

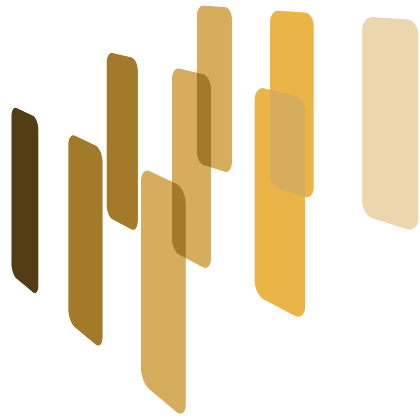
# EUROPEAN LANGUAGE GRID



**The European Cloud Platform for Language Technology,  
Natural Language Processing and Language-Centric AI**

Prof. Dr. Georg Rehm (DFKI GmbH, Germany)  
Coordinator ELG, Co-coordinator ELE

18-10-2021 ELG NCC Workshop Czech Republic & Slovakia  
<http://www.european-language-grid.eu>  
<http://www.european-language-equality.eu>

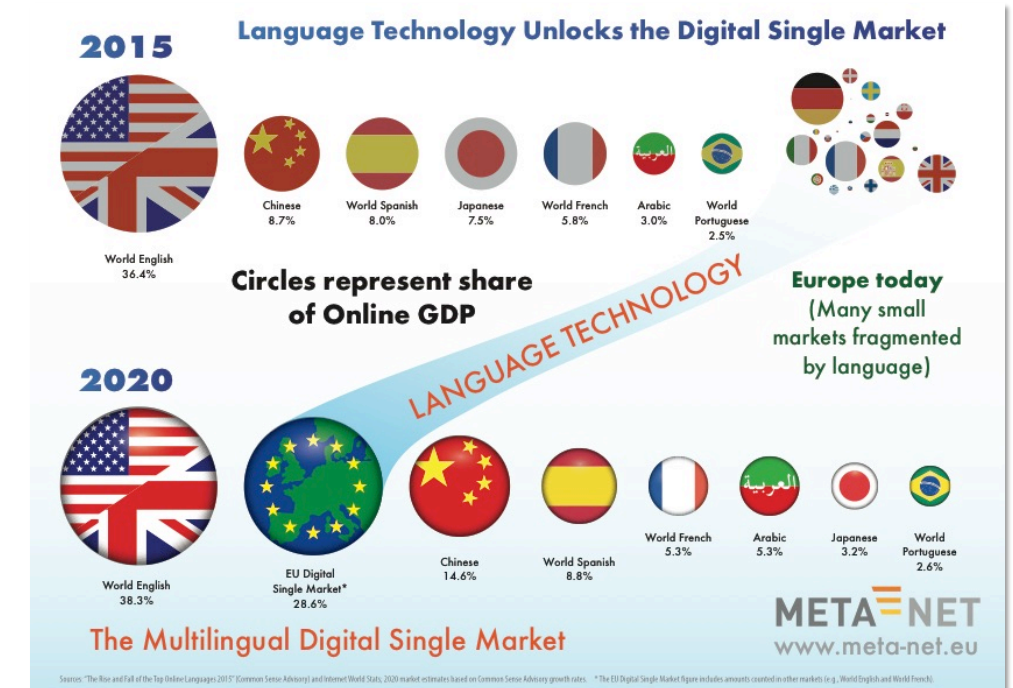
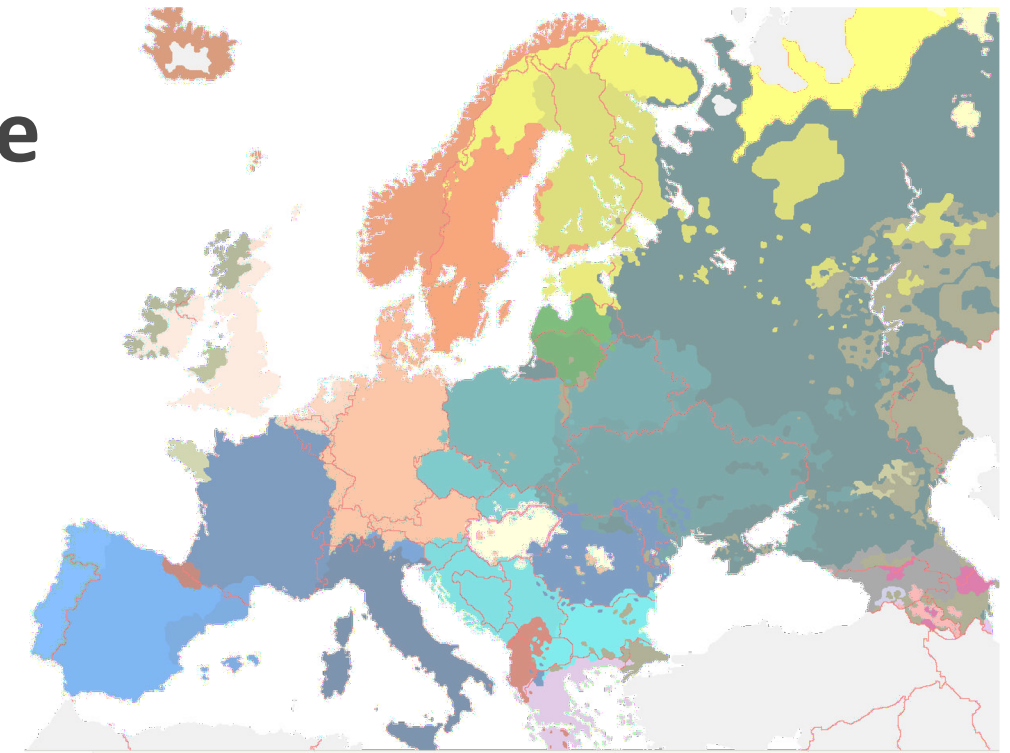


## European Language Grid

- History and context
- European Language Grid
- European Language Equality
- Summary and next steps

# 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 2010

T4ME

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

META-NET

**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 SHARE

## 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)

## META FORUM 2015



**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)

## META FORUM 2017

**"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 2011 2012 2013 2014 2015 2016 2017 2018

META-NORD

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

METANET4U

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

CESAR

**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)

**"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)

**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)

**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)



## EUROPEAN LANGUAGE GRID

2019-2022

## EUROPEAN LANGUAGE EQUALITY

2021-2022

## META FORUM 2011

## META FORUM 2013

## META FORUM 2016

ELG consortium



ELG European Language Grid

# META-NET Language White Papers “Europe’s Languages in the Digital Age”

Basque  
Bulgarian\*  
Catalan  
Croatian\*  
Czech\*  
Danish\*  
Dutch\*  
English\*  
Estonian\*  
Finnish\*  
French\*

Galician  
German\*  
Greek\*  
Hungarian\*  
Icelandic  
Irish\*  
Italian\*  
Latvian\*  
Lithuanian\*  
Maltese\*  
Norwegian

Polish\*  
Portuguese\*  
Romanian\*  
Serbian  
Slovak\*  
Slovene\*  
Spanish\*  
Swedish\*  
Welsh

