



EUROPEAN LANGUAGE GRID



European Language Grid and European Language Equality

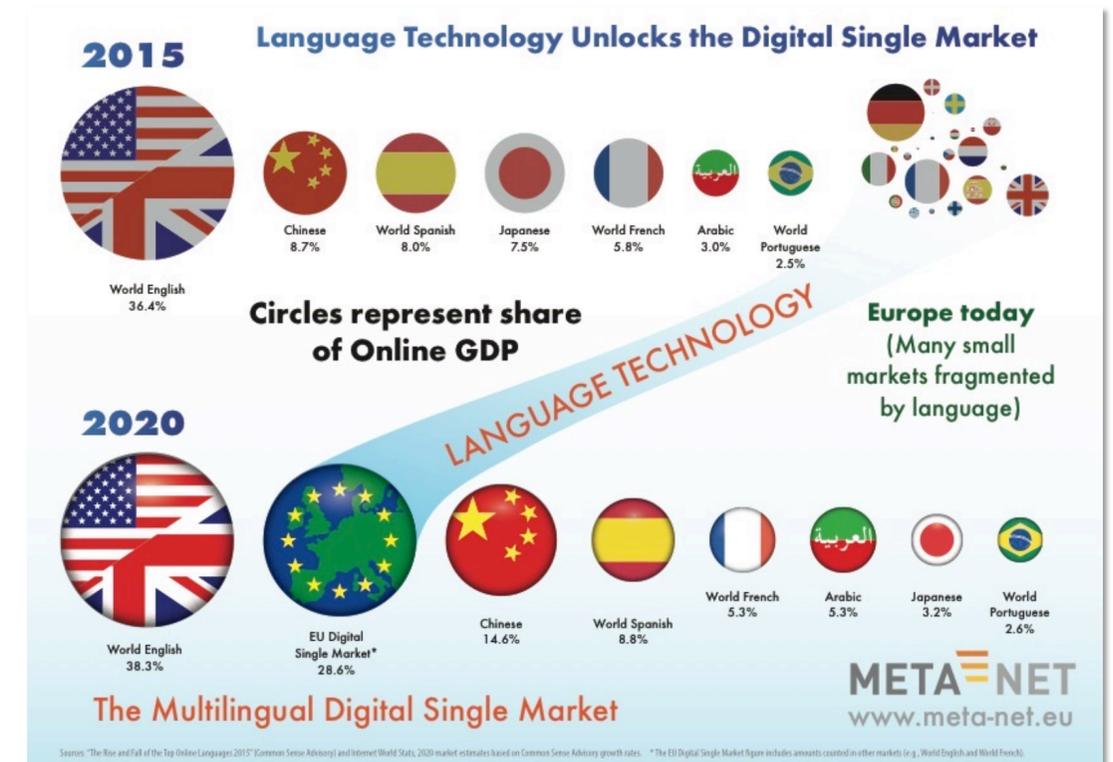
Prof. Dr. Georg Rehm (DFKI) – Coordinator ELG, Co-Coordinator ELE

24-03-2022 ELG Workshop Romania

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

Point of Departure: Multilingualism in Europe

- Multilingualism is at the heart of the European idea
- 24 official EU languages – they all have the same status
- Dozens of co-official, regional and minority languages as well as languages of immigrants and trade partners
- Many economic, social and technical challenges
 - The Digital Single Market needs to be multilingual
 - Cross-border, cross-lingual, cross-cultural communication
 - Fragmentation of the LT market and landscape





META FORUM 2015



META FORUM 2017



META FORUM 2010



META SHARE

T4ME
EU-funded project (Seventh Framework Programme) working on technologies for the Multilingual European Information Society



META-NET
Established in 2010, META-NET is a network of Excellence consisting of 60 research centres from 34 countries, building the technological foundations of a multilingual European information society

META FORUM 2012

META-FORUM 2012
Conference "A Strategy for Multilingual Europe" (Brussels – June 20/21, 2012)

META-NET White Papers
Release of 32 volumes on 31 languages, revealing that there is a severe threat of digital extinction for at least 21 European languages (December 2012)

CRACKER
EU-funded project CRACKER (Horizon2020) pushing towards an improvement of MT research in terms of efficiency and effectiveness (2015 – 2017)

Cracking the Language Barrier Federation
Founded in 2015, the federation has been assembling European research and innovation projects as well as all related community organisations working on multilingual technologies

META-FORUM 2015
Conference "Technologies for the Multilingual Digital Single Market" (Riga – April 27, 2015)

Strategic Agenda for the Multilingual Digital Single Market (Version 0.5)
Launch of the Strategic Agenda for the Multilingual Digital Single Market titled "Technologies for Overcoming Language Barriers towards a truly integrated European Online Market" (April 2015)

Riga Summit on the Multilingual Digital Single Market
Summit "Shape the future of the multilingual digital single market" (April 27–29, 2015)

"Language Equality in the Digital Age"
Workshop on "Language Equality in the Digital Age", commissioned by the EU Parliament's Science and Technology Options Assessment Committee (STOA) (January 2017)

"Language equality in the digital age: Towards a Human Language Project"
Launch of the study on "Language equality in the digital age: Towards a Human Language Project", commissioned by the EU Parliament (March 2017)

META-FORUM 2017
Conference "Towards a Human Language Project" (Brussels – November 13/14, 2017)

Strategic Research and Innovation Agenda (V1.0)
Launch of the Strategic Research and Innovation Agenda titled "Language Technologies for Multilingual Europe – Towards a Human Language Project" (December 2017)



2010

META-FORUM 2010
"Challenges for Multilingual Europe" (November 17/18, 2010)

2011

META-NORD
EU-funded (ICT PSP) Project to establish an open linguistic infrastructure in the Baltic and Nordic countries (2011 – 2013)



METANET4U
EU-funded project (ICT Policy Support Programme) to enhance the European Linguistic Infrastructure (2011 – 2013)



CESAR
EU-funded project (ICT Policy Support Programme) functioning as a part of META-NET to standardise language resources and tools (2011 – 2013)



META-FORUM 2011
Conference "Solutions for Multilingual Europe" (Budapest – June 27/28, 2011)

META FORUM 2011

2012

"State of the Art of Machine Translation – Current Challenges and Future Opportunities"
Workshop on "State of the Art of Machine Translation", commissioned by the EU Parliament (December 2013)

2014

META-SHARE
Initiated in 2013, META-SHARE has functioned as an open and secure network of repositories for sharing and exchanging language data, tools and services

Strategic Research Agenda for Multilingual Europe 2020
Launch of the Strategic Research Agenda for Multilingual Europe 2020 (January 2013)

META-FORUM 2013
Conference "Connecting Europe for New Horizons" (Berlin – September 19/20, 2013)

META FORUM 2013

Strategic Research and Innovation Agenda (Version 0.9)
Launch of the Strategic Research and Innovation Agenda titled "Language as a Data Type and Key Challenge for Big Data" (July 2016)

META-FORUM 2016
Conference "Beyond Multilingual Europe" (Lisbon – July 4/5, 2016)



META FORUM 2016

2016

ELG Final Proposal Submission
Final submission on Feb. 20, 2018

2018



EUROPEAN LANGUAGE GRID

2019-2022



EUROPEAN LANGUAGE EQUALITY

2021-2022

ELG European Language Grid

ELG consortium



European Parliament Resolution (2018)

EP Resolution “Language equality in the digital age”
P8_TA(2018)0332 – partially based on the STOA study

Voting (11 Sept. 2018): **592 yes** – 45 no

Recommendations addressed by European Language Grid:

29. Create a European LT platform for sharing of services

41. Enable and empower European SMEs to use LTs

26. ICT integrators should be given economic incentives for LT

27. Europe has to secure its leadership in language-centric AI

32. Set up LT financing platform; emphasise R&D in Deep NLU

40. Develop investment instruments and accelerator programs

European Parliament

2014-2019



TEXTS ADOPTED
Provisional edition

P8_TA-PROV(2018)0332

Language equality in the digital age

European Parliament resolution of 11 September 2018 on language equality in the digital age (2018/2028(INI))

The European Parliament,

- having regard to Articles 2 and 3(3) of the Treaty on the Functioning of the European Union (TFEU),
- having regard to Articles 21(1) and 22 of the Charter of Fundamental Rights of the European Union,
- having regard to the 2003 UNESCO Convention for the Safeguarding of the Intangible Cultural Heritage,
- having regard to Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information¹,
- having regard to Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information²,
- having regard to Decision (EU) 2015/2240 of the European Parliament and of the Council of 25 November 2015 establishing a programme on interoperability solutions and common frameworks for European public administrations, businesses and citizens (ISA2 programme) as a means for modernising the public sector³,
- having regard to the Council resolution of 21 November 2008 on a European strategy for multilingualism (2008/C 320/01)⁴,
- having regard to the Council decision of 3 December 2013 establishing the specific programme implementing Horizon 2020 – the Framework Programme for Research and

¹ OJ L 345, 31.12.2003, p. 90.

² OJ L 175, 27.6.2013, p. 1.

³ OJ L 318, 4.12.2015, p. 1.

⁴ OJ C 320, 16.12.2008, p. 1.



EUROPEAN LANGUAGE GRID

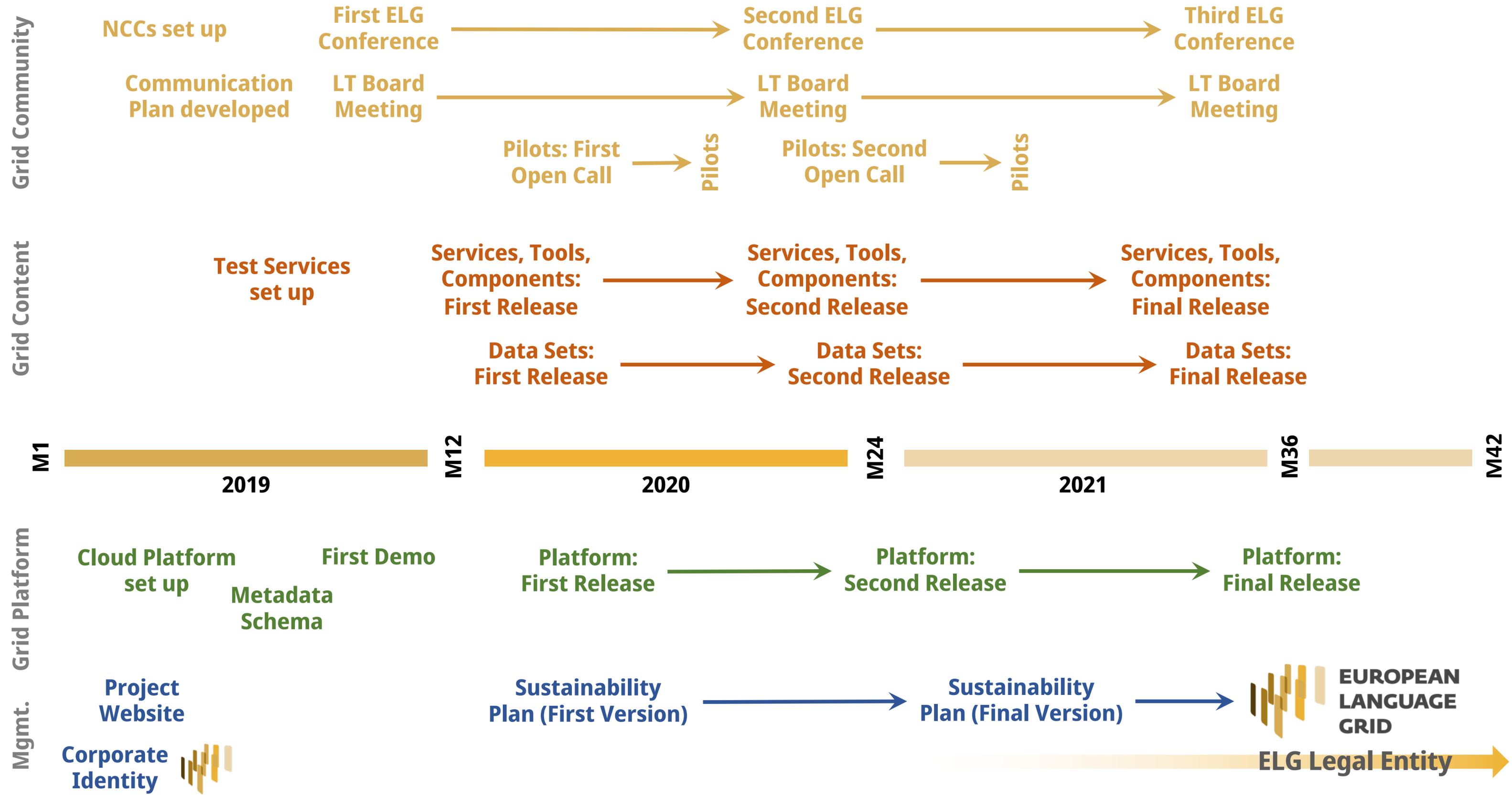
Objectives (Selection)

1. Establish the ELG as the primary Language Technology platform and market place in Europe to tackle the fragmentation of the European LT landscape.
2. ELG as a platform for commercial and non-commercial, industry-related LTs (functional and non-functional).
3. Enable the European LT community to upload services and data sets, to deploy them and to connect with, and make use of those resources made available by others.
4. Enable businesses to grow and benefit from scaling up.
5. Unleash enormous potential for innovation.

