/{ltServiceID}> for services that consume text
- The same API is also used by the ELG web-based UI
- Documentation: https://european-language-grid.readthedocs.io/en/release1.1.0/all/A2_API/LTPublicAPI.html
- A REST API will also be made available for the catalogue (search, download, upload metadata)
- A python-based command line client for the ELG platform is under development

ELG Platform – Provide to catalogue

- Outline
 - Register to the platform (create account)
 - You have to be assigned the “provider” role (send email to contact@european-language-grid.eu)
 - Register/Upload a resource
 - The resource undergoes the validation process
 - Finally the resource is published

ELG Platform – Register a resource

Dashboard

Clear all filters

Language resources & technologies **469 search results**

+ Corpus (249)



Option B: Use metadata editor

Dimitris

Organization
Companies, research organizations, academic organizations, etc. active in Language Technology in Europe.
[Go to form](#)

Project
Projects that have funded the development or used language resources and technologies.
[Go to form](#)

Corpus
Structured collections of pieces of data, such as collections of text documents, audio transcripts, audio and video recordings, parallel corpora, linguistically annotated corpora, treebanks, etc.
[Go to form](#)

Service or Tool
Services that run in the cloud, downloadable tools, source code, etc., that perform language processing and/or any Language Technology related operation, such as Machine Translation, Information extraction, linguistic annotation, automatic speech recognition, etc.
[Go to form](#)

Upload metadata record
Create item with upload of metadata file(s)
[Upload](#)

Option A: Upload a metadata record



Metadata editor is still under construction (beta)

Dimitris

Create a new Service or Tool

< Go to item type selection

LANGUAGE RESOURCE/TECHNOLOGY | TOOLSERVICE | DISTRIBUTION

Under construction
 Functional service

[Save](#)

IDENTITY
Resource name *
ILSP Named Entity Recognizer
language: English

CATEGORIES
Language Resource identifier
A string (e.g., PID, DOI, internal to an organization , etc.) used to uniquely identify a language resource
[Fill in](#)

CONTACT
Resource short name
ILSP NERC
language: English

DOCUMENTATION
Description *
This is a tool for ...
language: English

RELATED LRTS

Is this a functional service?

Yes
 No



ELG platform – validation process of an LT tool

- The LT Service is assigned to a set of “validators”
 - Technical, legal, metadata
- The LT tool is deployed to the ELG platform (@kubernetes)
- The LT Service is tested by the technical validator and the LT provider
 - Test via trial UIs
 - Check container logs ...
 - Troubleshoot ...
- When all validators approve the tool it is published to the catalogue.

Provide a functional LT service

- ELG supports the integration/registration of tools/services
- Currently, a tool must fall into one of the following broad categories:
 - Information Extraction (IE)
 - Text Classification (TC)
 - Machine Translation (MT)
 - Automatic Speech Recognition (ASR)
 - Text to Speech Generation (TTS)
- Documentation: <https://european-language-grid.readthedocs.io/en/release1.0.0/all/RegisterFunc.html>
- Requirements: Follow ELG JSON format, Dockerize

Provide non-functional content

- Non-functional content
 - Non-functional services
 - Lexical/conceptual resources,
 - Language descriptions
 - Datasets/Corpora
 - Projects, Organizations
- The process is similar to functional content
 - Documentation: <https://european-language-grid.readthedocs.io/>



**EUROPEAN
LANGUAGE
GRID**

Open Call for Pilot Projects

Open Call

- Call documentation
 - Annex 1 Guide for Applicants
 - Annex 2 Third Party Agreement
 - Annex 3 Project Proposal Template
 - Annex 4 Evaluation Criteria



Open Call



- **Goals of the European Language Grid Open Call for pilot projects**

- The pilot projects should demonstrate the usefulness of the ELG as a technology platform, by:
 - Broadening ELG's portfolio of Language Technologies (adding data, tools, services)
 - Developing solutions/applications using the ELG
- There are no priorities: best projects will be selected that fulfill the above, with your help

- **Outcome of the Pilot Projects**

- (a) Contribute resources, services, tools or data sets to the ELG to increase its coverage, or
- (b) Develop applications using language resources and technologies available in the ELG.