\* Official EU language



Excellent support	Good support	Moderate support	Fragmentary support	Weak or no support through LT
	English	French, Spanish	Catalan, Dutch, German, Hungarian, Italian, Polish, Romanian	Basque, Bulgarian, Croatian, Czech, Danish, Estonian, Finnish, Galician, Greek, Icelandic, Irish, Latvian, Lithuanian, Maltese, Norwegian, Portuguese, Serbian, Slovak, Slovene, Swedish, Welsh
	English	Dutch, French, German, Italian, Spanish	Basque, Bulgarian, Catalan, Czech, Danish, Finnish, Galician, Greek, Hungarian, Norwegian, Polish, Portuguese, Romanian, Slovak, Slovene, Swedish	Croatian, Estonian, Icelandic, Irish, Latvian, Lithuanian, Maltese, Serbian, Welsh
	English	Czech, Dutch, Finnish, French, German, Italian, Portuguese, Spanish	Basque, Bulgarian, Catalan, Danish, Estonian, Galician, Greek, Hungarian, Irish, Norwegian, Polish, Serbian, Slovak, Slovene, Swedish	Croatian, Icelandic, Latvian, Lithuanian, Maltese, Romanian, Welsh
	English	Czech, Dutch, French, German, Hungarian, Italian, Polish, Spanish, Swedish	Basque, Bulgarian, Catalan, Croatian, Danish, Estonian, Finnish, Galician, Greek, Norwegian, Portuguese, Romanian, Serbian, Slovak, Slovene	Icelandic, Irish, Latvian, Lithuanian, Maltese, Welsh