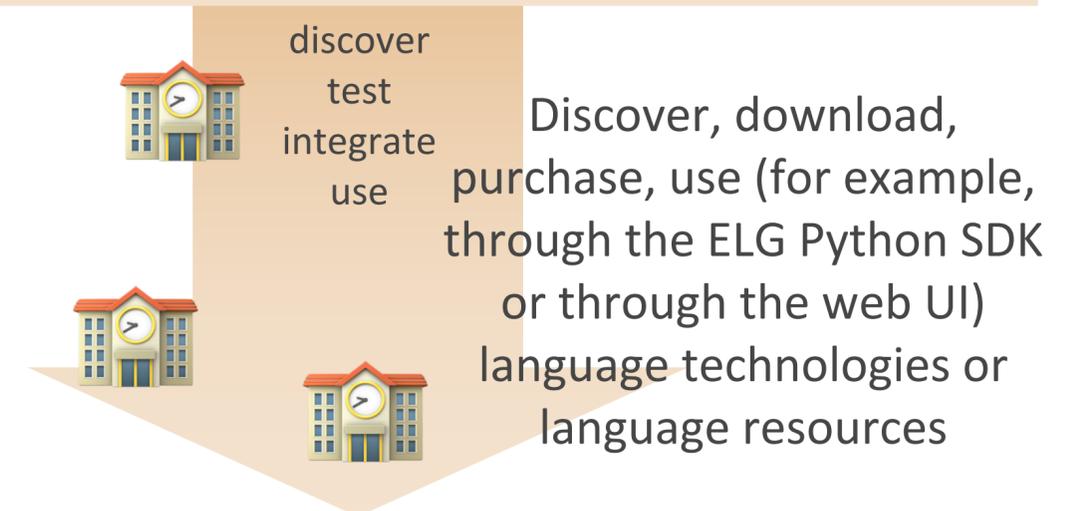
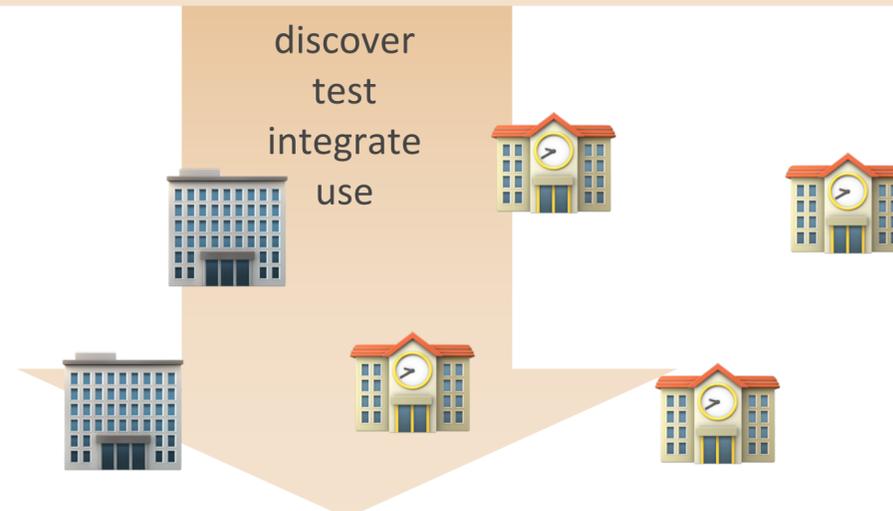
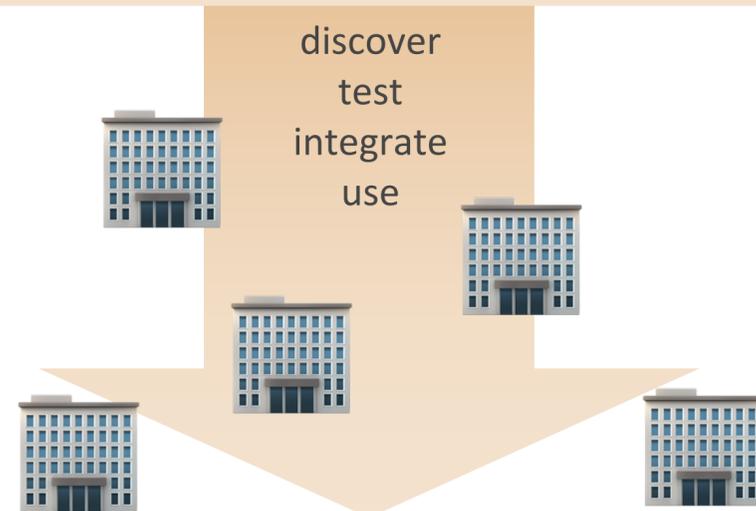
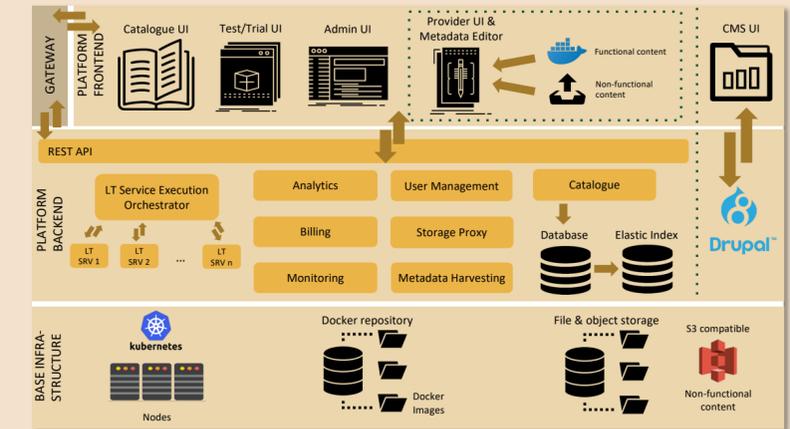
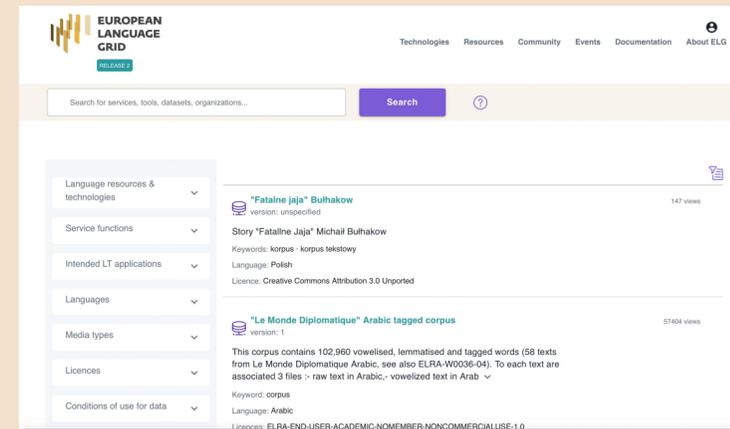
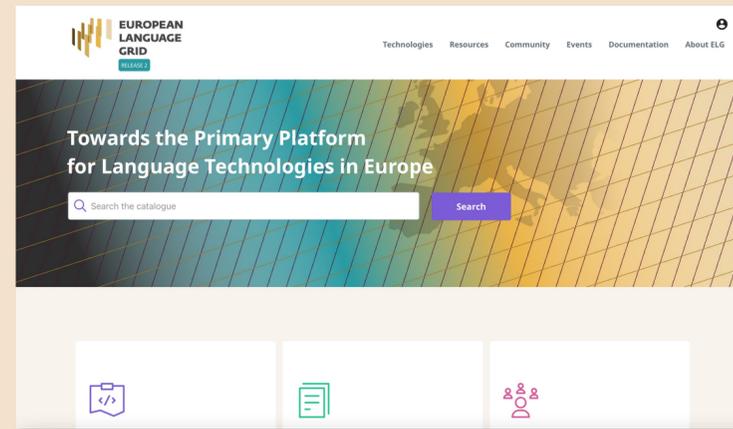
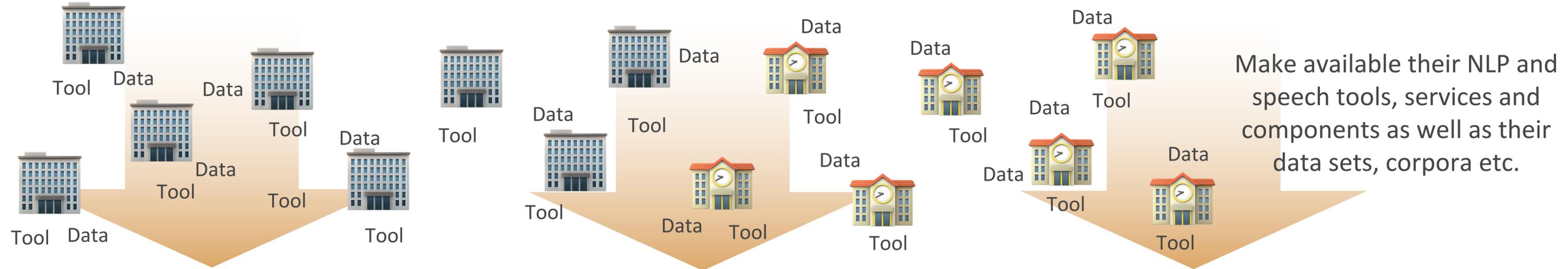


Kick-off meeting, 22/23 January 2019





Developers of Language Technologies: Companies, Universities, Research Centres (approx. 1800 organisations in total)



META-SHARE

ELRC-SHARE

OPUS

Zenodo

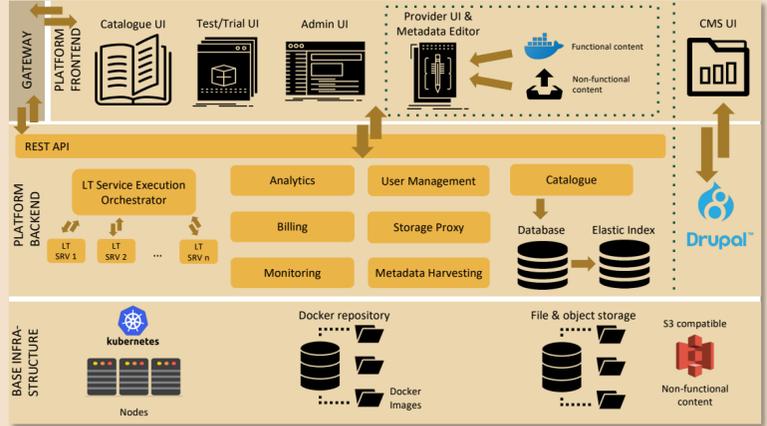
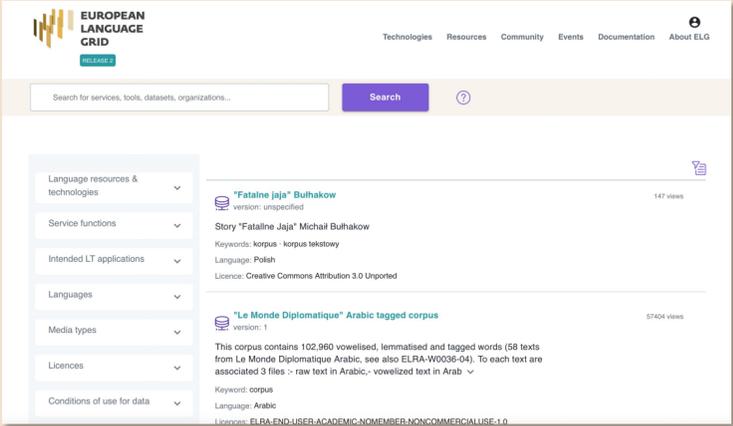
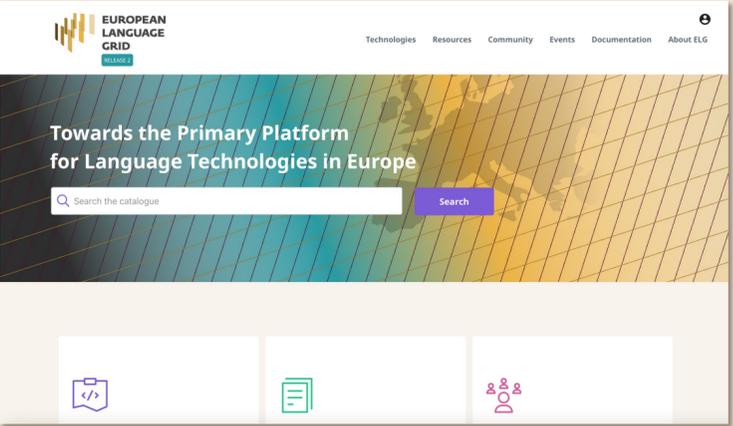
Harvest all relevant LT/LR repositories on a regular basis, collect metadata about resources and make them available through ELG for increased discovery and visibility

harvest and integrate

harvest and integrate

harvest and integrate

harvest and integrate



harvest and integrate



ELRA

HuggingFace

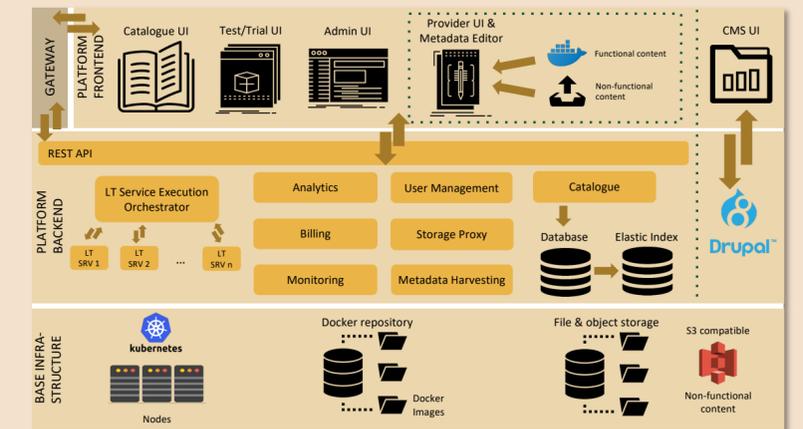
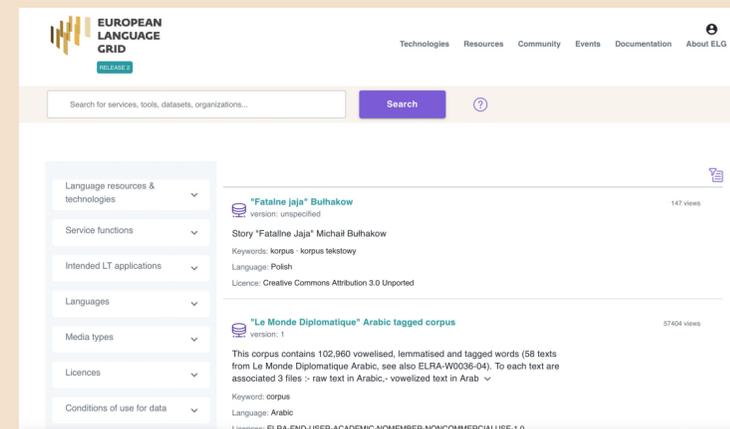
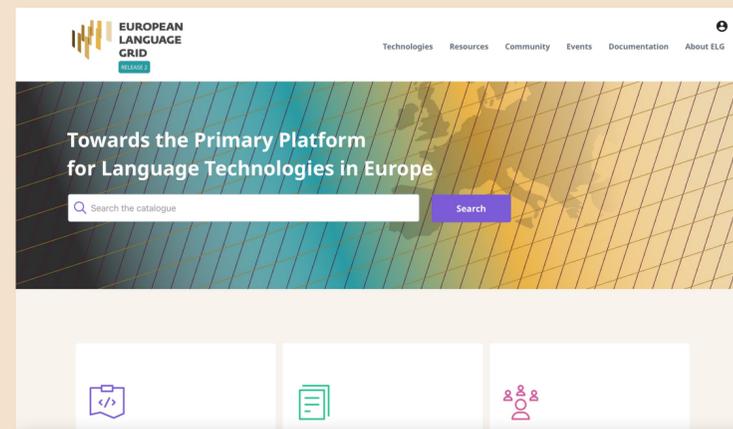
CLARIN

...

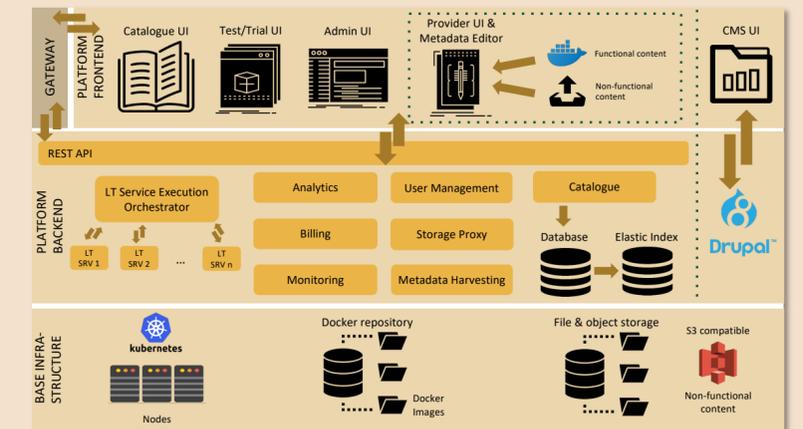
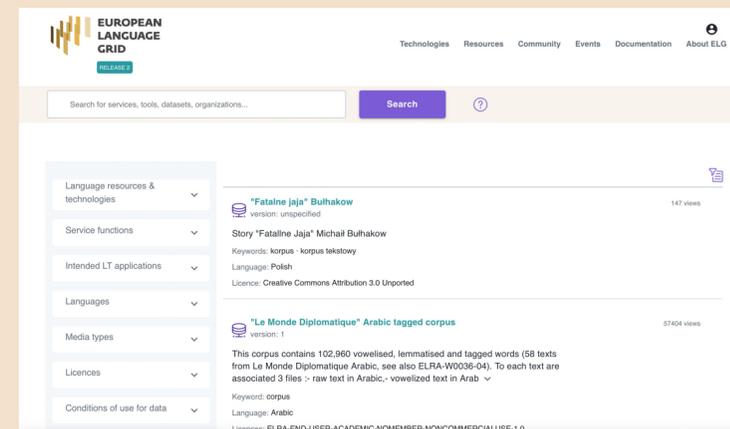
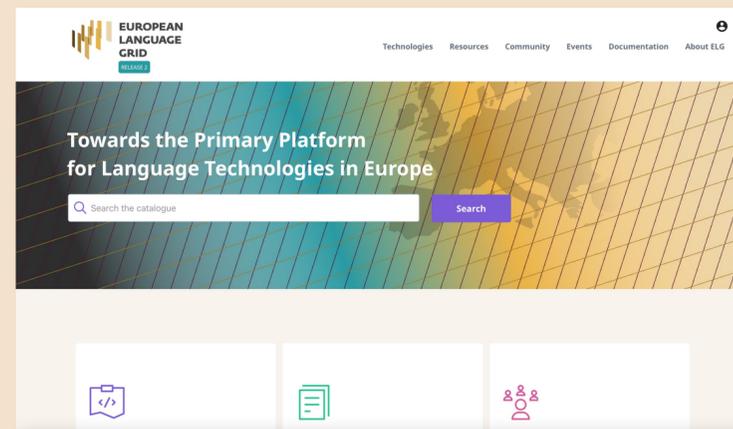
...