- **Key parameters**

- 50,000-200,000€, no co-financing, 9-12 month duration, SME & Research organizations, NGOs eligible
- project type "(b) develop applications using language technologies available in the grid" will get 30 bonus points in case the applicant is an SME

Open Call



- **Evaluation criteria**

- Objective fit (weight: 3)
- Technical approach (weight: 2)
- Business, Integration and Dissemination (“BID”) plan (weight: 3)
- Budget adequacy (weight: 1)
- Team (weight: 1)

- **Eligibility criteria**

- **Uniqueness:** No similar project, technology or application exists. Non-existence of the same proposal selected by ELG in the previous call.
- **Relevance for ELG:** Match of the actual proposal to the objective (a) or (b).
- **Project Phases:** Proposal includes all required phases (Experiment, Integration, Dissemination)

Open Call



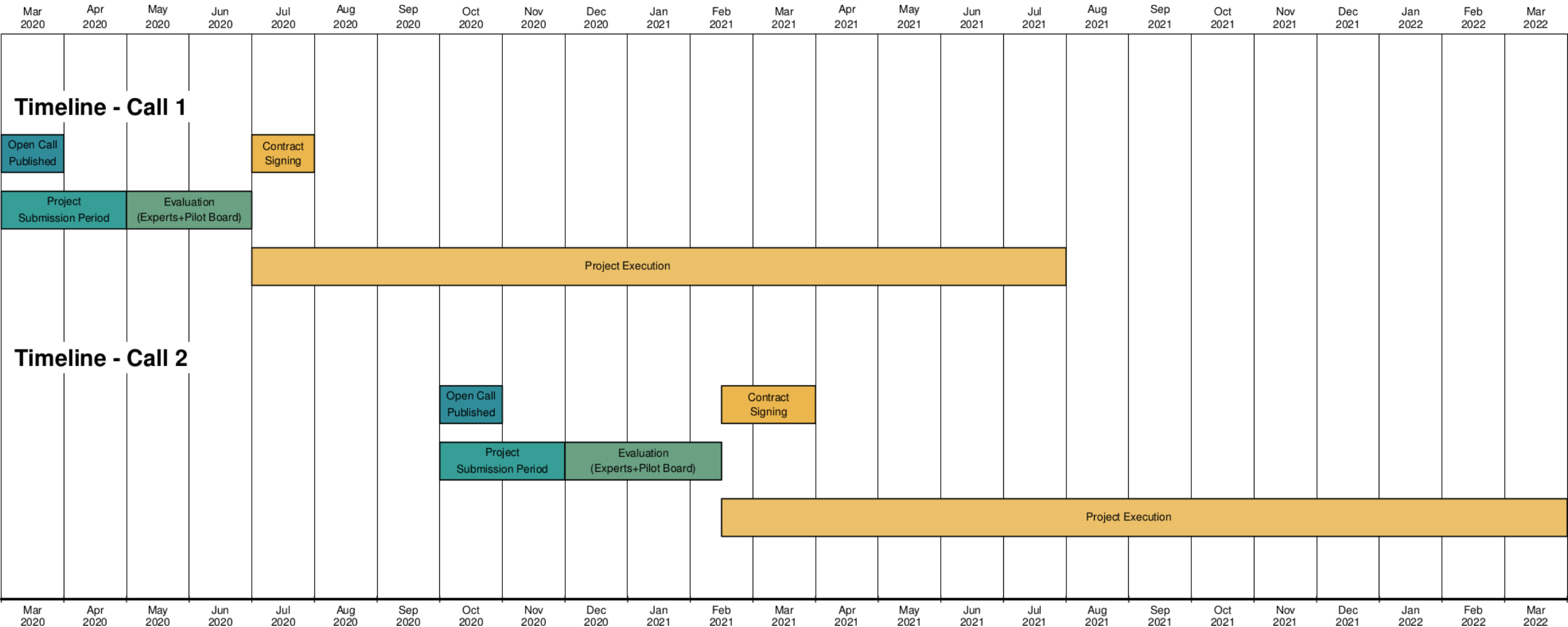
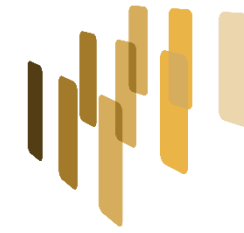
- **Schedule – second open call**