Machine Translation

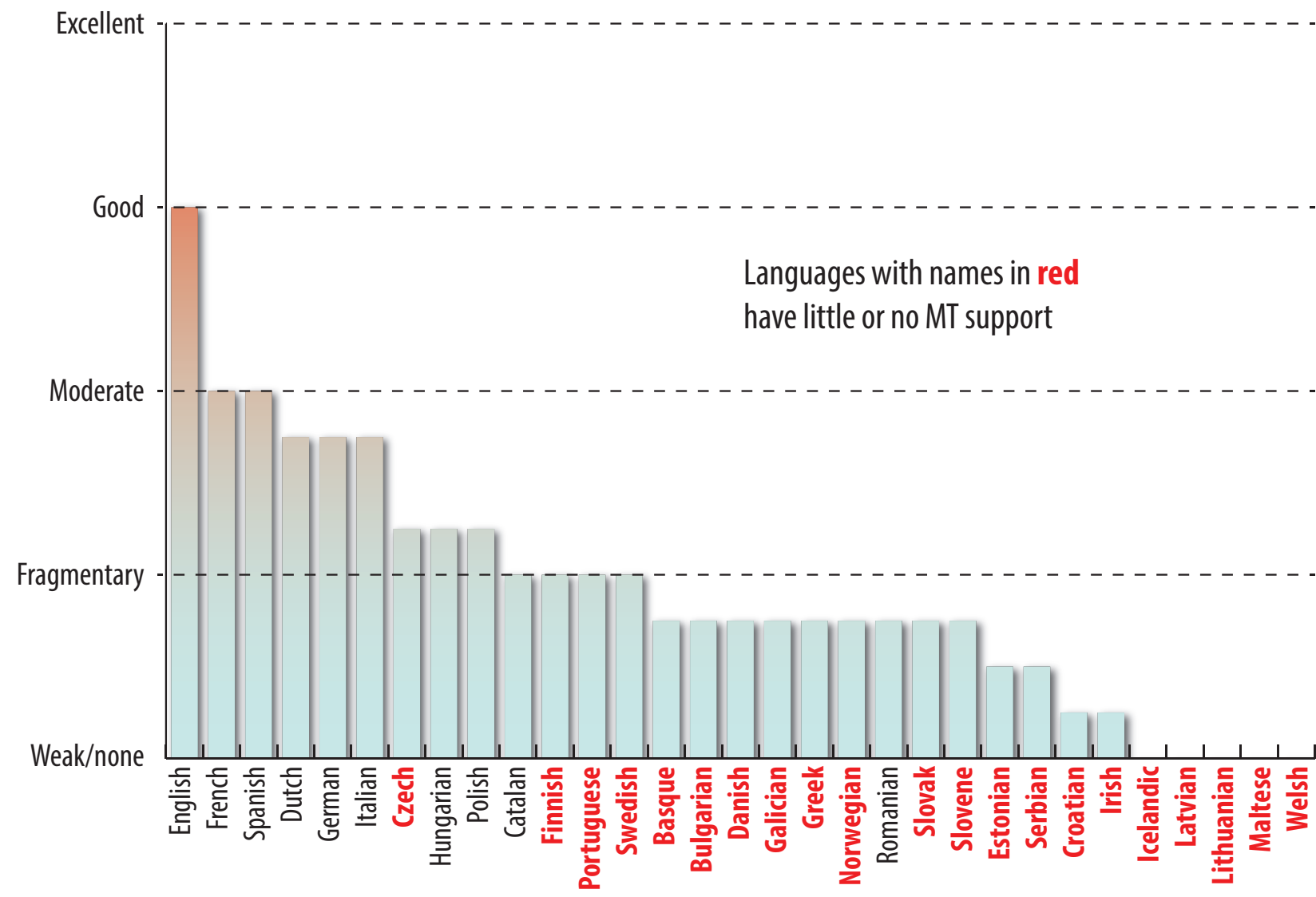
Text Analytics

Speech

Resources



Level of support

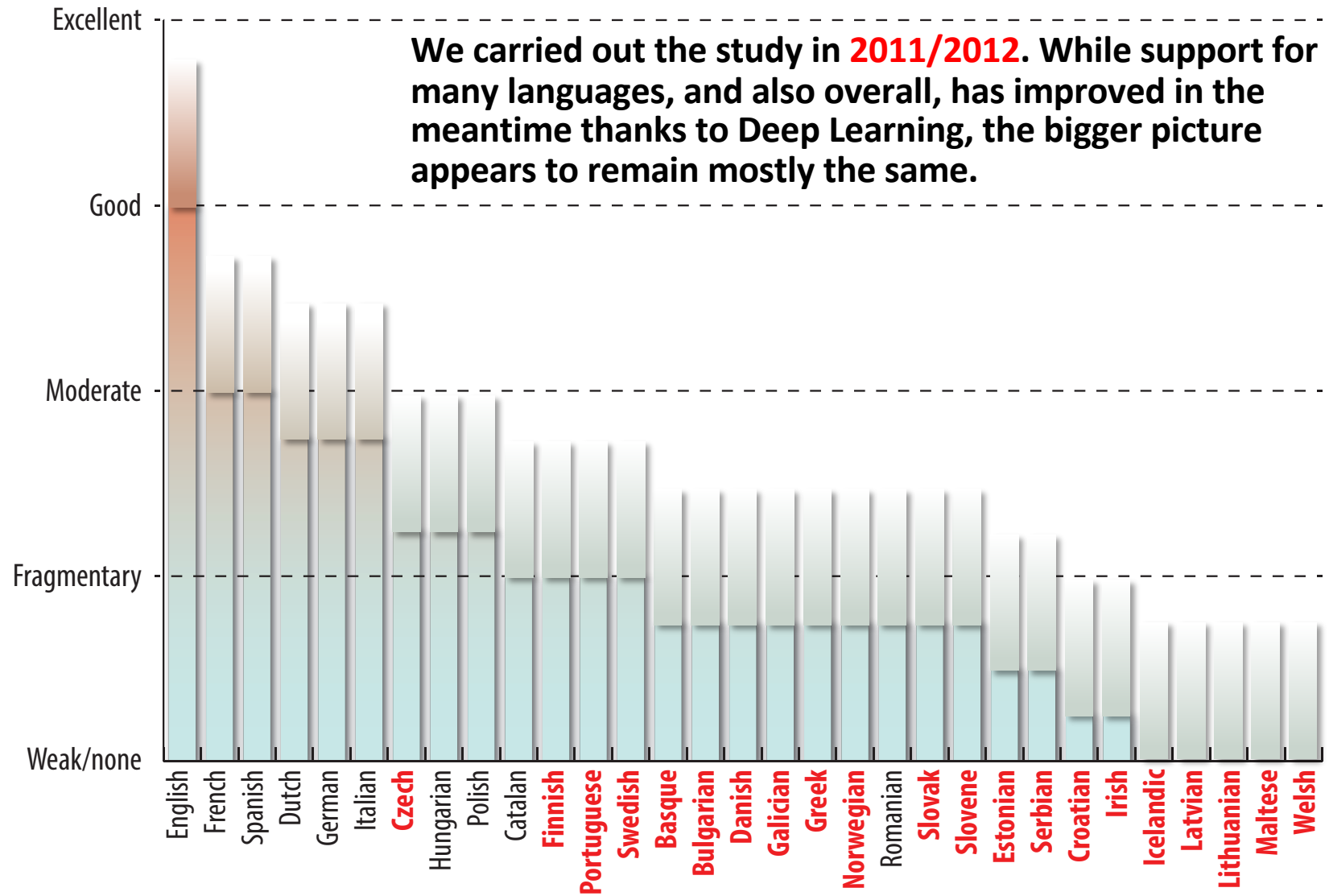


**META**  **NET**

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)



Level of support



We carried out the study in **2011/2012**. While support for many languages, and also overall, has improved in the meantime thanks to Deep Learning, the bigger picture appears to remain mostly the same.



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)



# 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,*

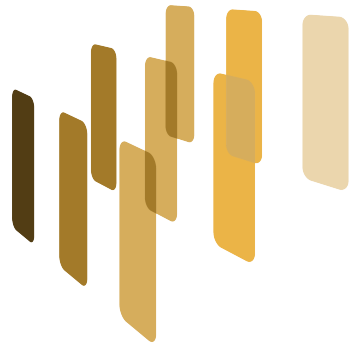
- having regard to Articles 2 and 3(3) of the Treaty on the Functioning of the European Union (TFEU),
- having regard to Articles 21(1) and 22 of the Charter of Fundamental Rights of the European Union,
- having regard to the 2003 UNESCO Convention for the Safeguarding of the Intangible Cultural Heritage,
- 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 information<sup>1</sup>,
- 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 information<sup>2</sup>,
- 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 sector<sup>3</sup>,
- having regard to the Council resolution of 21 November 2008 on a European strategy for multilingualism (2008/C 320/01)<sup>4</sup>,
- having regard to the Council decision of 3 December 2013 establishing the specific programme implementing Horizon 2020 – the Framework Programme for Research and

<sup>1</sup> OJ L 345, 31.12.2003, p. 90.

<sup>2</sup> OJ L 175, 27.6.2013, p. 1.

<sup>3</sup> OJ L 318, 4.12.2015, p. 1.

<sup>4</sup> OJ C 320, 16.12.2008, p. 1.



# EUROPEAN LANGUAGE GRID

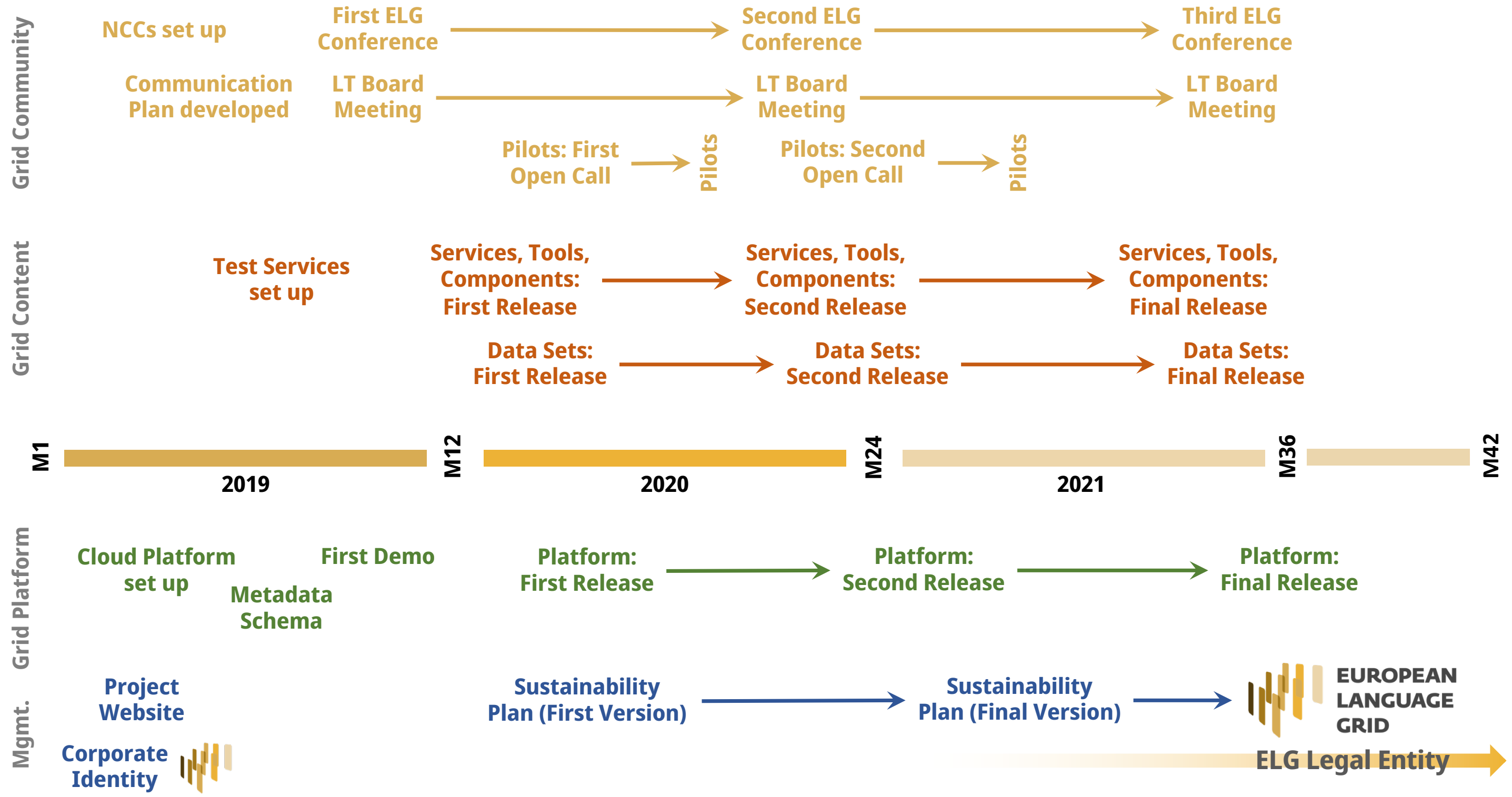
## Objectives (Selection)

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.