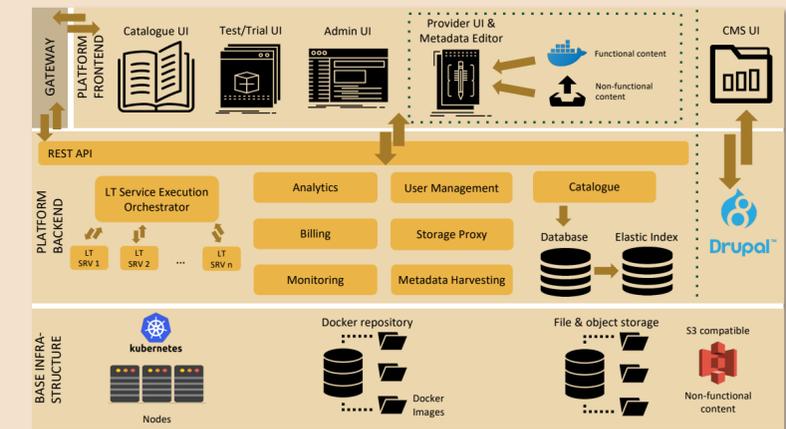
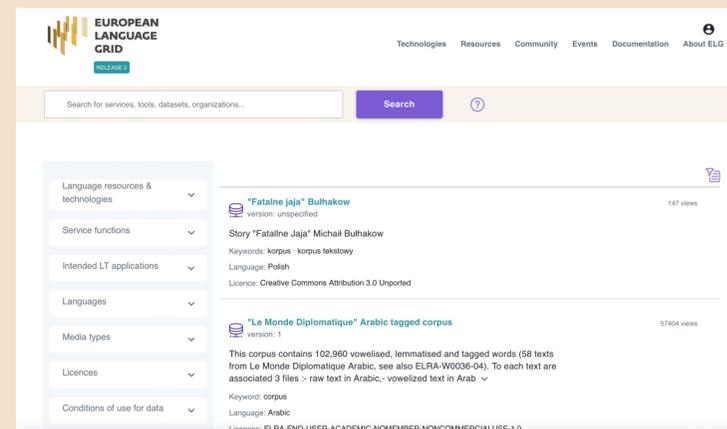
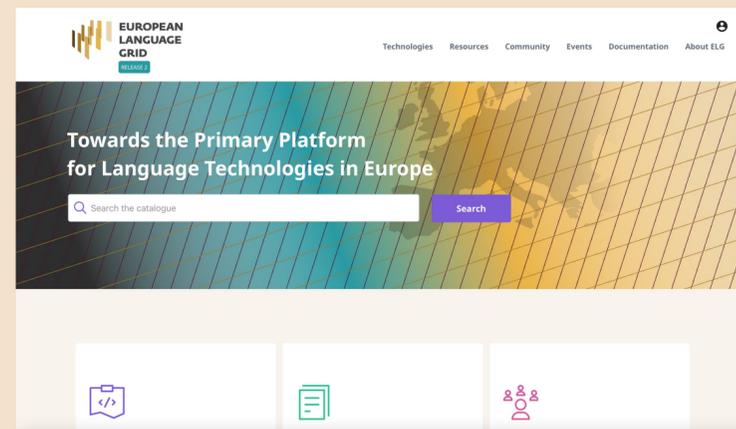
Harvesting of Language Technology-related Resource, Data, Tool and Model Repositories



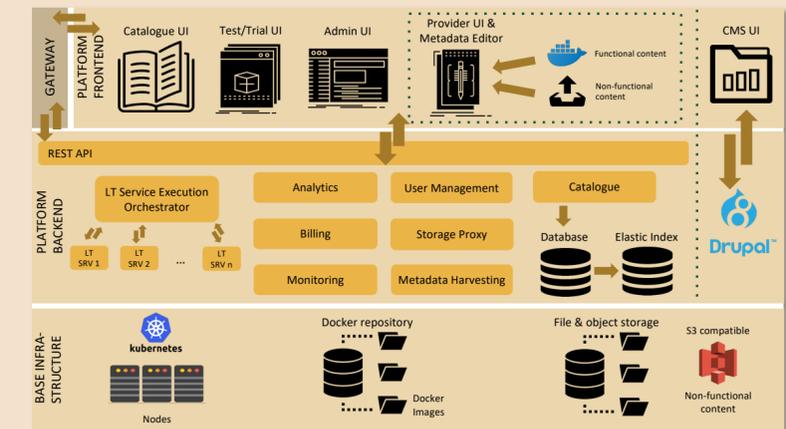
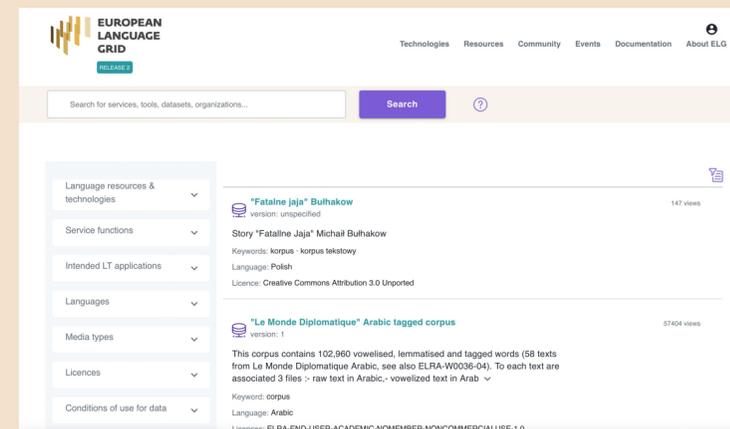
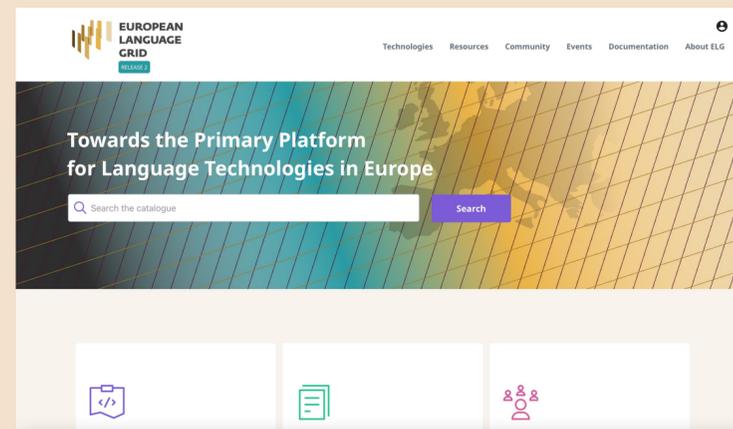
European Language Grid is a **joint technology platform** for the whole European Language Technology community (approx. 1800 organisations)



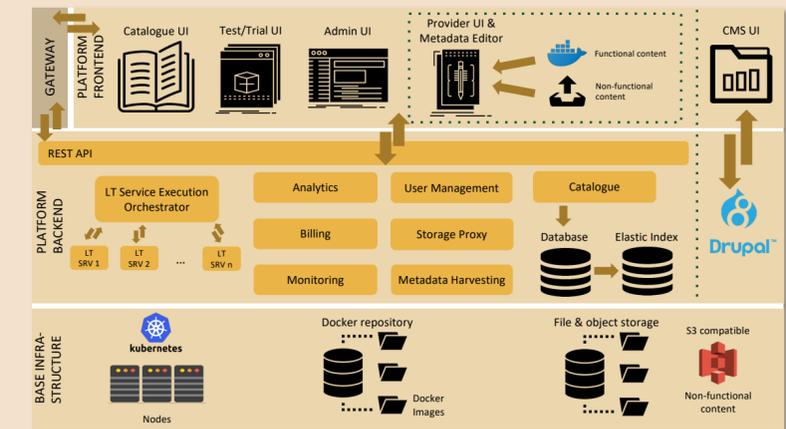
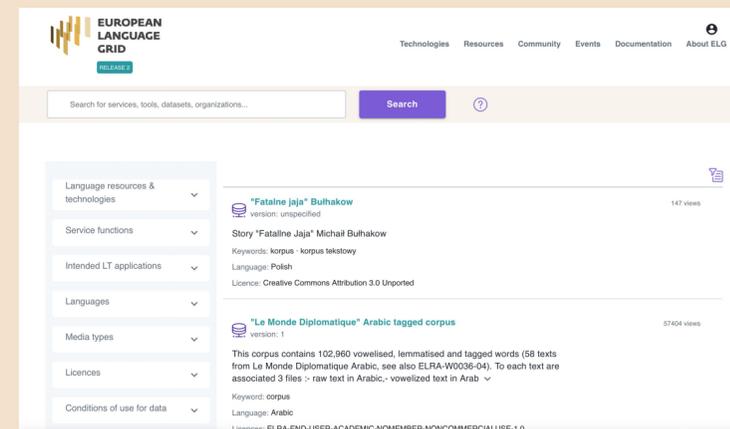
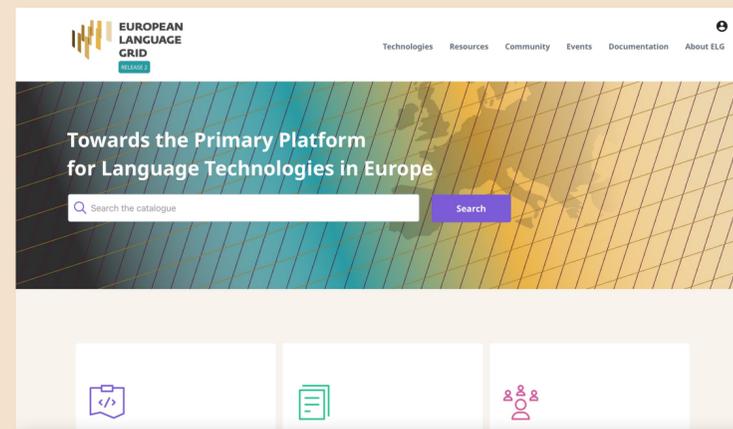
European Language Grid is a joint data, tool and resource sharing platform



European Language Grid is a **marketplace** for the whole European Language Technology community



European Language Grid is the **yellow pages** of the European Language Technology community



We want to make available all European Language Technologies and Language Resources in a one-stop-shop



EUROPEAN LANGUAGE GRID

Current State of Play (March 2022)

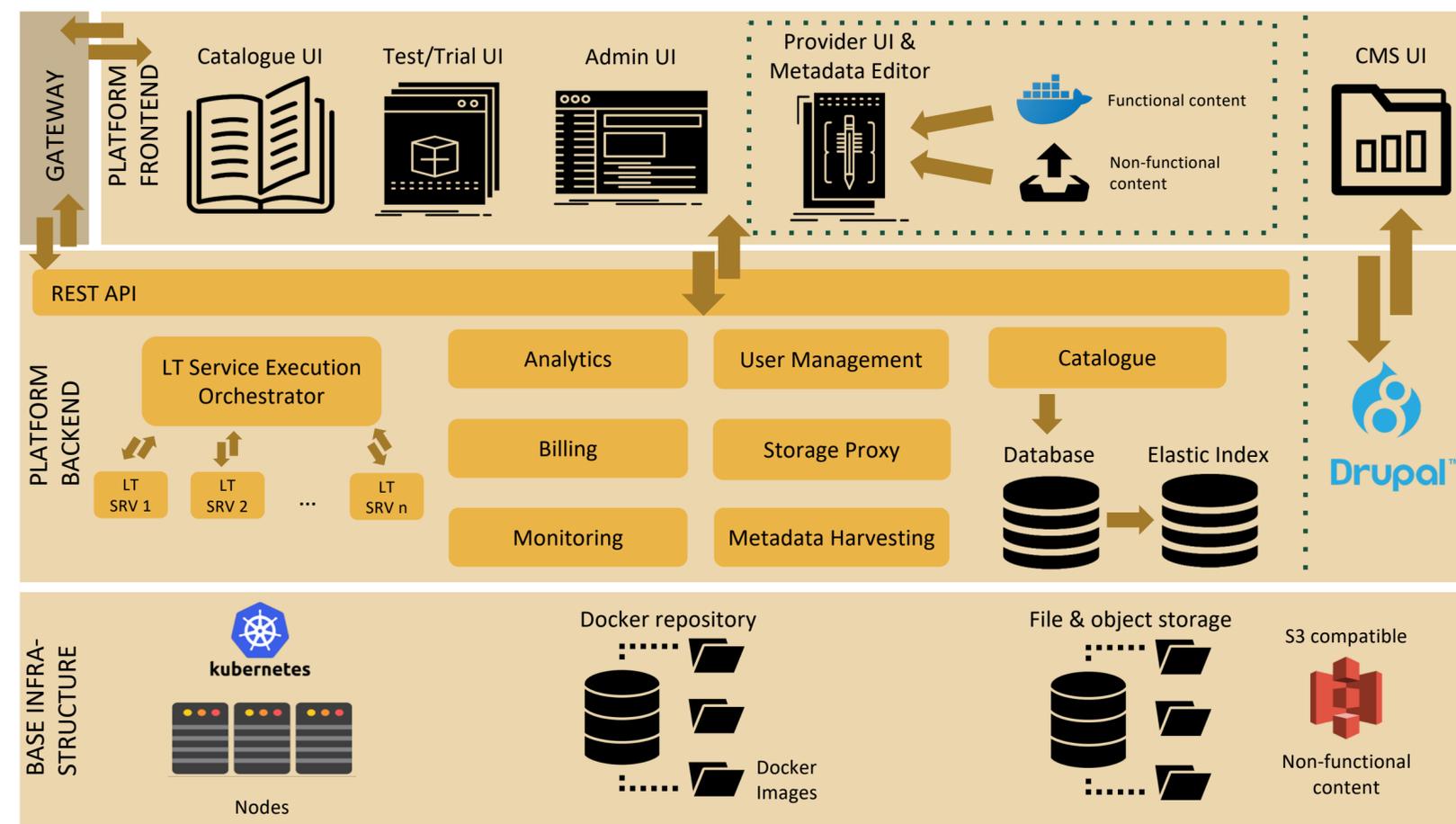
Release 2: 11,800+ Resources

- 6238 corpora and data sets
- 2957 services and tools
 - 730 functional services directly within ELG
- 2227 lexical/conceptual resources
- 407 lang. descriptions, grammars, models
- 1789 organisations (research organisations, companies)

Users can connect to the ELG cloud platform via ELG APIs, remote APIs, ELG GUI, Python SDK, download of containers or source code.

Source		Sources of data sets	
+ ELG / ELE	(8785)	+ Quantum Stat Datasets	(261)
+ ELRC-SHARE	(1299)	+ CLARIN-PL	(254)
+ ELRA Catalogue of Language Resources	(1180)	+ CLARIN.SI	(218)
+ Zenodo	(492)	+ LREC Shared LR (ELRA)	(144)
+ Hugging Face	(385)	+ META-SHARE/ILSP	(69)
+ LINDAT/CLARIAH-CZ	(353)	+ META-SHARE/DFKI	(2)
		+ TUdatalib	(1)
		+ labsai/eddi	(1)

Sources of data sets



Georg Rehm

Technologies Resources Community Events Documentation About ELG

Towards the Primary Platform for Language Technologies in Europe

Search the catalogue

Search

Language technologies
LT services, tools, components, downloadable or deployed directly through the grid

Language data and resources
The collection of data sets, corpora, language models and other language resources

Community
Organizations, projects, events etc in European Language technology field

Browsing the ELG catalogue

Georg Rehm

Technologies Resources Community Events Documentation About ELG

Towards the Primary Platform for Language Technologies in Europe

tweets written in Italiani

Search

Downloading a resource

Georg Rehm

Technologies Resources Community Events Documentation About ELG

UEDIN-MT-DeEn

Search

140 search results for UEDIN-MT-DeEn

UEDIN Machine Translation Service for German to English
1.0.0
111 views
2131 times used
ELG-compatible service

A machine translation (MT) service for German-to-English translation based on the Marian machine translation framework. The translation model is a basic transformer model trained on ca 13.3M sentence pairs using Marian N

Keywords: Machine Translation · German · English · Neural machine translation
Languages: German · English
Licence: Creative Commons Attribution Share Alike 4.0 International

UEDIN Machine Translation Service for Czech to English
1.0.0
25 views
152 times used
ELG-compatible service

An ensemble of 3 transformer-base models trained on WMT'19 data as a system for generating back-translations.

Keywords: Machine Translation · Czech · English · Neural machine translation
Languages: English · Czech
Licence: Creative Commons Attribution Share Alike 4.0 International

UEDIN Machine Translation Service for English to German
1.0.0
62 views
570 times used
ELG-compatible service

A machine translation (MT) service for English-to-German translation based on the Marian machine translation framework. The translation model is a basic transformer model trained on ca 13.3M sentence pairs using Marian N

Keywords: Machine Translation · German · English · Neural machine translation
Languages: German · English
Licence: Creative Commons Attribution Share Alike 4.0 International

Testing an MT service

Georg Rehm

Technologies Resources Community Events Documentation About ELG

Search for services, tools, datasets, organizations...