- published: October 2020
- project submission period: October – November 2020
- project start: February 2021
- project duration: 9 – 12 months

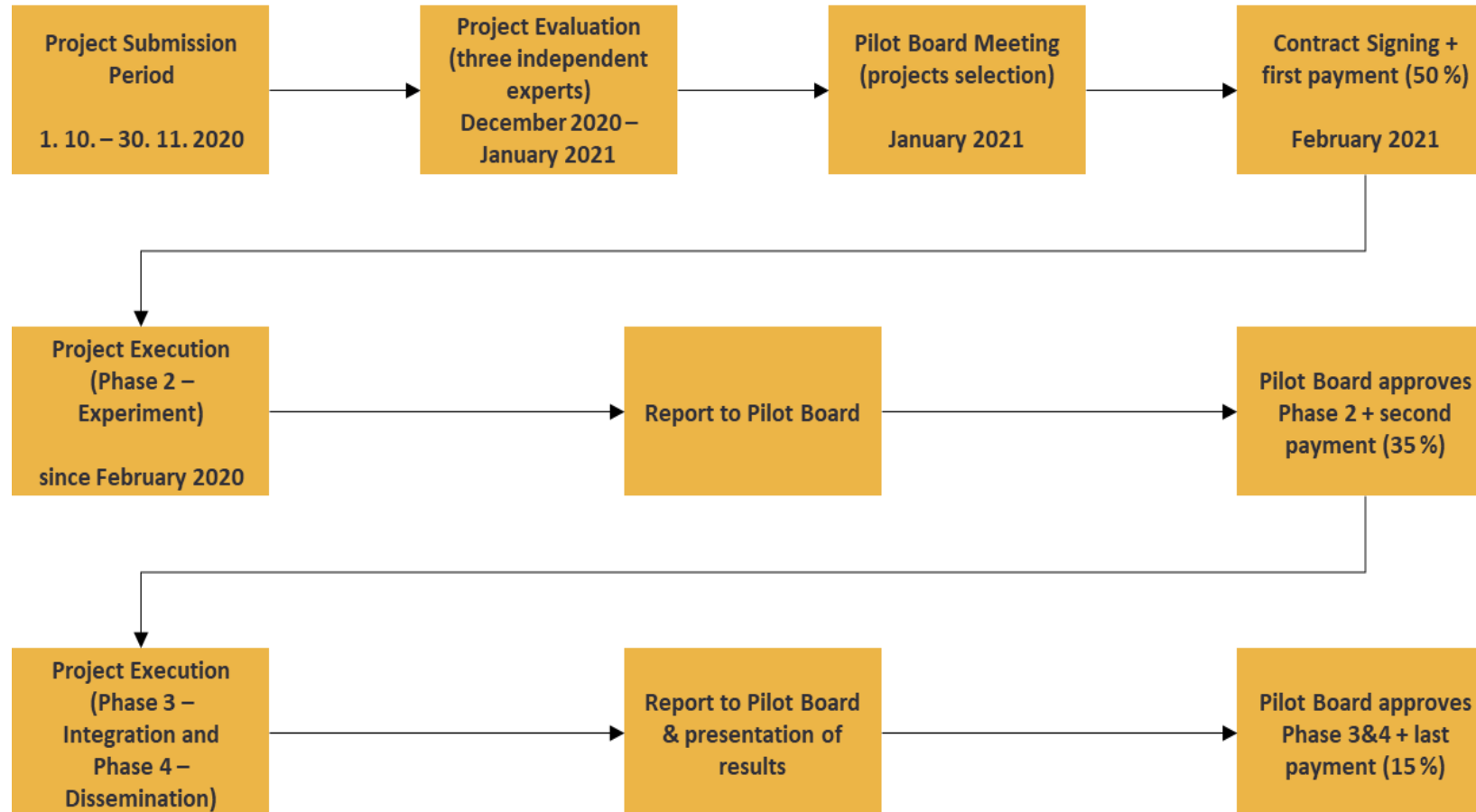
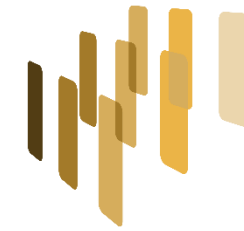
- **Funding Principle**