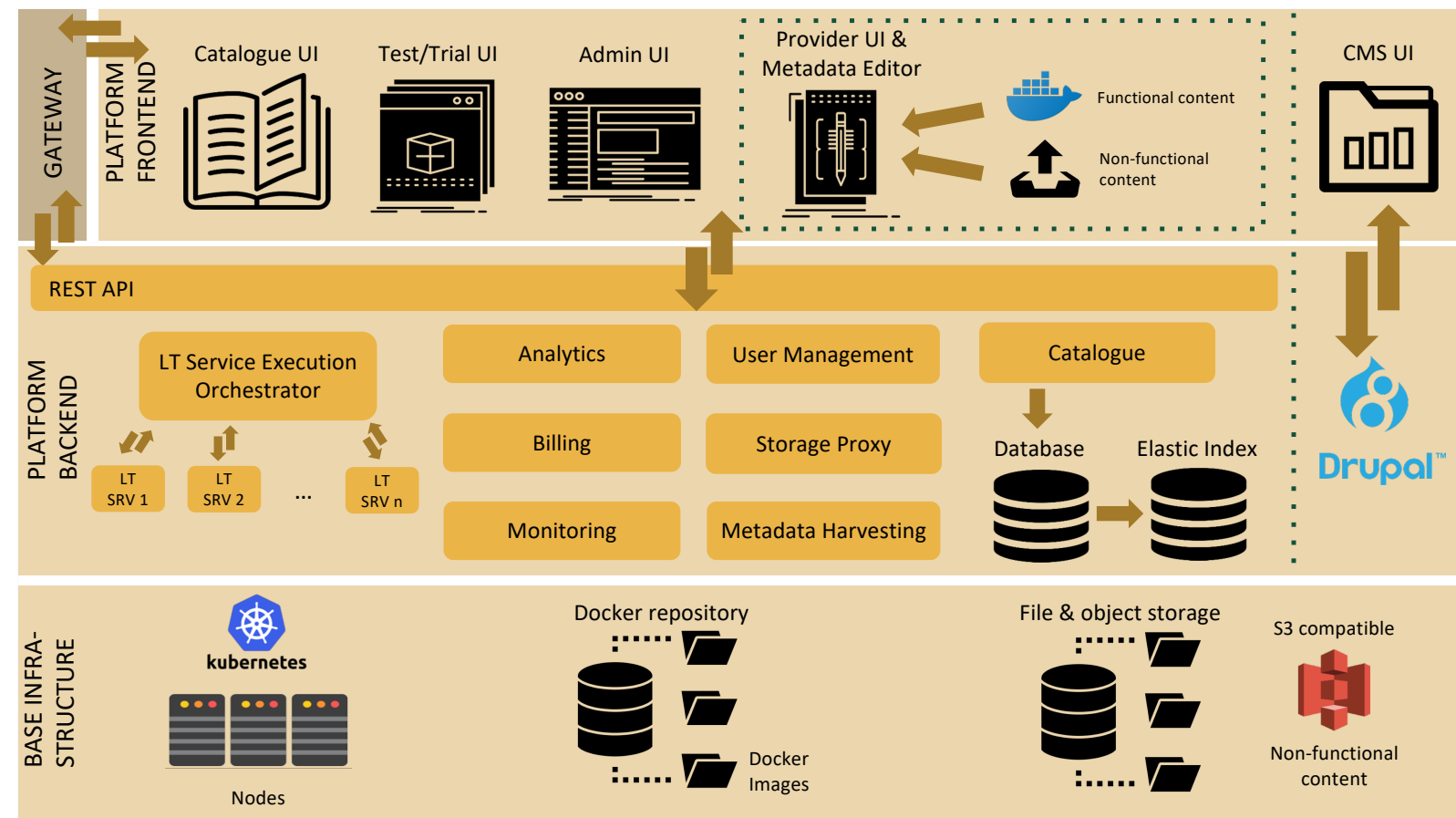
Kick-off meeting, 22/23 January 2019





# European Language Grid – Current State of Play (September 2021)

- User registration, authentication, authorisation
- User categories, respective rights and policies
- LT metadata upload
- Metadata conversion and harvesting
- LT service registration, integration
- LT service try out and execution
- LT data browsing, searching, downloading
- **Current state of play ELG Release 2 (Oct. 2021):**
  - 428 functional services and tools
  - 3500 corpora and data sets
  - 908 lexical/conceptual resources
  - 25 language descriptions, grammars, models
  - 1789 organisations (research orgs., companies)

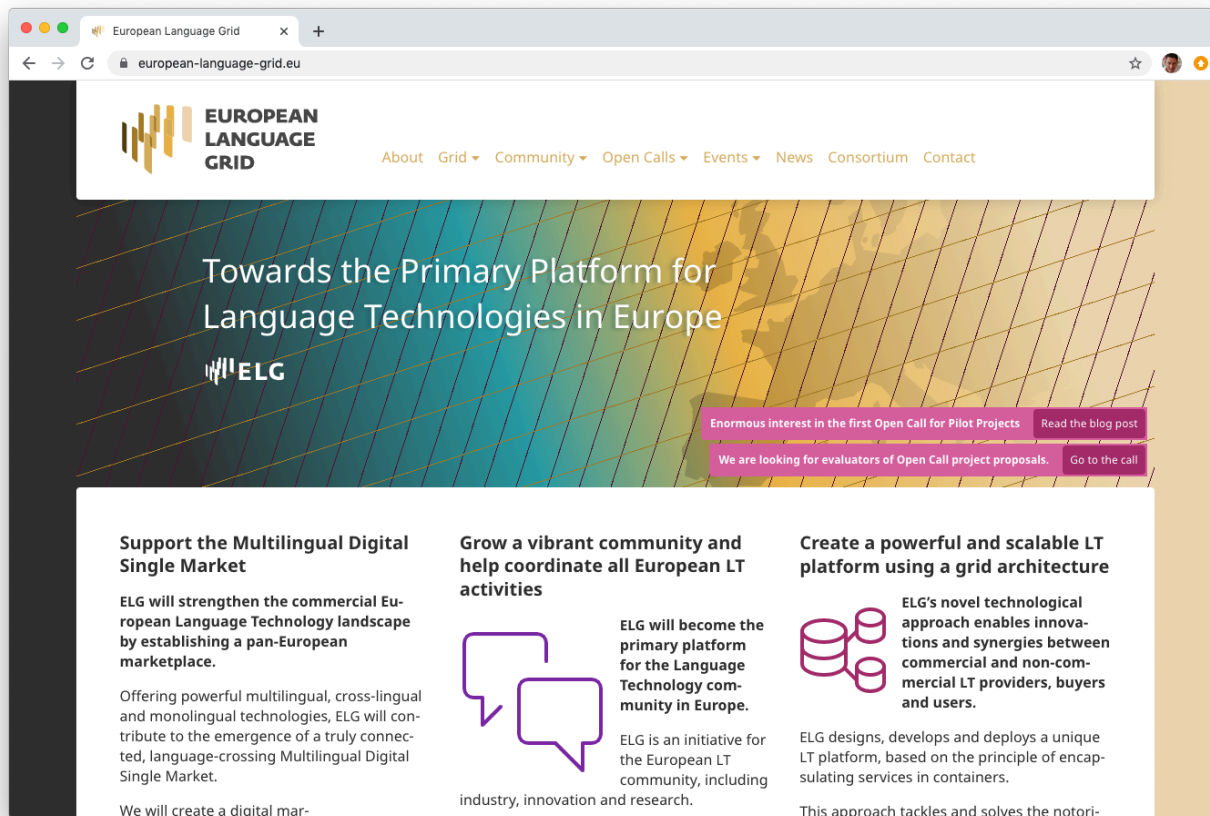


Users can connect to the ELG cloud platform via ELG APIs, remote APIs, ELG GUI, Python SDK, download of containers or source code.