Search

conll shared task
2007 conll shared task
hosted at elg
udpipe english

"Fatalne jaja" Bulhakow
unspecified
33 views
0 downloads

Story "Fatalne Jaja" Michail Bulhakow
Keywords: korpus · korpus tekstowy
Language: Polish
Licence: Creative Commons Attribution 3.0 Unported

"Le Monde Diplomatique" Arabic tagged corpus
1
28612 views
0 downloads

This corpus contains 102,960 vowelised, lemmatised and tagged words (58 texts from Le Monde Diplomatique Arabic, see also ELRA-W0036-04). To each text are associated 3 files: - raw text in Arabic, - vowelized text in

Keyword: corpus
Language: Arabic
Licences: ELRA-VAR-COMMERCIAL-MEMBER-COMMERCIALUSE-1.0 · ELRA-END-USER-COMMERCIAL-MEMBER-NONCOMMERCIALUSE-1.0

"Le Monde Diplomatique" Text corpus in Arabic
1
116 views
0 downloads

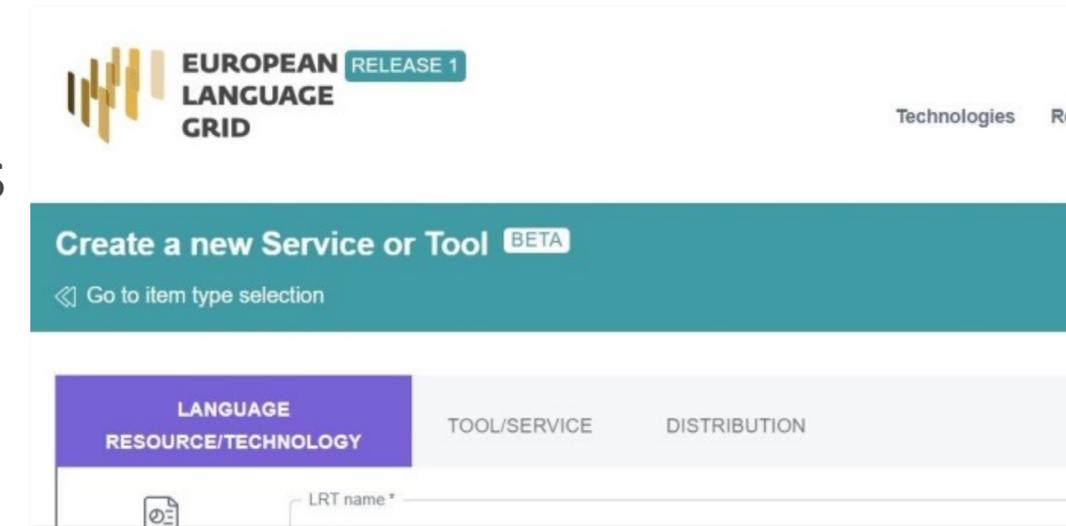
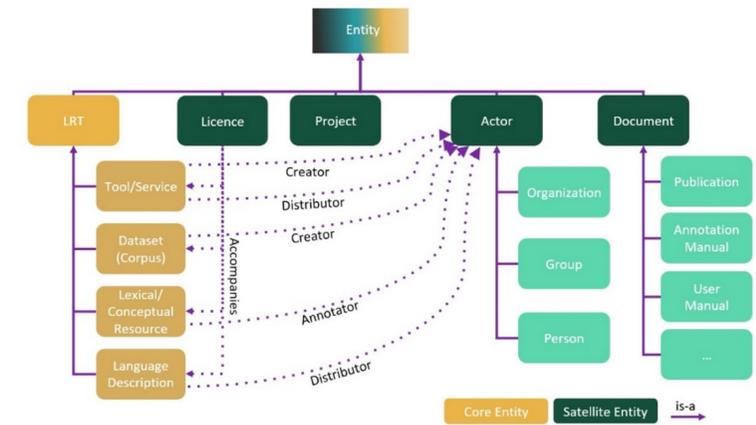
Electronic archiving of "Le Monde Diplomatique" articles in Arabic from 2000. The corpus is available in HTML. Each HTML file contains one article. Number of articles available per year: · 2000: 61 articles (November a

Keyword: corpus
Language: Arabic
Licences: ELRA-END-USER-ACADEMIC-MEMBER-NONCOMMERCIALUSE-1.0 · ELRA-END-USER-COMMERCIAL-MEMBER-NONCOMMERCIALUSE-1.0

Testing a dependency parser

European Language Grid – What can you do with it?

- **Data consumers** can search and browse the ELG catalogue
 - for different types of data, language processing services, projects and organisations
 - download data (depending on access conditions)
- **Service consumers** can try out and test language processing services
 - call a service from the command line, integrate it into workflow
 - view code samples
 - current API support: MT, IE, ASR, TTS, text classification
- **Data and service consumers** can use a Python-based API for accessing the ELG catalogue, searching and directly fetching datasets to feed them into, e.g., their model training pipeline
- **Data providers** can upload resources and register metadata descriptions using a semantic schema
 - a dashboard provides an overview of own resources including their status (*draft* → *published*)
- **Language Technology providers** can provide their service/tool as a Docker container
 - Different options: (a) Docker image in ELG, can be run from ELG; (b) Docker image in ELG as proxy talks to remote service



ELG Python SDK – Overview

- Brings main ELG functionalities to Python.
- Integrate ELG services into pipelines.
- Repository available on Gitlab:
<https://gitlab.com/european-language-grid/platform/python-client>

```
from elg import Service # `pip install elg` to install the package

lt = Service.from_id(4800) # 4800 is the id of a Polish-English translation service

# At this step, the user will need to authenticate on the ELG.

result = lt("Pandemia powoduje, ze telepraca staje sie obowiazkowa dla wielu osob.") # it is also
possible to translate a text file by passing the path

print(result)
# {'response': {'type': 'texts', 'texts': [{'content': 'The pandemic makes teleworking compulsory for
many people.', 'score': 0}]}}
```

- Currently available functionalities:
 - Search the ELG catalogue
 - Download ELG resources
 - Call ELG services
 - Compare multiple ELG services on various inputs (benchmarking)
- Coming up:
 - Contribute resources/services to ELG using Python through automated deployment
 - Enable running ELG services locally (by running Docker images locally)
 - Improve search experience
 - Many additional features ...

LT Services: From the consortium

- ELG Release 1 in April 2020 finalized APIs for major classes of services (ASR, IE, MT, TTS)
- Concentrated on a subset of EU languages – Czech, English, French, German, Greek, Latvian, Spanish (native languages of the ELG consortium)
 - 9 ASR services, approx. 150 distinct services for IE & Text Analytics, 24 MT, 2 TTS
- Release 2 (early 2021) added support for more EU and related languages
 - 8 further ASR services, approx. 300 IE & Text Analytics, 24 MT, 17 TTS
- ... and APIs for additional service types
 - 12 services doing keyword spotting in audio data
- Release 3 (2022) will add more languages beyond the EU (13 ASR, approx. 150 IE, 9 MT), and new service types including OCR



Stakeholders and Users

Companies that

- ... *develop* Language Technologies
- ... *integrate* Language Technologies
- ... *purchase* Language Technologies

Universities and research centres that

- ... *develop* Language Technologies
- ... *use* Language Technologies

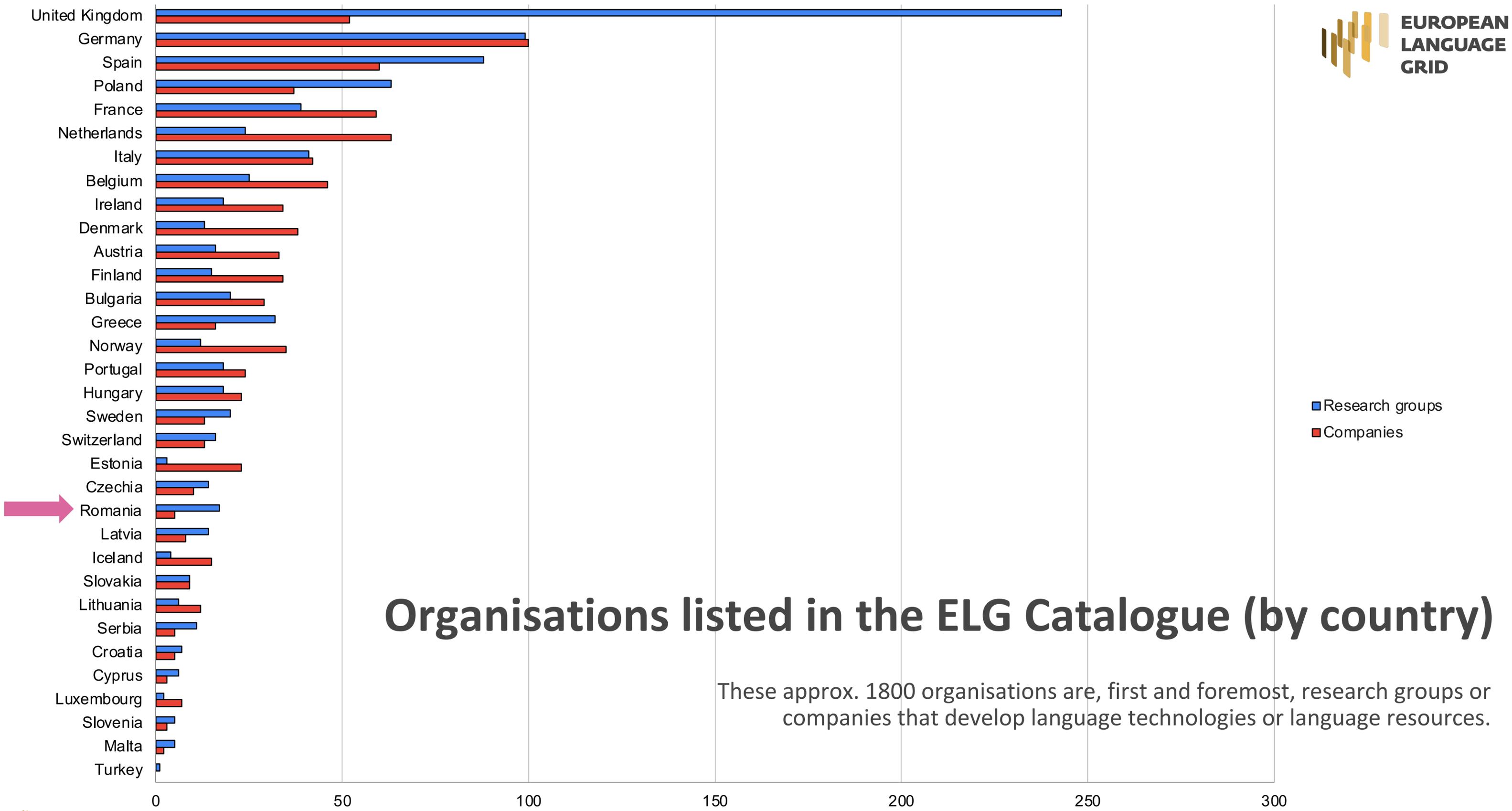
Public administrations that *purchase* or *use* Language Technologies

Other organisations (e.g., NGOs) that *purchase* or *use* Language Technologies

Funding agencies that support the development of Language Technologies



META-FORUM 2019 (8/9 October) – Brussels, Belgium



Organisations listed in the ELG Catalogue (by country)

These approx. 1800 organisations are, first and foremost, research groups or companies that develop language technologies or language resources.

META-FORUM Conference Series

META-FORUM 2022 – June 08-10, Brussels, Belgium

Save the date!

META-FORUM 2021 – November 15-17, virtual conference

Using the European Language Grid

META-FORUM 2020 – December 01-03, virtual conference

Piloting the European Language Grid

META-FORUM 2019 – October 08/09, Brussels, Belgium

Introducing the European Language Grid

META-FORUM 2017 – November 13/14, Brussels, Belgium

Towards a Human Language Project

META-FORUM 2016 – July 04/05, Lisbon, Portugal

Beyond Multilingual Europe

META-FORUM 2015 – April 27, Riga, Latvia

Technologies for the Multilingual Digital Single Market

META-FORUM 2013 – September 19/20, Berlin, Germany

Connecting Europe for New Horizons

META-FORUM 2012 – June 20/21, Brussels, Belgium

A Strategy for Multilingual Europe

META-FORUM 2011 – June 27/28, Budapest, Hungary

Solutions for Multilingual Europe

META-FORUM 2010 – November 17/18, Brussels, Belgium

Challenges for Multilingual Europe

All META-FORUM 2020 and META-FORUM 2021 sessions are available on the European Language Grid YouTube channel:

<https://www.youtube.com/channel/UCarEHmsWT2JslcvvWkbhL4A>

The screenshot shows the YouTube channel page for the European Language Grid, which has 61 subscribers. The channel is categorized under 'VIDEOS'. The page displays a grid of video uploads, including:

- META-FORUM 2021 (21 views, Streamed 20 hours ago)
- META-FORUM 2021 - Session 5: European Language... (24 views, Streamed 22 hours ago)
- META-FORUM 2021 - Session 4: Language-centric AI in... (37 views, Streamed 1 day ago)
- META-FORUM 2021 - Session 3: European language... (59 views, Streamed 1 day ago)
- META-FORUM 2021 - Session 2: ELG Pilot Projects (54 views, Streamed 2 days ago)
- META-FORUM 2021 - Session 1: Opening (317 views, Streamed 2 days ago)
- 3-Year-Old Moraima speaks Galician with Siri - European... (83 views, 2 days ago)
- Georg Rehm: "The Role of the European Language Grid for..." (21 views, 2 months ago)
- Browsing the ELG catalogue (39 views, 5 months ago)
- What is the European Language Grid? (107 views, 6 months ago)
- EUROPEAN LANGUAGE GRID (January 2021) (4:52)
- Context: "Language Equality" EP Resolution (34:19)
- Europe's wider Multilingual Europe Community (9:16)
- ELG: Goals and Objectives (42:20)
- ONDEWO GmbH (1:20:53)

Stakeholders and Users and Collaborators

ELG Open Calls: 15 pilot projects

ICT-29b research and innovation projects

ELE – European Language Equality

AI4EU – European AI on demand platform

Other related projects: Elexis, Lynx, MeMAD, CEF INEA projects (MAPA, NTEU etc.), OpenGPT-X, NFDI4DataScience etc.

Core networks: META-NET, LT Innovate

Additional initiatives: CLAIRE, AI PPP, ELRC, ECSPM, EFNIL, BDVA, NPLD, CLARIN, W3C, NGI, RDA, EOSC, OpenAIRE etc.