- Lump Sum in three payments:
 1. 50 % after signing the contract
 2. 35 % after demonstrating the results after the end of activity Experiment
 3. 15 % after completing the activities Integration and Dissemination

Open Calls: Timeline



Project execution scheme



Open Call



- **Proposals are to be submitted via the ELG open calls platform**
 - <https://opencalls.european-language-grid.eu/user>
 - For more information on the proposal submission see Annex 1 – Guide for Applicants.



**EUROPEAN
LANGUAGE
GRID**

Q & A

Q&A

- Q: How much money is allocated for pilot projects?
 - A: Approx. €1,950,000 to be distributed in total in two open calls. There is € 585,000 left for the second call.
- Q: How many calls will you open?
 - A: There are two open calls. First was opened in March 2020, this is the second and last call. We expect that 3-6 projects to be selected in Call 2, up to the maximum funding available.
- Q: Can I receive internal feedback on my project proposal?
 - A: All project proposals are going to be evaluated by three independent evaluators and the Pilot Board using the evaluation criteria specified in the Call Documentation. Please note that we cannot help you with the content of your project as we need to maintain equal access to information for all applicants (i.e., no individual consultations are possible).

Q&A

- Q: Can we submit more than one proposal?
 - A: Yes. Up to two proposals per applicant will be accepted for evaluation, one for objective A and one for objective B (see Call Documentation). Others will be discarded. If the project proposal is not accepted in the first call, resubmission to the second call is allowed.
 - Please check within your organization so that you coordinate with other teams that might be interested in this call.

- Q: Can we collaborate with partners in our project?
 - A: No. You are allowed to apply as an individual project only, i.e. one organisation per project only.

- Q: Are there any priority topics in the open call?
 - A: There are no priorities. We want to select the best projects based on the evaluation of three independent experts and the Pilot Board.

Q&A

- Q: For a type A project, how are we supposed to share our services?
 - A: You will share your services via ELG (see examples in the grid). For type A projects (extending the ELG capabilities), the resource, tool or service has to be integrated into ELG. A license will be signed between ELG and the Pilot Project to specify the terms of use. It must be available at least as a demo and/or for research purposes for free. Any commercial use will either be specified in the license agreement, or it will be postponed until the ELG will have an established model for full commercial exploitation.
 - Regarding the IP rights – please see the Third Party Agreement Model (<https://www.european-languagegrid.eu/open-calls/open-call-2/>), section 6.
- Q: For a type B project, how are we supposed to set up the business model?
 - A: For type B projects (developing an application using the existing ELG resources, tools or services), the resulting application must be described in the ELG catalogue, pointing to a place where it is available (e.g., a company website), and which acknowledges the ELG Open Call funding. The integration of the resulting service is also possible. A service level agreement negotiated during the execution of the project will be signed between the Third Party and ELG.

Q&A

- Q: Is my project type A or type B?
 - A: Sometimes there is no sharp line between the two. We recommend to describe and explain your project idea well in your proposal. The categorization of your project might be changed during the evaluation process. This does not have any impact on the chance of acceptance. Please see the Call Documentation for more information.

- Q: What is meant by “results of your project must be integrated into the ELG”?
 - A: Your resulting services, tools and resources will be shared via the ELG platform. The ELG Technical Team will guide you through the integration process. Also, please check the [ELG user manual](#).



**EUROPEAN
LANGUAGE
GRID**

pilot-projects@european-language-grid.eu