ELG Release 2 made available at the end of March 2021

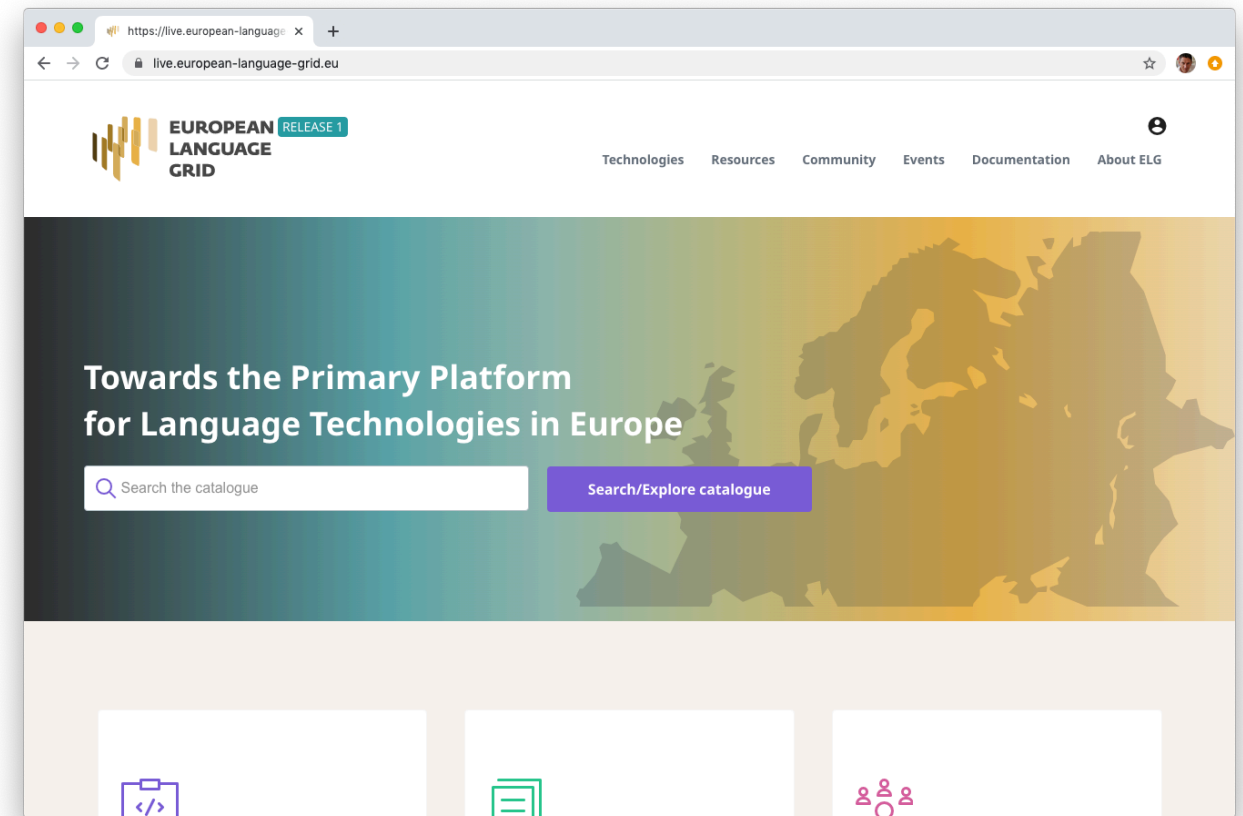
ELG Release 3 to be made available in early 2022.

# European Language Grid – Quick Demo



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

- ELG project website
- Typical information about our EU project plus a lot of additional content (events, open calls, community, NCCs etc.)
- The whole content of the project website will be integrated into the European Language Grid website (proper) in the second half of 2021



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

- The actual European Language Grid platform and website
- The next revision of the ELG project website (left) will include dynamic content that will be pulled from the ELG repository (right), e.g., languages, number of services, categories of services etc.



Georg Rehm

- Technologies
- Resources
- Community
- Events
- Documentation
- About ELG

# Towards the Primary Platform for Language Technologies in Europe

Search/Explore catalogue



### Language technologies

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

Browse Technologies



### Language data and resources

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

View Resources



### Community

Organizations, projects, events etc in European Language technology field

Explore Community



Georg Rehm

- Technologies
- Resources
- Community
- Events
- Documentation
- About ELG

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

Search

Clear all filters

48 search results

Language resources & technologies

- + Corpus (40)
- + Lexical/Conceptual resource (5)
- + Tool/Service (3)

Service functions

- + Dependency parsing (1)
- + Lemmatization (1)
- + Morphological annotation (1)
- + Part-of-Speech Tagging (1)
- + Speech Recognition (1)

Show more

Languages

- Official EU Languages
- + Spanish (33)
- + English (32)
- + Portuguese (31)
- + Bulgarian (30)
- + Czech (30)

Show more

Other EU/European languages

- Basque (48)
- + Turkish (28)
- + Catalan (27)
- + Russian (25)
- + Norwegian (24)

Show more

Other languages

**2007 CoNLL Shared Task - Basque, Catalan, Czech & Turkish** Corpus

2007 CoNLL Shared Task - Basque, Catalan, Czech & Turkish consists of dependency treebanks in four languages used as part of the CoNLL 2007 shared task on multi-lingual dependency parsing and domain adaptation. The langu ...

54 views  
2 downloads  
hosted at ELG

Keyword: corpus

Languages: Basque Czech Turkish Catalan

Licences: ELRA-END-USER-ACADEMIC-MEMBER-NONCOMMERCIALUSE-1.0  
ELRA-END-USER-COMMERCIAL-NOMEMBER-NONCOMMERCIALUSE-1.0 ...

**Annotated corpora and tools of the PARSEME Shared Task on Automatic Identification of Verbal Multiword Expressions (edition 1.1)** Corpus

This multilingual resource contains corpora in which verbal MWEs have been manually annotated. VMWEs include idioms (let the cat out of the bag), light-verb constructions (make a decision), verb-particle constructions (g ...

8 views

Keywords: Multiword expressions verbal multiword expressions light-verb constructions verb-particle constructions

Languages: English Modern Greek (1453-) Bulgarian German Polish ...

Licence: PARSEME Shared Task Data (v. 1.1) Agreement

**Annotated corpora and tools of the PARSEME Shared Task on Semi-Supervised Identification of Verbal Multiword Expressions (edition 1.2)** Corpus

This multilingual resource contains corpora in which verbal MWEs have been manually annotated, gathered at the occasion of the 1.2 edition of the PARSEME Shared Task on semi-supervised Identification of Verbal MWEs (2020 ...