Community: NCCs and LTC



32 National Competence Centres (NCCs)

- Strong international network of national networks to broaden ELG's reach, identify content for the ELG and interest companies in using the ELG.
- Main goal: *support the mission of the ELG project.*

European LT Council (LTC) (work in progress)

- A pan-European body in which strategic LT-related matters can be discussed and coordinated.
- Main goal: *represent and support European LT community.*



Austria



DAGMAR GROMANN

Centre for Translation Studies, University of Vienna

Belgium



WALTER DAELEMANS

Computational Linguistics and Psycholinguistics, University of Antwerp

Bulgaria



SVETLA KOEVA

Institute for Bulgarian Language, Bulgarian Academy of Sciences

Croatia



MARKO TADIĆ

Department of Linguistics, University of Zagreb

Cyprus



DORA LOIZIDOU

University of Cyprus

Czech Republic



JAN HAJIČ

Institute of Formal and Applied Linguistics, Charles University Prague

Denmark



BOLETTE SANDFORD PEDERSEN

Centre for Language Technology, Department of Nordic Studies and Linguistics, University of Copenhagen

1. **Austria:** Dagmar Gromann, Zentrum für Translationswissenschaft, Universität Wien
2. **Belgium:** Walter Daelemans, Comp. Ling. and Psycholing. Res. Centre (CLiPS), University of Antwerp
3. **Bulgaria:** Svetla Koeva, Bulgarian Academy of Sciences
4. **Croatia:** Marko Tadic, Inst. of Linguistics, Faculty of Hum. and Social Science, University of Zagreb
5. **Cyprus:** Dora Loizidou, University of Cyprus
6.  **Czech Republic:** Jan Hajic, Inst. of Formal and Applied Linguistics, Charles University in Prague
7. **Denmark:** Bolette Sandford Pedersen, Centre for Lang. Technology, Department of Nordic Research, University of Copenhagen
8. **Estonia:** Kadri Vare, Department of Language, Estonian Ministry of Education and Research
9. **Finland:** Krister Linden, Department of Modern Languages, University of Helsinki
10. **France:** François Yvon, CNRS-LIMSI
11.  **Germany:** Georg Rehm, Speech and Language Technology Lab, DFKI
12.  **Greece:** Maria Gavriilidou, ILSP, R.C. "Athena"
13. **Hungary:** Tamás Várardi, Research Institute for Linguistics, Hungarian Academy of Sciences
14. **Iceland:** Eiríkur Rögnvaldsson, School of Humanities, University of Iceland
15. **Ireland:** Andy Way, ADAPT Centre and School of Computing, Dublin City University
16. **Italy:** Bernardo Magnini, Human Language Technology, Fondazione Bruno Kessler (FBK)
17. **Latvia:** Inguna Skadina, Institute of Mathematics and Computer Science, University of Latvia
18. **Lithuania:** Albina Aukšoriūtė, Institute of the Lithuanian Language
19. **Luxembourg:** Dimitra Anastasiou, Luxembourg Institute of Science and Technology
20. **Malta:** Michael Rosner, Department Intelligent Computer Systems, University of Malta
21. **Netherlands:** Jan Odijk, Utrecht Institute of Linguistics, Universiteit Utrecht
22. **Norway:** Kristine Eide, The Language Council of Norway – Språkrådet
23. **Poland:** Maciej Ogrodniczuk, Institute of Computer Science, Polish Academy of Sciences
24. **Portugal:** António Branco, Department of Informatics, University of Lisbon
25. **Romania:** Dan Tufis, Research Institute for Artificial Intelligence, Romanian Academy of Sciences
26. **Serbia:** Cvetana Krstev, Faculty of Mathematics, Belgrade University (UBG)
27. **Slovakia:** Radovan Garabik, Ludovít Stur Institute of Linguistics, Slovak Academy of Sciences
28. **Slovenia:** Simon Krek, Jozef Stefan Institute
29. **Spain:** Marta Villegas, Barcelona Supercomputing
30. **Sweden:** Jens Edlund, Speech, Music & Hearing/Språkbanken Tal, KTH Royal Institute of Technology
31. **Switzerland:** Hervé Boudlard, Idiap Research Institute
32.  **UK:** Kalina Bontcheva, Department of Computer Science, University of Sheffield



European Language Grid: Sustainable Operational Model & Legal Entity

- ELG is supposed to be a long-term, sustainable initiative – a legal entity is needed.
- The technical and operational requirements – high availability and performance, SLAs, billing, support etc. – create non-trivial costs: hosting; bandwidth; ELG team; legal; etc.
- We've identified several ways and approaches of covering the costs on a long-term basis.
- Establish consensus for a sustainable operational model.
- Options: a) for-profit or b) *not-for-profit company*, c) association, d) foundation
- Business Model Canvas prepared
- First set of approx. 20 products defined
- Next up: market validation of the product ideas
- Q2 2022: Establish legal entity



Open Calls for Pilot Projects

Two open calls for pilot projects

- **Open Call #1:** 03/04 2020
- **Open Call #2:** 10/11 2020

Pilot projects shall

- **Type A:** broaden ELG's portfolio or
- **Type B:** demonstrate usefulness of ELG

Up to €200,000 per project

Approx. €2,000,000 FSTP in total

Available in Open Call #1: 1.3M€

Available in Open Call #2: 585k€

Project duration: 9-12 months

Eligibility: SMEs, research organisations

Open Call #1 – Statistics

110 project proposals evaluated !

- **Type A:** 79 – **Type B:** 31 proposals

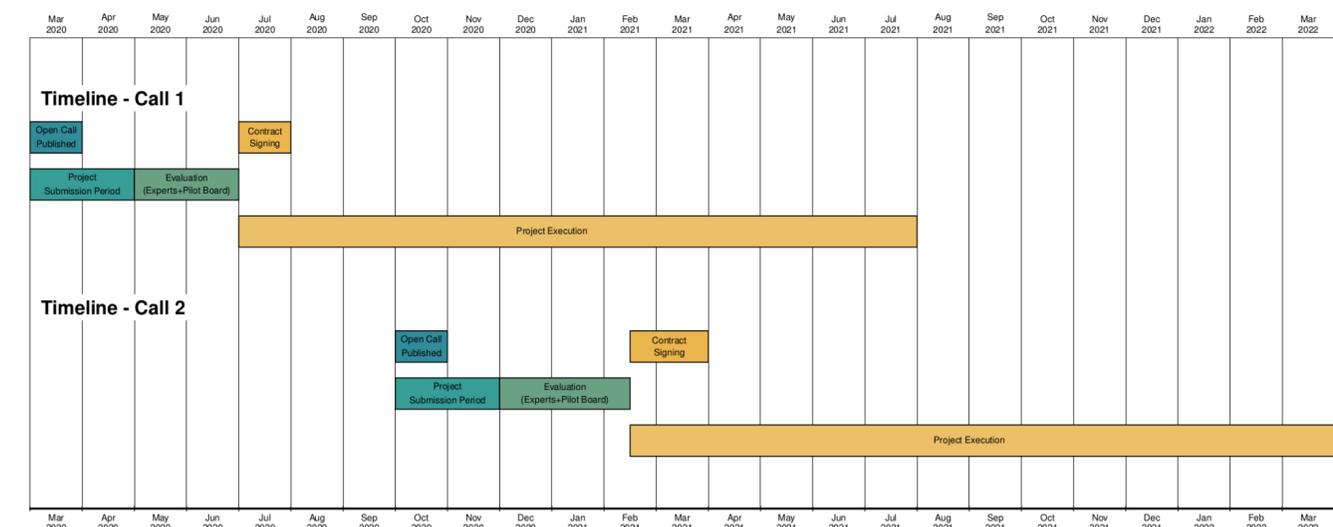
Total amount requested: 16.9M€

→ 10 projects selected

Open Call #2 – Statistics

106 project proposals evaluated !

→ 5 projects selected



Open Call #1

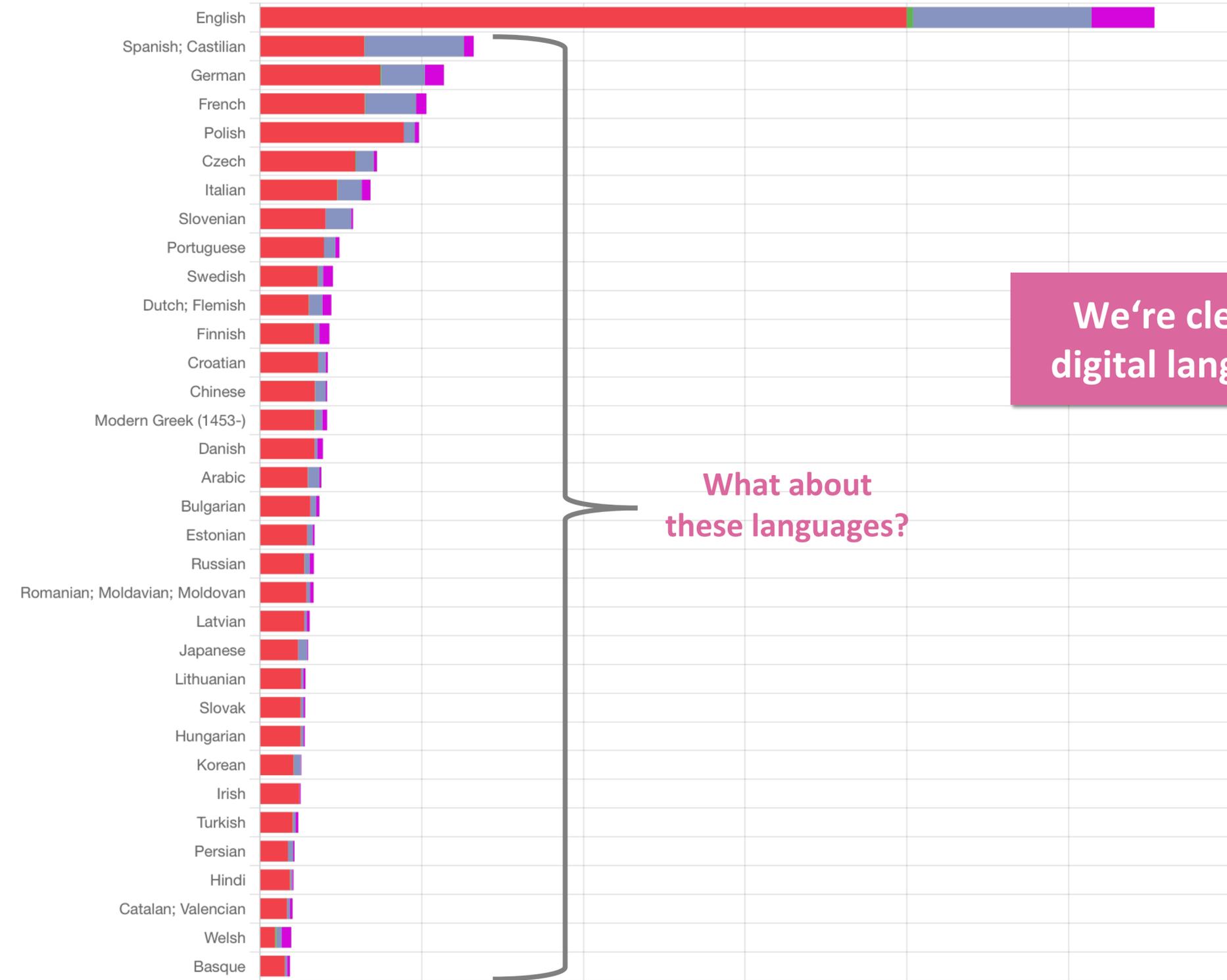
Fondazione Bruno Kessler	European Clinical Case Corpus ✓	IT	139,370€
Lingsoft, Inc.	Lingsoft Solutions as Distributable Containers ✓	FI	140,625€
Coreon GmbH	MKS as Linguistic Linked Open Data ✓	DE	167,375€
Elhuyar Fundazioa	Basque-speaking smart speaker based on Mycroft AI ✓	ES	117,117€
Universita' Degli Studi di Torino	Italian EVALITA Benchmark Linguistic Resources, NLP Services and Tools ✓	IT	126,125€
University of Helsinki	Open Translation Models, Tools and Services ✓	FI	154,636€
University of Vienna	Extracting Terminological Concept Systems from Natural Language Text ✓	AT	132,977€
University of Turku	Textual paraphrase dataset for deep language modelling ✓	FI	166,085€
Weber Consulting KG	Virtual Personal Assistant Prototype ✓	AT	87,445€
FZI Research Center for IT	Streaming Language Processing in Manufacturing ✓	DE	132,160€

Open Call #2

Institute for Bulgarian Language	Multilingual Image Corpus 2021	BG	110,960€
EDIA BV	CEFR Labelling and Assessment Services	NL	137,560€
University of West Bohemia	Motion Capture 3D Sign Language Resources	CZ	85,421€
Sapienza University of Rome	Universal Semantic Annotator	IT	113,228€
Sign Time GmbH	Sign language explanations for terms in a text	AT	137,227€

Languages of the Resources

■ Corpus
 ■ LanguageDescription
 ■ LexicalConceptualResource
 ■ Organization
 ■ Project
 ■ ToolService



We're clearly very far away from digital language equality in Europe.

What about these languages?



EUROPEAN LANGUAGE EQUALITY

Consortium: 52 partners from all over Europe

Coordinator: ADAPT Centre (Dublin City University)

Objective: *development of a strategic research, innovation and deployment agenda to achieve digital language equality in Europe by 2030*

Runtime: 18 months – ELE & ELG will both finish in June 2022

Start on 1 January 2021

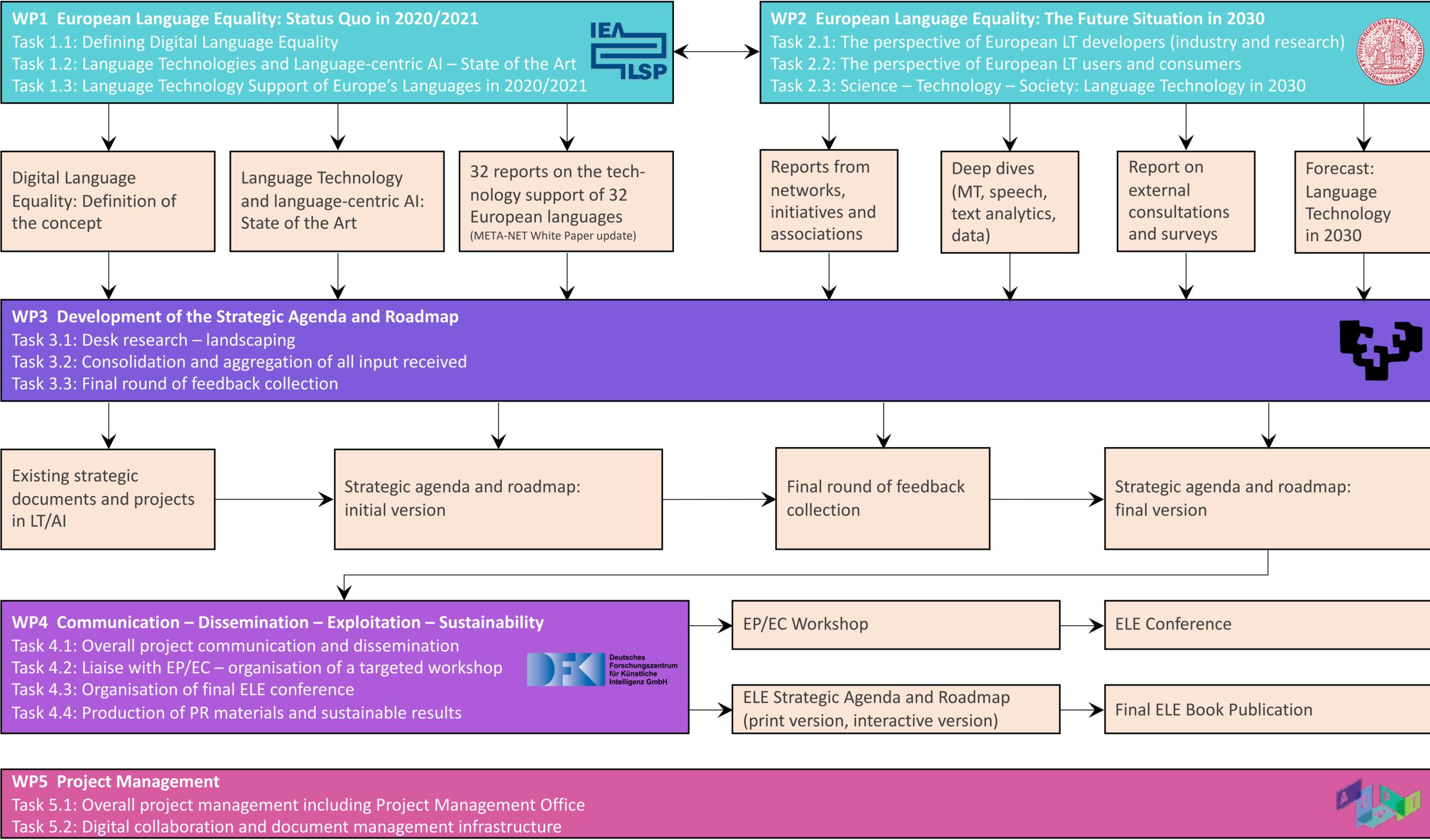
PP/PA (*not* Horizon 2020) – Pilot Project/Preparatory Action

Budget: 1,8M€ (Coordination and Support Action)

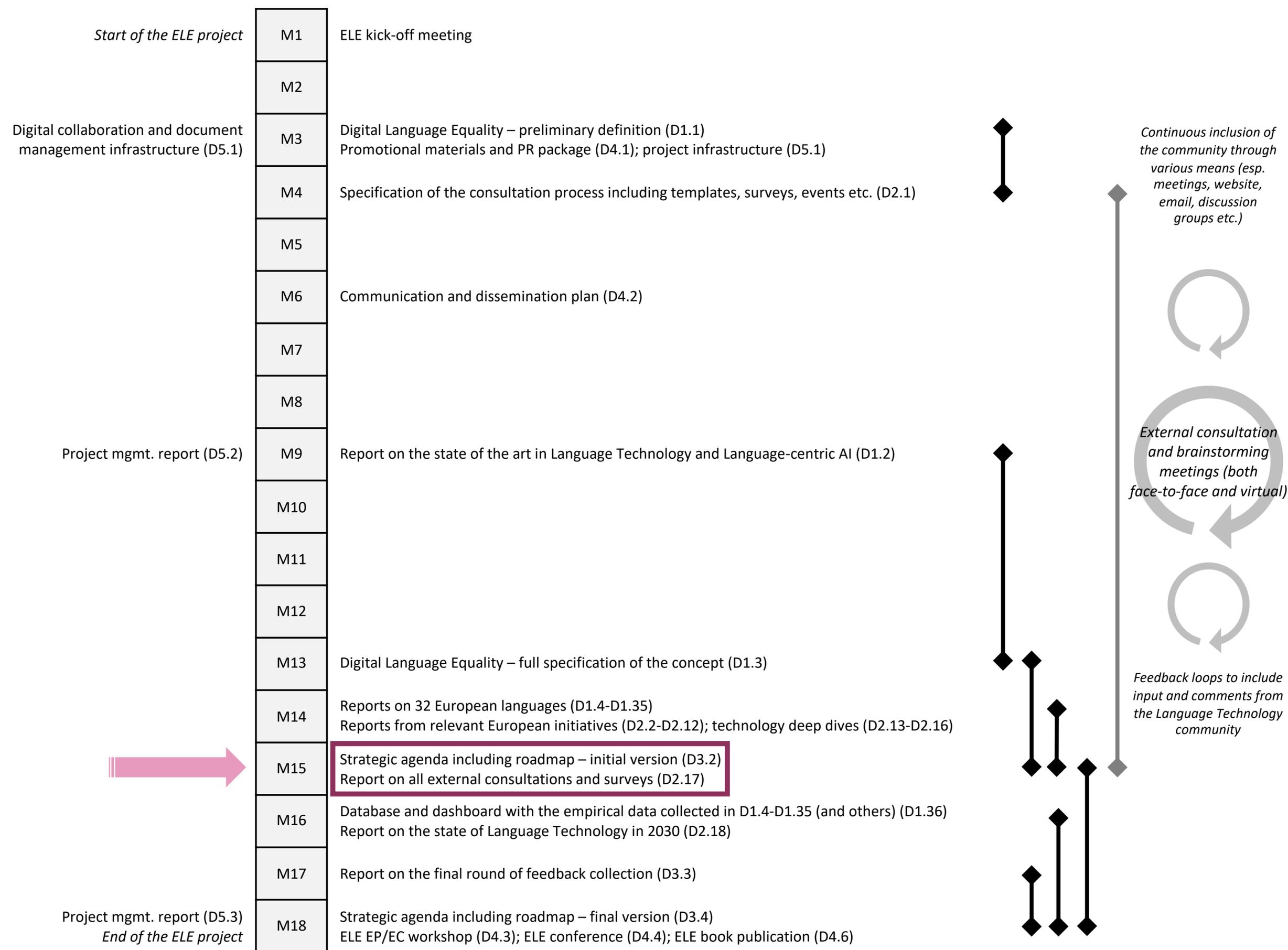
<http://www.european-language-equality.eu>



EUROPEAN LANGUAGE GRID



**Work
Packages and
main
Deliverables**



Timeline

Current State of Play in the Project

- **Digital Language Equality:** concept defined, *technological* and *contextual factors* specified, currently working on the implementation of the DLE metric
- **Update of META-NET White Papers** successfully concluded – technology support of Europe’s languages now assessed *dynamically and fully empirically*
- **Collection of LRs/LTs:** with the whole ELE consortium we collected 6000+ additional records
- **ELG:** these 6000+ records have been included in ELG, arriving at *11,000+ resources* in total
- **Dynamic DLE computation:** based on these 11,000 metadata records we can compute the DLE metric for all European languages and visualise the results through a dedicated dashboard
- **Continuously:** this approach enables us to monitor the development of DLE in Europe over time
- **ELG:** Making a new resource available through ELG will increase the DLE metric for that language
- **EP:** Workshop on DLE in the European Parliament in approx. Q2/2022 (tbc)
- Almost all analyses to be ready in approx. April 2022 – ELE book to be published in July/Aug. 2022

Digital Language Equality: Preliminary Working Definition

Digital Language Equality (DLE) is the state of affairs in which all languages have the technological support and situational context necessary for them to continue to exist and to prosper as living languages in the digital age.



D1.1
Digital Language Equality
(preliminary definition)

Authors	Federico Gaspari (DCU), Andy Way (DCU), Jane Dunne (DCU), Georg Rehm (DFKI), Stelios Piperidis (ILSP), Maria Giagkou (ILSP)
Dissemination level	Public
Date	31-03-2021

Towards the DLE Metric

Technological Factors

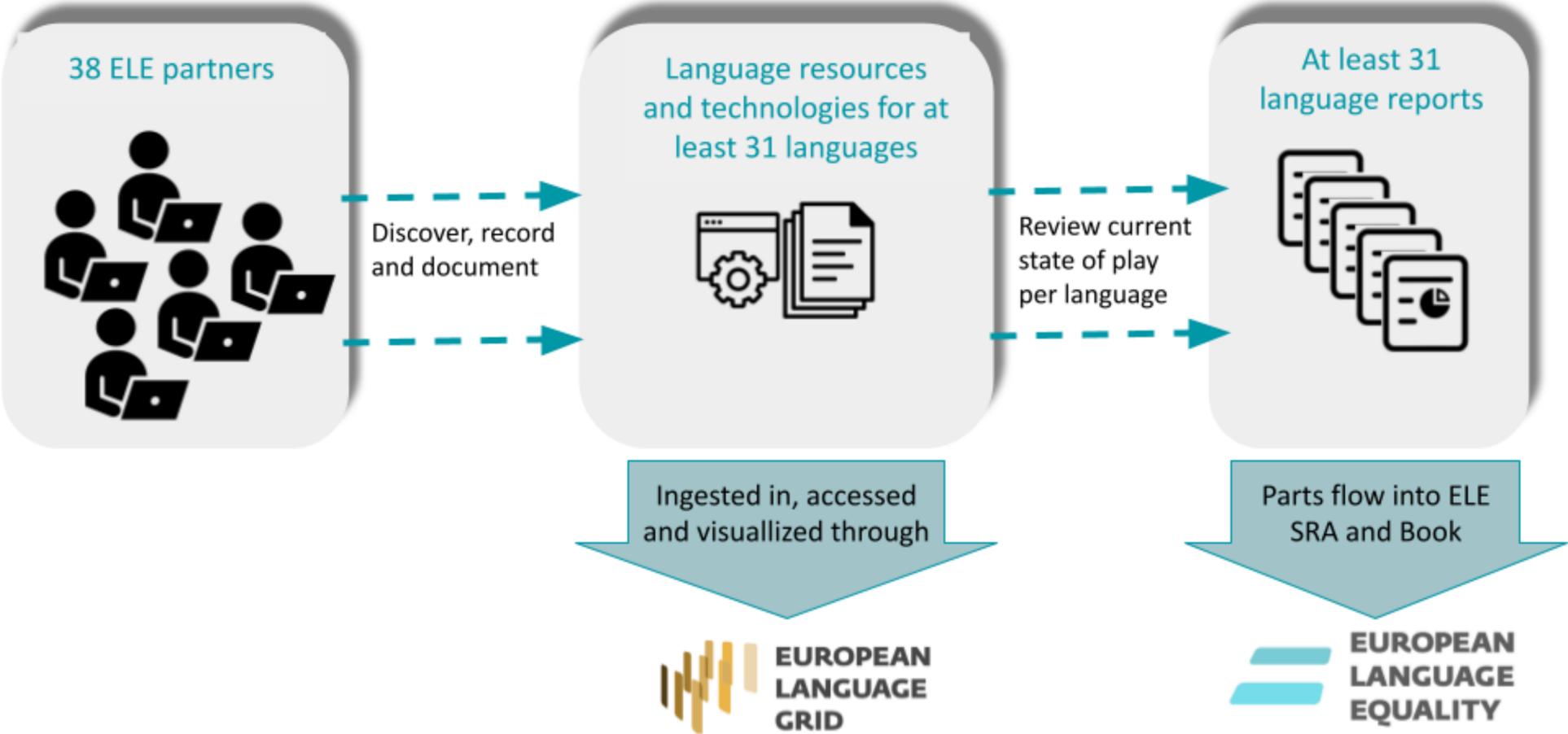
- Tools and Services
- Corpora
- Language Descriptions and Models
- Lexical and Conceptual Resources
- Projects
- Organisations



Contextual Factors

- Industry
- Online
- Research & Development & Innovation
- Technology
- Economy
- Education
- Funding
- Law
- Media
- Policy
- Public administration
- Society

Compiling a collection of LRTs representative of the current state of digital (in)equality of Europe's languages



List of all languages investigated by the ELE consortium:

<https://european-language-equality.eu/languages/>

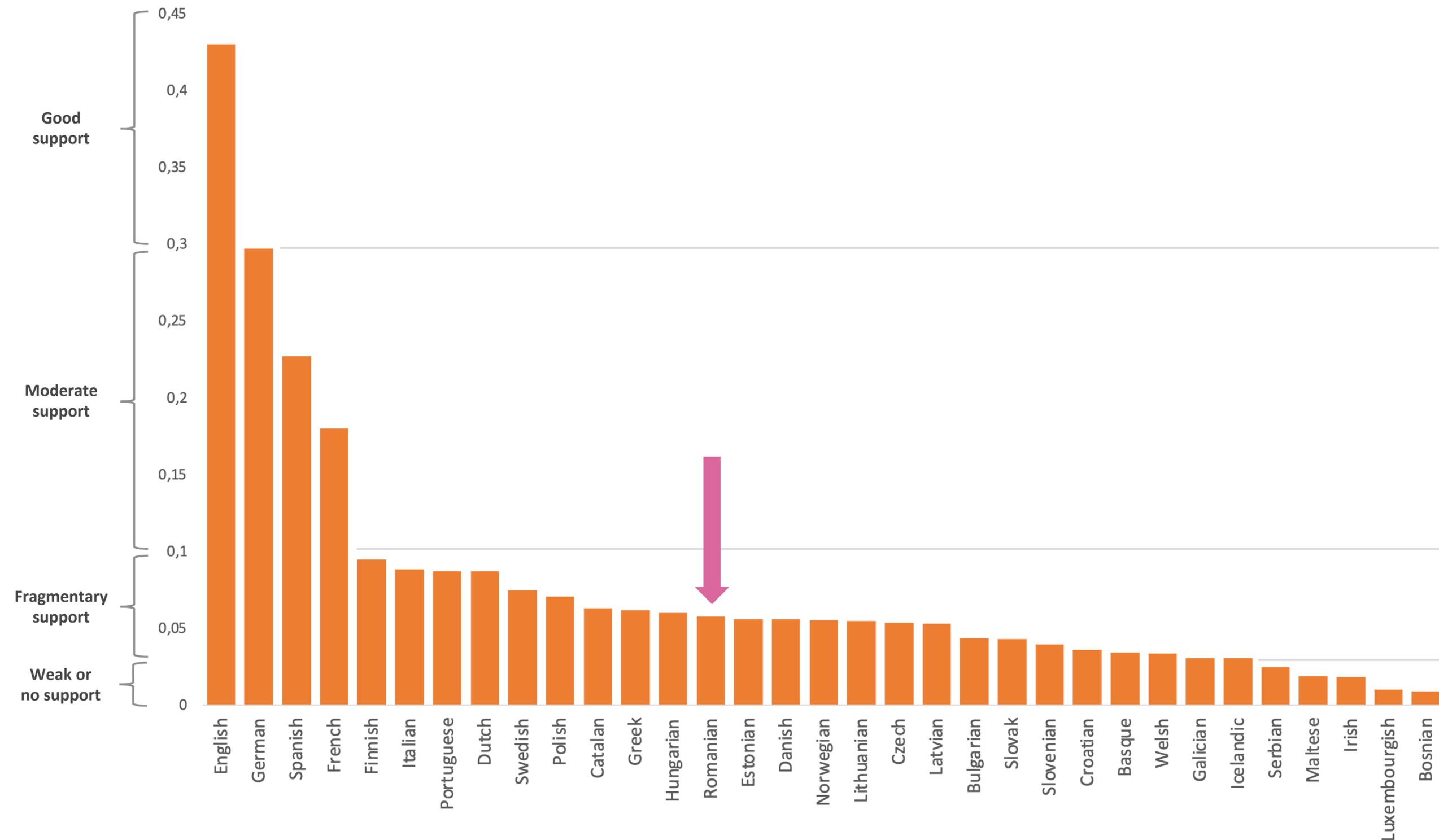
Preliminary Results

		Tools and services							Data (corpora, models, lexical resources)					Overall level of support	
		Text Processing	Speech Processing	Image/Video processing	Information extraction and Information retrieval	Human-Computer interaction	Translation technologies	Natural Language Generation	Text corpora	Multimodal corpora	Parallel corpora	Models	Lexical resources		
EU Official languages	Bulgarian														
	Croatian														
	Czech														
	Danish														
	Dutch														
	English														
	Estonian														
	Finnish														
	French														
	German														
	Hungarian														
	Irish														
	Italian														
	Latvian														
	Lithuanian														
	Maltese														
	Greek														
	Polish														
	Portuguese														
	Romanian														
Slovak															
Slovenian															
Spanish															
Swedish															
Other (co-)official languages at the national level	Albanian														
	Bosnian														
	Icelandic														
	Luxembourgish														
	Macedonian														
	Norwegian														
	Serbian														
(Co-)official languages at regional level	Basque														
	Catalan														
	Faroese														
	Frisian, Western														
	Galician														
	Jerriais														
	Low German														
	Manx														
	Mirandese														
	Occitan														
	Sorbian, Upper														
Welsh															
All other languages under investigation															

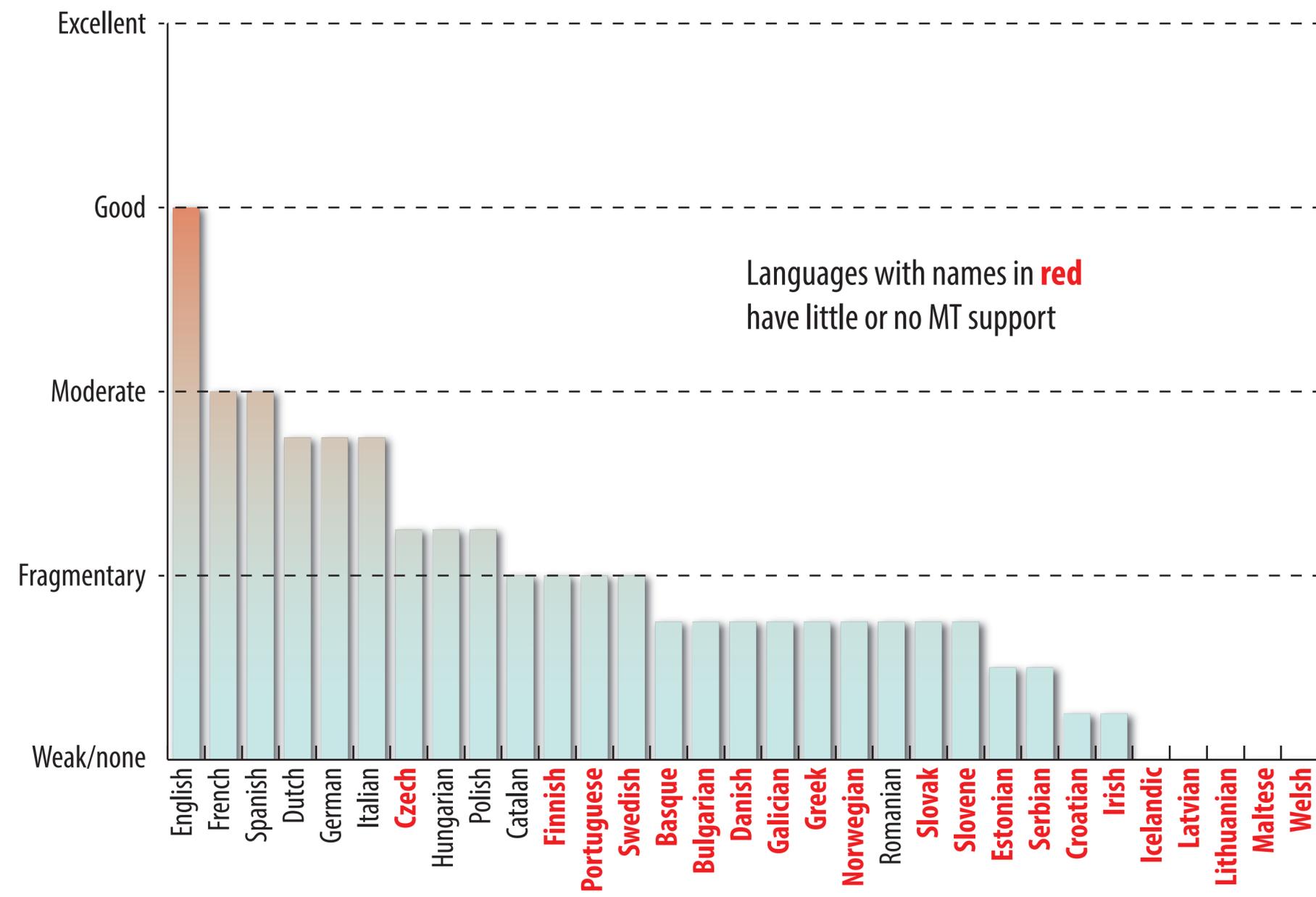
Legend:	Weak/no support	Fragmentary	Moderate	Good
----------------	-----------------	-------------	----------	------

Results based on raw counts of the 11,000+ language resources and language technologies currently described with metadata records in the ELG platform.