3 views

Keywords: multiword expressions verbal multiword expressions light verb construction



Georg Rehm

- Technologies
- Resources
- Community
- Events
- Documentation
- About ELG

udpipe basque

Search

Clear all filters

Language resources & technologies

- + Tool/Service (62)
- + Corpus (16)
- + Lexical/Conceptual resource (4)

Service functions

- + Dependency parsing (60)
- + Lemmatization (60)
- + Morphological annotation (60)
- + Part-of-Speech Tagging (60)
- + Tokenization (60)

Show more

Languages

- Official EU Languages
- + Spanish (10)
- + Portuguese (9)
- + French (8)
- + Bulgarian (7)
- + Czech (7)

Show more

Other EU/European languages

- + Basque (22)
- + Turkish (7)
- + Catalan (5)
- + Russian (3)
- + Ukrainian (3)

Show more

85 search results for udpipe basque

UDPipe Basque: Morphosyntactic Analysis of Raw Text

UDPipe performs segmentation, tokenization, POS tagging, morphological analysis, lemmatization and dependency parsing of raw Basque texts.

Keywords: Tokenization Part of Speech tagging Lemmatisation Morphological analyser

Language: Basque

Licences: Mozilla Public License 2.0

Creative Commons Attribution Non Commercial Share Alike 4.0 International

Tool/Service

19 views

3 times used

Functional service

Basque WordNet

The Basque WordNet is a lexical database including information about Basque words. It is an extension of WordNet 1.6, a lexical database for English developed at the Princeton University. The Basque WordNet is tightly al ...

Keyword: lexicalconceptualresource

Language: Basque

Licences: ELRA-END-USER-COMMERCIAL-MEMBER-NONCOMMERCIALUSE-1.0

ELRA-END-USER-ACADEMIC-NOMEMBER-NONCOMMERCIALUSE-1.0 ...

Lexical/Conceptual resource

0 views

Elhuyar Basque ASR

Automatic Speech Recognition service for Basque

Keywords: automatic speech recognition Basque

Language: Basque

Licence: Apache License 2.0

Tool/Service

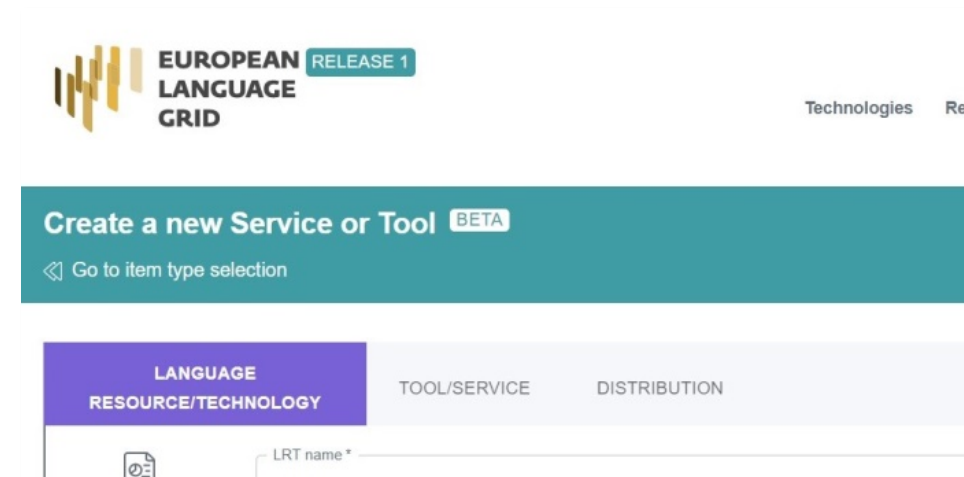
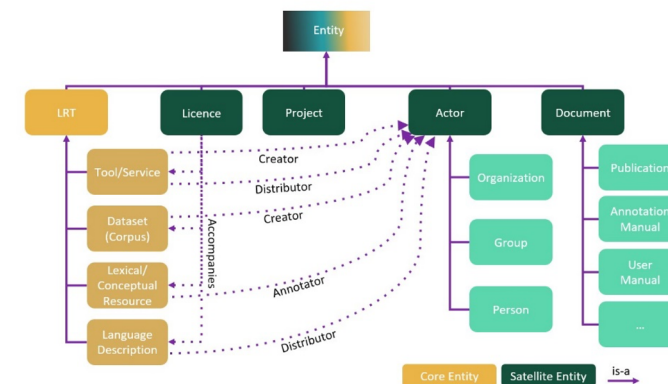
24 views

under construction



# 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>

- 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 ...

```
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}]}}
```

# 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

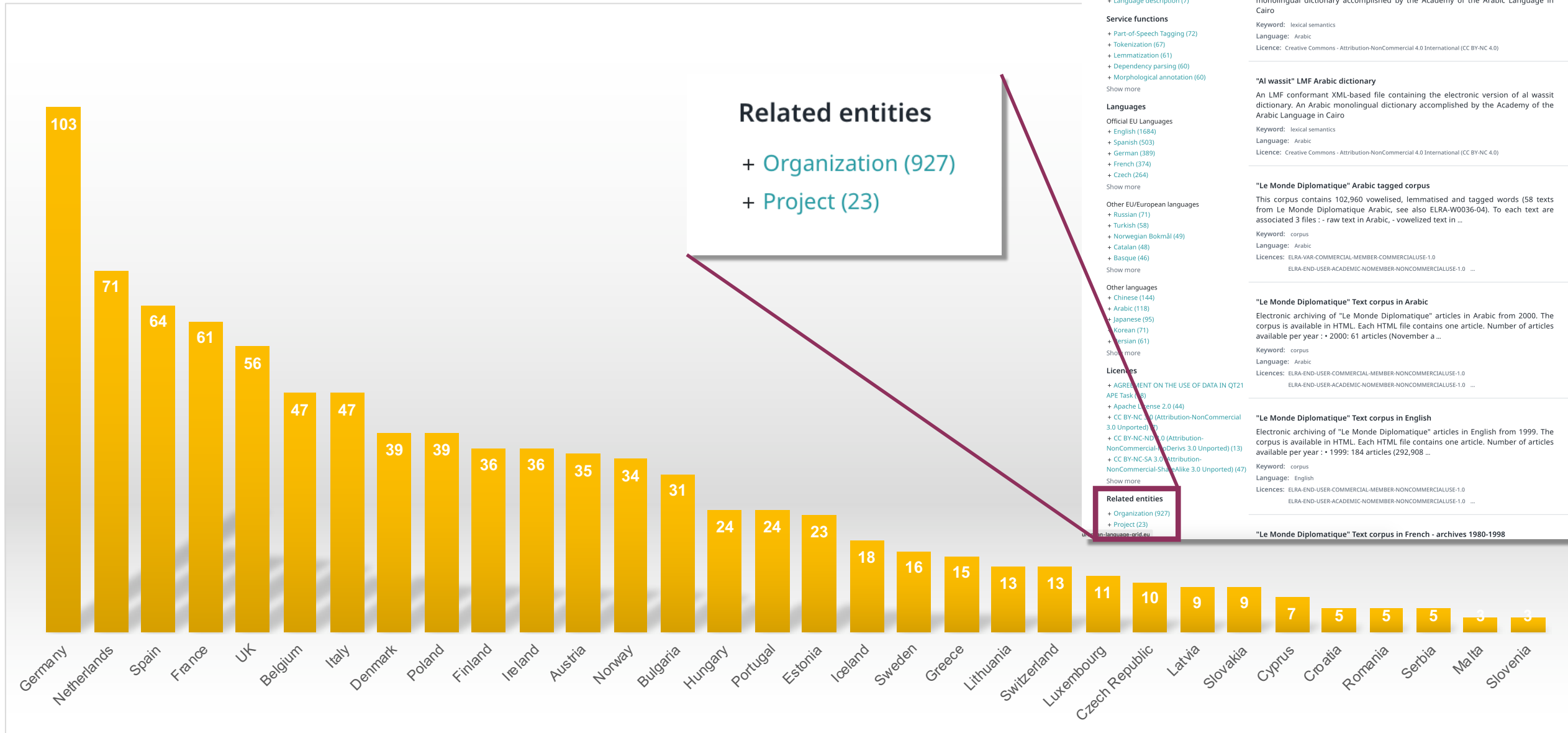
## Other organisations (e.g., NGOs) that *purchase* or *use* Language Technologies