Preliminary Results



Level of support

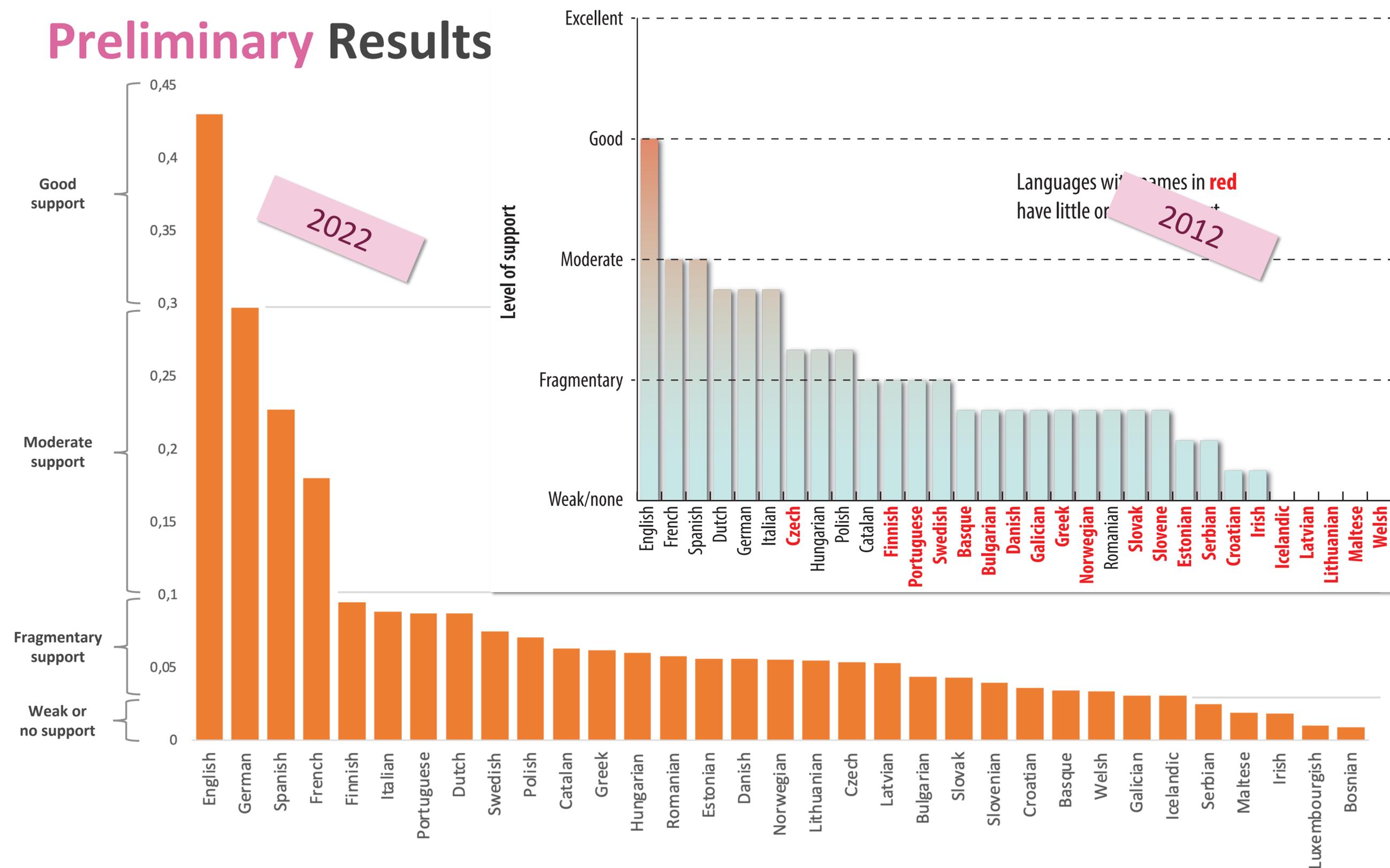


- THE SWEDISH LANGUAGE IN THE DIGITAL AGE – SVENSKA SPRÅKET I DEN DIGITALA TIDSÅDEREN
- THE SPANISH LANGUAGE IN THE DIGITAL AGE – LA LENGUA ESPAÑOLA EN LA ERA DIGITAL
- THE SLOVENE LANGUAGE IN THE DIGITAL AGE – SLOVENSKI JEZIK V DIGITALNI DOBI
- THE SLOVAK LANGUAGE IN THE DIGITAL AGE – SLOVENSKÝ JAZYK V DIGITÁLNOM VEKU
- THE SERBIAN LANGUAGE IN THE DIGITAL AGE – СРПСКИ ЈЕЗИК У ДИГИТАЛНОМ ДОБУ
- THE ROMANIAN LANGUAGE IN THE DIGITAL AGE – LIMBA ROMÂNĂ ÎN ERA DIGITALĂ
- THE PORTUGUESE LANGUAGE IN THE DIGITAL AGE – A LINGUA PORTUGUESA NA ERA DIGITAL
- THE NORWEGIAN LANGUAGE IN THE DIGITAL AGE – NORSK I DEN DIGITALE TIDSÅDEREN
- THE POLISH LANGUAGE IN THE DIGITAL AGE – JEZYK POLSKI W ERZE CYFROWEJ
- THE ITALIAN LANGUAGE IN THE DIGITAL AGE – LA LINGUA ITALIANA NELL'ERA DIGITALE
- THE LITHUANIAN LANGUAGE IN THE DIGITAL AGE – LIetuvių kalba skaitmeniniame amžiuje
- THE LATVIAN LANGUAGE IN THE DIGITAL AGE – LATVIEŠU VALODA DIGITĀLAJĀ LAIKMĒTĀ
- THE HUNGARIAN LANGUAGE IN THE DIGITAL AGE – A MAGYAR NYELV A DIGITÁLIS KORBAN
- THE ICELANDIC LANGUAGE IN THE DIGITAL AGE – ISLÆNSK TUNGA Á STARFÆRNINI ÖLD
- THE IRISH LANGUAGE IN THE DIGITAL AGE – AN GHAELIGE SA RÉ DHIGITEACH
- THE GREEK LANGUAGE IN THE DIGITAL AGE – Η ΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΣΤΗΝ ΨΗΦΙΑΚΗ ΕΠΟΧΗ
- THE GALICIAN LANGUAGE IN THE DIGITAL AGE – O IDIOMA GALEGO NA ERA DIGITAL
- THE FINNISH LANGUAGE IN THE DIGITAL AGE – LA LANGUE FRANÇAISE À L'ÈRE DU NUMÉRIQUE
- THE ESTONIAN LANGUAGE IN THE DIGITAL AGE – SUOMEN KIELI DIGITAALISELÄ AIKAKAUDELTA
- THE DUTCH LANGUAGE IN THE DIGITAL AGE – HET NEDERLANDS IN HET DIGITALE TIJDPERK
- THE DANISH LANGUAGE IN THE DIGITAL AGE – DET DANSKE SPROG I DEN DIGITALE TIDSÅDER
- THE CZECH LANGUAGE IN THE DIGITAL AGE – ČEŠTINA V DIGITÁLNÍM VEKU
- THE CROATIAN LANGUAGE IN THE DIGITAL AGE – HRVATSKI JEZIK U DIGITALNOJ DOBU
- THE CATALAN LANGUAGE IN THE DIGITAL AGE – LA LLINGÜA CATALANA A L'ERA DIGITAL
- THE BULGARIAN LANGUAGE IN THE DIGITAL AGE – БЪЛГАРСКИЯТ ЕЗИК В ДИГИТАЛНАТА ЕПОХА
- THE BASQUE LANGUAGE IN THE DIGITAL AGE – EUSKARA ARO DIGITALEAN
- THE META-NET STRATEGIC RESEARCH AGENDA FOR MULTILINGUAL EUROPE 2020



Source: META-NET White Paper Series: Europe's Languages in the Digital Age. Springer, Heidelberg, New York, Dordrecht, London, September 2012. Georg Rehm and Hans Uszkoreit (series editors)

Preliminary Results



European Language Grid: An Overview

Georg Rehm¹, Maria Berger¹, Ela Elsholz¹, Stefanie Hegele¹, Florian Kintzel¹, Katrin Marheinecke¹, Stelios Piperidis², Miltos Deligiannis², Dimitrios Galanis², Katerina Gkirtzou², Penny Labropoulou², Kalina Bontcheva³, David Jones³, Ian Roberts³, Jan Hajič⁴, Jana Hamrlová⁴, Lukáš Kačena⁴, Khalid Choukri⁵, Victoria Arranz⁵, Andrejs Vasiljevs⁶, Orians Anvari⁶, Andis Lagzdins⁶, Jūlija Melņika⁶, Gerhard Backfried⁷, Erinc Dikiçi⁷, Miroslav Janosik⁷, Katja Prinz⁷, Christoph Prinz⁷, Severin Stampfer⁷, Dorothea Thomas-Aniola⁷, José Manuel Gómez Pérez⁸, Andres Garcia Silva⁸, Cristian Berrio⁸, Ulrich Germann⁹, Steve Renals⁹, Ondrej Klejch⁹

¹DFKI GmbH, Germany • ²ILSP/Athena RC, Greece • ³University of Sheffield, UK • ⁴Charles University, Czech Republic • ⁵Evaluations and Language Resources Distribution Agency (ELDA), France • ⁶Tilde, Latvia • ⁷SAIL LABS Technology GmbH, Austria • ⁸Expert System Iberia SL, Spain • ⁹University of Edinburgh, UK

Corresponding author: Georg Rehm – georg.rehm@dfki.de

Abstract

Making Metadata Fit for Next Generation Language Technology Platforms: The Metadata Schema of the European Language Grid

Penny Labropoulou¹, Katerina Gkirtzou¹, Maria Gavrilidou¹, Miltos Deligiannis¹, Dimitrios Galanis¹, Stelios Piperidis¹, Georg Rehm², Maria Berger², Valérie Mapelli³, Mickaël Rigault³, Victoria Arranz³, Khalid Choukri³, Gerhard Backfried⁴, José Manuel Gómez Pérez⁵, Andres Garcia Silva⁵

¹ILSP/Athena RC, Greece, ²DFKI GmbH, Germany, ³Evaluations and Language Resources Distribution Agency (ELDA), France, ⁴SAIL LABS Technology GmbH, Austria, ⁵Expert System Iberia SL, Spain
{penny.gkirtzou, maria.mdl, galanis.d, spip}@athenarc.gr, {georg.rehm, maria.berger}@dfki.de, {mapelli, mickael, arranz, choukri}@elda.org, Gerhard.Backfried@sail-labs.com, {agarcia, jmgomez}@expertsystem.com

Abstract

The current scientific and technological landscape is characterised by the increasing availability of data resources and processing tools and services. In this setting, metadata have emerged as a key factor facilitating management, sharing and usage of such digital assets. In this paper we present ELG-SHARE, a rich metadata schema catering for the description of Language Resources and Technologies (processing and generation services and tools, models, corpora, term lists, etc.), as well as related entities (e.g., organizations, projects, supporting documents, etc.). The schema powers the European Language Grid platform that aims to be the primary hub and marketplace for industry-relevant Language Technology in Europe. ELG-SHARE has been based on various metadata schemas, vocabularies, and ontologies, as well as related recommendations and guidelines.

Keywords: metadata, language technology, language technology services, language resources

1. Introduction

The rise of data-driven approaches that use Machine Learning (ML), and especially the breakthroughs in the Deep Learning field, has put data into a central place in all scientific and technological areas. Natural Language Processing (NLP) being no exception. Datasets and NLP tools and services are made available through various repositories (institutional, disciplinary, general purpose, etc.), which makes it hard to find the appropriate resources for one's purposes. Even if they are brought together in one catalogue, such as the European Open Science Cloud¹ or the Google dataset search service², the difficulty of spotting the right resources and services among thousands still remains. Metadata plays an instrumental role in solving this puzzle, as it becomes the intermediary between consumers (humans and machines) and digital resources.

In addition, in the European Union, with the 24 official and many additional languages, multilingualism, cross-lingual and cross-cultural communication in Europe as well as an inclusive Digital Single Market³ can only be enabled and firmly established through Language Technologies (LT). The boosting of the LT domain is thus of utmost importance. To this end, the European LT industry needs to be strengthened, promote its products and services, integrate them into applications, and collaborate with academia into advancing research and innovation, and bringing research outcomes to a mature level of entering the market. The European Language Grid (ELG) project⁴ aims to drive forward the European LT sector by creating a platform and establishing it as the primary hub and marketplace for the LT community. The ELG is developed to be a scalable

¹<https://www.eosc-portal.eu>

²<https://toolbox.google.com/datasetsearch>

³<https://ec.europa.eu/digital-single-market/en>

⁴<https://www.european-language-grid.eu>

cloud platform, providing in an easy-to-integrate way, access to hundreds of commercial and non-commercial LTs for all European languages, including running tools and services as well as data resources. Discovery of and access to these resources can only be achieved through an appropriate metadata schema. We present here the ELG-SHARE schema, which is used for the description of LT-related resources shared through the ELG platform and its contribution to the project goals.

2. Objectives

The ELG project (Rehm et al., 2020a) aims to foster European LT by addressing the fragmentation that hinders its development; see indicatively (Rehm and Hegele, 2018; Rehm et al., 2016). To this end, it builds a platform dedicated to the *distribution and deployment of Language Resources and Technologies (LRT)*, aspiring to establish it as the primary platform and marketplace for industry-relevant LT in Europe. The *promotion of LT stakeholders and activities* and growth of their visibility and outreach is also one of its goals. Together with complementary material in the portal (e.g., training material, information on events, job offerings, etc.), ELG offers a comprehensive picture of the European LT sector.

The ELG platform⁵ will offer access to hundreds of *commercial and non-commercial LTs* and ancillary *data LRs* for all European languages and more; these include processing and generation services, tools, applications for written and spoken language, corpora, lexicons, ontologies, term lists, models, etc. All resources are accessed through their descriptions in the ELG catalogue. LRT providers can describe, upload, and integrate their assets in ELG, and LRT

⁵The ELG platform has just been launched (alpha release) and will continue to be updated with new resources and functionalities (official release dates are on April of 2020, 2021 and 2022).

The European Language Technology Landscape in 2020: Language-Centric and Human-Centric AI for Cross-Cultural Communication in Multilingual Europe

Georg Rehm¹, Katrin Marheinecke¹, Stefanie Hegele¹, Stelios Piperidis², Kalina Bontcheva³, Jan Hajič⁴, Khalid Choukri⁵, Andrejs Vasiljevs⁶, Gerhard Backfried⁷, Christoph Prinz⁷, José Manuel Gómez Pérez⁸, Luc Meertens⁹, Paul Lukowicz¹⁰, Josef van Genabith¹⁰, Andrea Lösch¹¹, Philipp Slusalek¹¹, Morten Irgens¹⁰, Patrick Gatellier¹¹, Joachim Köhler¹², Laure Le Bars¹³, Dimitra Anastasiou¹⁴, Albina Auksoyri¹⁵, Nária Bel¹⁶, António Branco¹⁷, Gerhard Budin¹⁸, Walter Daelemans¹⁹, Koenraad De Smedt²⁰, Radovan Garabik²¹, Maria Gavrilidou², Dagmar Gromann¹⁹, Svetla Koeva²², Simon Krek²³, Cvetana Krstev²⁴, Krister Lindén²⁵, Bernardo Magnini²⁶, Jan Odijk²⁷, Maciej Ogrodniczuk²⁸, Eirikur Rögnvaldsson²⁹, Mike Rosner³⁰, Bolette Sandford Pedersen³¹, Inguna Skadiņa³², Marko Tadić³³, Dan Tufiş³⁴, Tamás Váradi³⁵, Kadri Vider³⁶, Andy Way³⁷, François Yvon³⁸

¹DFKI GmbH, Germany • ²ILSP/Athena RC, Greece • ³University of Sheffield, UK • ⁴Charles University, Czech Republic • ⁵ELDA, France • ⁶Tilde, Latvia • ⁷SAIL LABS Technology GmbH, Austria • ⁸Expert System Iberia SL, Spain • ⁹CrossLang, Belgium • ¹⁰Oslo Metropolitan University, Norway • ¹¹Thales Group, France •

¹²Fraunhofer IIS, Germany

¹⁵University of Jyväskylä, Finland

²¹University of Turku, Finland

²⁷University of Helsinki, Finland

³⁶University of Tartu, Estonia

³⁷University of Cambridge, UK

³⁸University of Edinburgh, UK

³⁹University of Edinburgh, UK

⁴⁰University of Edinburgh, UK

⁴¹University of Edinburgh, UK

⁴²University of Edinburgh, UK

⁴³University of Edinburgh, UK

⁴⁴University of Edinburgh, UK

⁴⁵University of Edinburgh, UK

⁴⁶University of Edinburgh, UK

⁴⁷University of Edinburgh, UK

⁴⁸University of Edinburgh, UK

⁴⁹University of Edinburgh, UK

⁵⁰University of Edinburgh, UK

⁵¹University of Edinburgh, UK

⁵²University of Edinburgh, UK

⁵³University of Edinburgh, UK

⁵⁴University of Edinburgh, UK

⁵⁵University of Edinburgh, UK

⁵⁶University of Edinburgh, UK

⁵⁷University of Edinburgh, UK

⁵⁸University of Edinburgh, UK

⁵⁹University of Edinburgh, UK

⁶⁰University of Edinburgh, UK

⁶¹University of Edinburgh, UK

⁶²University of Edinburgh, UK

⁶³University of Edinburgh, UK

⁶⁴University of Edinburgh, UK

⁶⁵University of Edinburgh, UK

⁶⁶University of Edinburgh, UK

⁶⁷University of Edinburgh, UK

⁶⁸University of Edinburgh, UK

⁶⁹University of Edinburgh, UK

⁷⁰University of Edinburgh, UK

⁷¹University of Edinburgh, UK

⁷²University of Edinburgh, UK

⁷³University of Edinburgh, UK

⁷⁴University of Edinburgh, UK

⁷⁵University of Edinburgh, UK

⁷⁶University of Edinburgh, UK

⁷⁷University of Edinburgh, UK

⁷⁸University of Edinburgh, UK

⁷⁹University of Edinburgh, UK

⁸⁰University of Edinburgh, UK

⁸¹University of Edinburgh, UK

⁸²University of Edinburgh, UK

⁸³University of Edinburgh, UK

⁸⁴University of Edinburgh, UK

⁸⁵University of Edinburgh, UK

⁸⁶University of Edinburgh, UK

⁸⁷University of Edinburgh, UK

⁸⁸University of Edinburgh, UK

⁸⁹University of Edinburgh, UK

⁹⁰University of Edinburgh, UK

⁹¹University of Edinburgh, UK

⁹²University of Edinburgh, UK

⁹³University of Edinburgh, UK

⁹⁴University of Edinburgh, UK

⁹⁵University of Edinburgh, UK

⁹⁶University of Edinburgh, UK

⁹⁷University of Edinburgh, UK

⁹⁸University of Edinburgh, UK

⁹⁹University of Edinburgh, UK

¹⁰⁰University of Edinburgh, UK

European Language Grid:

A Joint Platform for the European Language Technology Community

Georg Rehm¹, Stelios Piperidis², Kalina Bontcheva³, Jan Hajič⁴, Victoria Arranz⁵, Andrejs Vasiljevs⁶, Gerhard Backfried⁷, José Manuel Gómez-Pérez⁸, Ulrich Germann⁹, Rémi Calizzano¹, Nils Feldhus¹, Stefanie Hegele¹, Florian Kintzel¹, Katrin Marheinecke¹, Julián Moreno-Schneider¹, Dimitrios Galanis², Penny Labropoulou², Miltos Deligiannis², Katerina Gkirtzou², Athanasia Kolovou², Dimitris Gkoumas², Leon Voukoutis², Ian Roberts³, Jana Hamrlová⁴, Dušan Variš⁴, Lukáš Kačena⁴, Khalid Choukri⁵, Valérie Mapelli⁵, Mickaël Rigault⁵, Jūlija Melņika⁶, Miroslav Jánošík⁷, Katja Prinz⁷, Andrés García-Silva⁸, Cristian Berrio⁸, Ondrej Klejch⁹, Steve Renals⁹

¹DFKI GmbH, Berlin, Germany; ²R. C. Athena, ILSP, Greece; ³University of Sheffield, UK; ⁴Charles University, Czechia; ⁵ELDA, France; ⁶Tilde, Latvia; ⁷HENSOLDT Analytics, Austria; ⁸Expert System, Spain; ⁹University of Edinburgh, UK

Corresponding author: georg.rehm@dfki.de

Abstract

Europe is a multilingual society, in which dozens of languages are spoken. The only option to enable and to benefit from multilingualism is through Language Technologies (LT), i. e., Natural Language Processing and Speech Technologies. We describe the European Language Grid (ELG), which is targeted to evolve into the primary platform and marketplace for LT in Europe by providing one umbrella platform for the European LT landscape, including research and industry, enabling all stakeholders to upload, share and distribute their services, products and resources. At the end of our EU project, which will establish a legal entity in 2022, the ELG will provide access to approx. 1300 services for all European languages as well as thousands of data sets.

1 Introduction

Europe is a multilingual society with 24 EU Member State languages and dozens of additional languages including regional and minority languages and languages spoken by immigrants, trade partners and tourists. The only option to enable and to benefit from multilingualism is through Language Technologies (LT) including Natural Language Processing (NLP) and Speech Technologies (Rehm, 2017). While the European LT landscape is world class, it is also massively fragmented (Vasiljevs et al., 2019; Rehm et al., 2020d).

We describe Release 2 of the European Language Grid (ELG) cloud platform.¹ This scalable system is targeted to evolve into the primary

¹<https://www.european-language-grid.eu>. We provide a screencast demo video at <https://youtu.be/LD6QadkZiM>.

platform for LT in Europe. It will provide one umbrella platform for all LTs developed by the European LT landscape, including research and industry, addressing a gap that has been repeatedly raised by the European LT community for many years (Rehm and Uszkoreit, 2013; Rehm et al., 2016b; STOA, 2017; Rehm, 2017; Rehm and Hegele, 2018; European Parliament, 2018). ELG is meant to be a virtual home and marketplace for all products, services and organisations active in the LT space in Europe (Rehm et al., 2020a). The platform can be used by all stakeholders to upload, share and distribute their products, tools and resources. At the end of the EU ELG (2019-2022), which will establish a legal entity in early 2022, the platform will provide access to approx. 1300 commercial and non-commercial services for all European languages as well as thousands of language resources. ELG will enable the European LT community to deposit and upload their technologies and data and to deploy them through the grid. It is also meant to support *digital language technology* in Europe (STOA, 2017; European Parliament, 2018), i. e., to create a situation in which languages are supported through technologies well. The current imbalance is characterised by the stark predominance of LTs for English, while most other languages are only marginally supported and, thus, in danger of digital language extinction (Rehm and Uszkoreit, 2012; Kornai, 2013; Rehm et al., 2014, 2016a; ELRC, 2019).

Section 2 gives an overview of the ELG platform and related activities. Section 3 touches upon related work. Section 4 concludes the paper.

Towards an Interoperable Ecosystem of AI and LT Platforms: A Roadmap for the Implementation of Different Levels of Interoperability

Georg Rehm¹, Dimitrios Galanis², Penny Labropoulou², Stelios Piperidis², Martin Welß³, Ricardo Usbeck³, Joachim Köhler³, Miltos Deligiannis², Katerina Gkirtzou², Johannes Fischer⁴, Christian Chiaros⁵, Nils Feldhus¹, Julián Moreno-Schneider¹, Florian Kintzel¹, Elena Montiel⁶, Victor Rodríguez Doncel⁶, John P. McCrae⁷, David Laqua³, Irina Patricia Theile³, Christian Dittmar⁴, Kalina Bontcheva⁸, Ian Roberts⁸, Andrejs Vasiljevs⁹, Andis Lagzdins⁹

¹DFKI GmbH, Germany • ²ILSP/Athena RC, Greece • ³Fraunhofer IAIS, Germany • ⁴Fraunhofer IIS, Germany • ⁵Goethe University Frankfurt, Germany • ⁶Universidad Politécnica de Madrid, Spain • ⁷National University of Ireland Galway, Ireland • ⁸University of Sheffield, UK • ⁹Tilde, Latvia

Corresponding author: Georg Rehm – georg.rehm@dfki.de

Abstract

LREC 2020 Workshop
Language Resources and Evaluation Conference
11–16 May 2020

IWLTP 2020 – 1st International Workshop on
Language Technology Platforms

PROCEEDINGS

Edited by:

Georg Rehm, Kalina Bontcheva, Khalid Choukri,
Jan Hajič, Stelios Piperidis, Andrejs Vasiljevs

Coming up:

- European Language Grid book (2022)
- European Language Equality book (2022)
- Plus: ELG Online Documentation (next slide)

ELG Support services

The screenshot displays the European Language Grid documentation website. The main content area shows the 'Contribute an ELG compatible service' page, which includes a table of contents with sections like 'Service', 'Benchmark', 'Pipeline', and 'ANNEX 2: METADATA SCHEMA'. The sidebar menu on the left lists various chapters and sections, including 'Contribute an ELG compatible service', 'ANNEX 2: METADATA SCHEMA', 'ANNEX 3: API SPECIFICATIONS', 'ANNEX 4: PROCESSES AND POLICIES', 'ANNEX 5: PUBLICATIONS AND REPORTS', and 'ANNEX 6: OPEN CALL PROJECTS'. The main content area also features examples of services, such as 'Example 1: Information Extraction service' and 'Example 2: Machine Translation service', with links to their respective documentation pages.

- **Documentation:**
<https://european-language-grid.readthedocs.io>
 - Structured with the user in mind
 - Using, Contributing, Validating
 - Examples and technical guidance
 - Recommendations on split of metadata records, data packaging and integration of services
 - Continuously updated
- **YouTube: ELG tutorial video**
- **Schema documentation** and ready-to-use templates and examples:
<https://gitlab.com/european-language-grid/platform/ELG-SHARE-schema>
- Provide **feedback** and get help:
<https://gitlab.com/european-language-grid/platform/elg-platform>
- **Helpdesk:**
<https://www.european-language-grid.eu/contact/>




EUROPEAN LANGUAGE TECHNOLOGY
ELE ELG

European Language Technology

Fostering the European Language Technology community towards digital language equality in Europe by 2030!
Research · Brussels · 408 followers

[Following](#) [Learn more](#) [More](#)

[Home](#) [About](#) [Posts](#) [Jobs](#) [People](#) [Videos](#)

About

This is the combined channel of the EU projects European Language Equality (ELE) and European Language Grid (ELG) – together working towards a joint network of language technology for Europe’s languages and digital language equality by 2030. Follow us for project updates, events and news from the European artificial...
[see more](#)

[See all details](#)



← **European Language Technology**
74 Tweets



EUROPEAN LANGUAGE TECHNOLOGY
ELE ELG

[Edit profile](#)

European Language Technology
@EuroLangTech

The channel of the EU projects European Language Equality and European Language Grid, fostering the LT community towards digital language equality by 2030!

📍 Europe european-language-technology.eu 📅 Joined June 2021

902 Following **303** Followers



Welcome to European Language Technology!
|||ELT

[Follow ELT on Twitter](#)
[Follow ELT on LinkedIn](#)

European Language Technology is the combined communication channel for the sister projects European Language Grid (ELG) and European Language Equality (ELE) – funded by the European Commission.

For further information about either project – their goals, consortium partners, and contact details etc. – please click below:



<https://www.linkedin.com/company/european-language-technology>

<https://twitter.com/EuroLangTech>

<https://www.european-language-technology.eu>

Subscribe to our newsletter (4000+ subscribers already)



European Language Grid

Summary and Next Steps

- Establish ELG as the primary platform and marketplace for Language Technology in Europe.
- An initiative *from* the European LT community *for* the European LT community.
- European LT landscape is **highly fragmented**: ELG aims to provide just the right **umbrella platform**.
- Global market size by 2025 is enormous: we want the European LT community to be a **key player**.
- We want to **increase the visibility and reach of all members of the European LT landscape**.
- ELG is a long-term initiative: we will establish a **legal entity** for sustainability, which will operate and maintain the technology platform for the whole LT community as a joint marketplace.
- Contribute to **Digital Language Equality** in Europe by giving all our languages one virtual home and umbrella platform that collects **all** services and resources (**ELE**).  **EUROPEAN
LANGUAGE
EQUALITY**
- **Next steps**: validation of **ELG products**; attach more **data repositories**; include ELG in **relevant infrastructures** (e.g., NFDI, GAIA-X); **raise awareness** of LT for Europe and influence **policy making**; **META-FORUM 2022**; develop **ELG Release 3** (early 2022); establish **ELG legal entity** (2022).



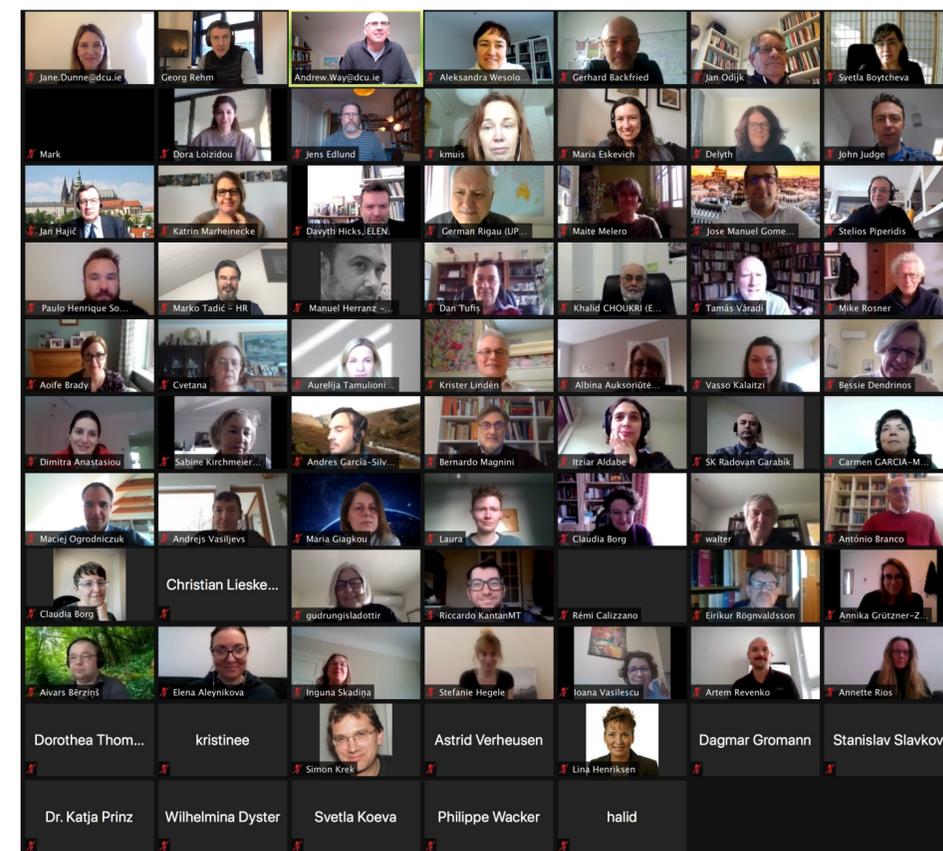
With many thanks to all speakers
and all colleagues in the ELG and
ELE teams and consortia!

Thank you!



The European Language Grid has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825627 (ELG).

The European Language Equality project has received funding from the European Union under grant agreement No LC-01641480 – 101018166 (ELE),



Prof. Dr. Georg Rehm (DFKI) – Coordinator ELG, Co-Coordinator ELE

Katrin Marheinecke (DFKI) – Project Manager ELG

24-03-2022 ELG Workshop Romania

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

<https://european-language-equality.eu>