## Funding agencies that support the development of Language Technologies



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

# European LT Industry in the ELG



**Related entities**

- + Organization (927)
- + Project (23)

Clear all filters

- Language resources & technologies
  - + Corpus (1857)
  - + Lexical/Conceptual resource (728)
  - + Tool/Service (168)
  - + Language description (7)
- Service functions
  - + Part-of-Speech Tagging (72)
  - + Tokenization (67)
  - + Lemmatization (61)
  - + Dependency parsing (60)
  - + Morphological annotation (60)
- Languages
  - Official EU Languages
    - + English (1684)
    - + Spanish (503)
    - + German (389)
    - + French (374)
    - + Czech (264)
  - Other EU/European languages
    - + Russian (71)
    - + Turkish (58)
    - + Norwegian Bokmål (49)
    - + Catalan (48)
    - + Basque (46)
  - Other languages
    - + Chinese (144)
    - + Arabic (118)
    - + Japanese (95)
    - + Korean (71)
    - + Persian (61)
- Licences
  - + AGREEMENT ON THE USE OF DATA IN QT21 APE Task (8)
  - + Apache License 2.0 (44)
  - + CC BY-NC 3.0 (Attribution-NonCommercial 3.0 Unported) (7)
  - + CC BY-NC-ND 3.0 (Attribution-NonCommercial-NoDerivs 3.0 Unported) (13)
  - + CC BY-NC-SA 3.0 (Attribution-NonCommercial-ShareAlike 3.0 Unported) (47)
- Related entities**
  - + Organization (927)
  - + Project (23)

3710 search results

- "Al wassit" Arabic dictionary**

Lexical/Conceptual resource  
114 views

An XML-based file containing the electronic version of al wassit dictionary. An Arabic monolingual dictionary accomplished by the Academy of the Arabic Language in Cairo

Keyword: lexical semantics  
Language: Arabic  
Licence: Creative Commons - Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
- "Al wassit" LMF Arabic dictionary**

Lexical/Conceptual resource  
21 views

An LMF conformant XML-based file containing the electronic version of al wassit dictionary. An Arabic monolingual dictionary accomplished by the Academy of the Arabic Language in Cairo

Keyword: lexical semantics  
Language: Arabic  
Licence: Creative Commons - Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
- "Le Monde Diplomatique" Arabic tagged corpus**

Corpus  
1 views

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-ACADEMIC-NOMEMBER-NONCOMMERCIALUSE-1.0 ...
- "Le Monde Diplomatique" Text corpus in Arabic**

Corpus  
1 views

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-COMMERCIAL-MEMBER-NONCOMMERCIALUSE-1.0 ... ELRA-END-USER-ACADEMIC-NOMEMBER-NONCOMMERCIALUSE-1.0 ...
- "Le Monde Diplomatique" Text corpus in English**

Corpus  
1 views

Electronic archiving of "Le Monde Diplomatique" articles in English from 1999. The corpus is available in HTML. Each HTML file contains one article. Number of articles available per year : • 1999: 184 articles (292,908 ...

Keyword: corpus  
Language: English  
Licences: ELRA-END-USER-COMMERCIAL-MEMBER-NONCOMMERCIALUSE-1.0 ... ELRA-END-USER-ACADEMIC-NOMEMBER-NONCOMMERCIALUSE-1.0 ...
- "Le Monde Diplomatique" Text corpus in French - archives 1980-1998**

Corpus

# Stakeholders and Users and Collaborators

ELG Open Calls: 15 pilot projects

Six ICT-29b research and innovation projects

ELE – European Language Equality

AI4EU – European AI on demand platform

Related projects: ELRC, Elexis, Lynx, MeMAD, Fandango, CEF INEA projects (MAPA, NTEU etc.) etc.

Core networks: META-NET, LT Innovate

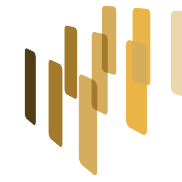
Additional initiatives: CLAIRE, AI PPP, ECSPM, EFNIL, BDVA, NPLD, CLARIN, W3C, NGI, RDA, EOSC, OpenAIRE etc., ICT-48 projects



Global Under Resourced MEdia Translation



# 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.*

**EUROPEAN  
LANGUAGE  
EQUALITY**



### 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 TADIC

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



# 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 – Save the date!*

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

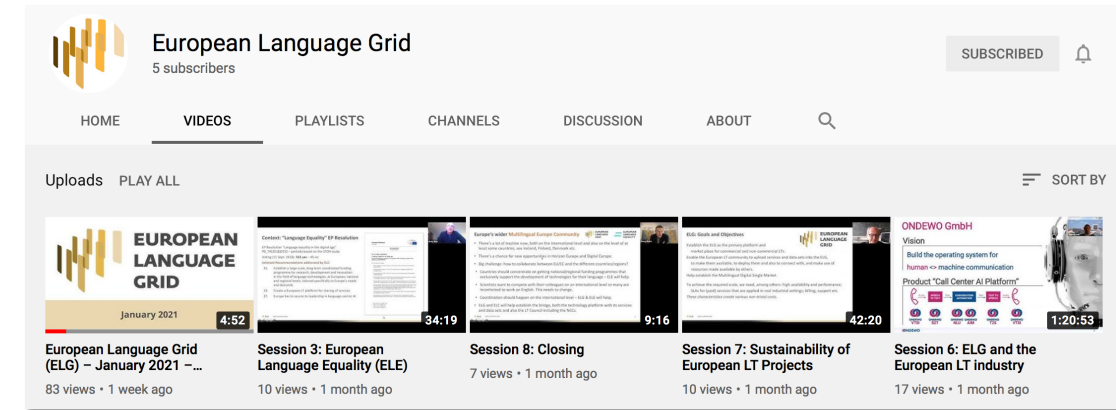
*Solutions for Multilingual Europe*

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

*Challenges for Multilingual Europe*

All META-FORUM 2020 sessions and a five-minute screencast video explaining the current state of play of the ELG platform are available on the European Language Grid YouTube channel:

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





# 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

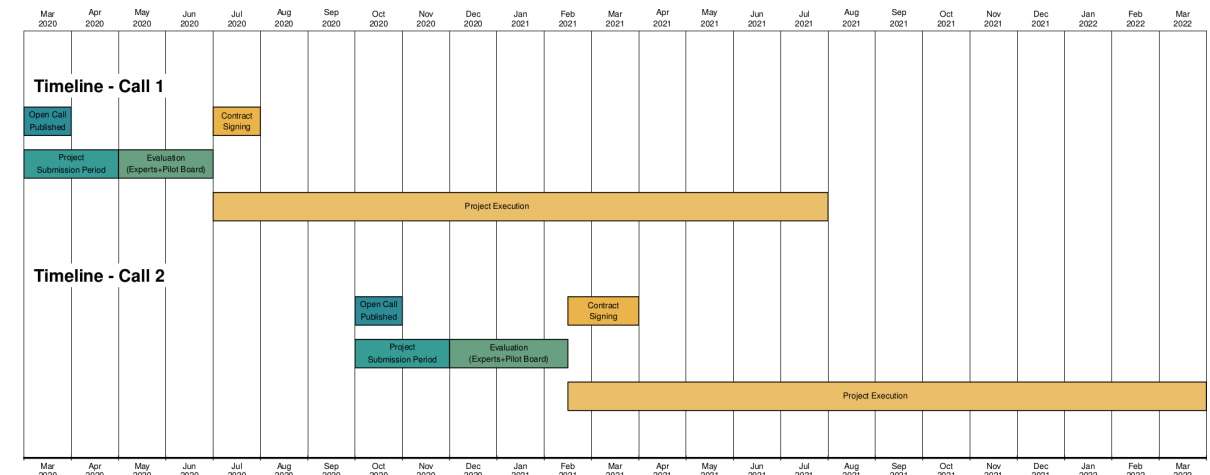
Total amount requested: 16.9M€

→ 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€

# Example: Integration of the results of the project QURATOR into the ELG platform.

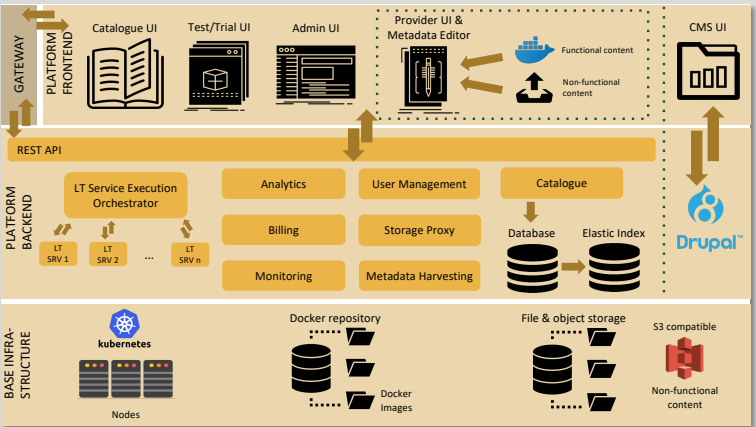
What does that mean?

- We created a page for the QURATOR project in the ELG catalogue
- We created for all QURATOR partners, too
- Two QURATOR services have already been made available through the ELG platform – their metadata records are associated with QURATOR
- More QURATOR resources will be made available in the future

**Qurator**  
Curation Technologies



**EUROPEAN LANGUAGE GRID**



Go to catalogue

**QURATOR – Curation Technologies. Flexible KI-Verfahren für die adaptive Analyse und kreative Generierung digitaler Inhalte in branchenübergreifenden Kontexten**

Overview Related entities

In all domains and sectors, the demand for intelligent systems to support the processing and generation of digital content is rapidly increasing. The availability of vast amounts of content and the pressure to publish new content quickly and in rapid succession requires faster, more efficient and smarter processing and generation methods. With a consortium of ten partners from research and industry and a broad range of expertise in AI, Machine Learning and Language Technologies, the QURATOR project, funded by the G... [Read More](#)

**Keyword**  
Language technology services  
Multilingualism  
Curation Technologies

**LT area**  
Language Technology

**Coordinator**  
German Research Center for Artificial Intelligence  
[Website](#)

- Participants**
- ART+COM AG [Website](#)
  - Condat AG [Website](#)
  - 3pc GmbH [Website](#)
  - Fraunhofer FOKUS [Website](#)
  - Semtation GmbH [Website](#)
  - Stiftung Preußischer Kulturbesitz [Website](#)
  - uberMetrics Technologies GmbH [Website](#)
  - Wikimedia Deutschland e.V. [Website](#)
  - Ada Health GmbH [Website](#)

**Project information**

[Website](#)  
Project start date  
2018-11-01  
Project end date  
2021-10-31

**Funder**  
Federal Ministry of Education and Research  
[Website](#)

**Funding scheme category**  
IA

**Funding country**  
Germany

**Funding type**  
national funds  
Grant number: Wachstumskern no.–03WKDA1A  
Status  
SIGNED  
Related call  
Wachstumskerne – Unternehmen Region  
Related programme  
Innovative regionale Wachstumskerne

Go to catalogue

**QURATOR – Curation Technologies. Flexible KI-Verfahren für die adaptive Analyse und kreative Generierung digitaler Inhalte in branchenübergreifenden Kontexten**

Overview Related entities

Funded Language Resources and Technologies

- BERTNER German v1.0.0
- BERTNER English v1.0.0

Similar integration of services or resources can be done by companies or research groups: all members of the European LT community can make their assets available via the ELG platform!

Go to catalogue

**BERTNER German BERTNER-de**

Version: v1.0.0  
Functional service

Overview Download/Run Test/Try out Code samples

The Named Entity Recognition we are using two different approaches to recognise entities: (1) based on models and (2) based on dictionaries. This module has been implemented using the Apache OpenNLP tool. The model based approach is using the NameFinderME module that is provided with a specific model (language-dependent). We have trained separate models for each entity type we distinguish. This means that we look for persons, organisations and locations serially, with three corresponding models applied consecutively... [Read More](#)

**Resource provider**  
German Research Center for Artificial Intelligence  
[Website](#)

**Funded by**  
QURATOR – Curation Technologies. Flexible KI-Verfahren für die adaptive Analyse und kreative Generierung digitaler Inhalte in branchenübergreifenden Kontexten  
[Website](#)

**Keyword**  
Named Entity Recognition  
BERT  
Language Models

**Intended application**  
Named Entity Recognition

**Input content resource**

Language  
German  
Data format  
JSON  
Processing resource type  
file  
[more](#)

**Function**

Function  
Named Entity Recognition  
Language dependent  
true

**Output resource**

Language  
German  
Data format  
JSON  
Processing resource type  
file  
[more](#)

**Resource creator** Evaluated: false

German Research Center for Artificial Intelligence  
[Website](#)

Publication date  
2019-08-23

# European Language Grid: Sustainable Operational Model & Legal Entity

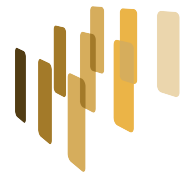
- ELG is supposed to be a long-term, sustainable initiative – a legal entity is needed.
- The technical and operational requirements – high availability and performance, SLAs, billing, support etc. – create non-trivial costs: hosting; bandwidth; ELG team; legal; etc.
- We've identified several ways and 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
- Q1/Q2 2022: Establish legal entity



# ELG's Sister Project European Language Equality (ELE)



**EUROPEAN  
LANGUAGE  
EQUALITY**



**EUROPEAN  
LANGUAGE  
GRID**

**Consortium:** 52 partners from all over Europe  
**Coordinator:** ADAPT Centre (Dublin City University)

**Co-Coordinator:** DFKI

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

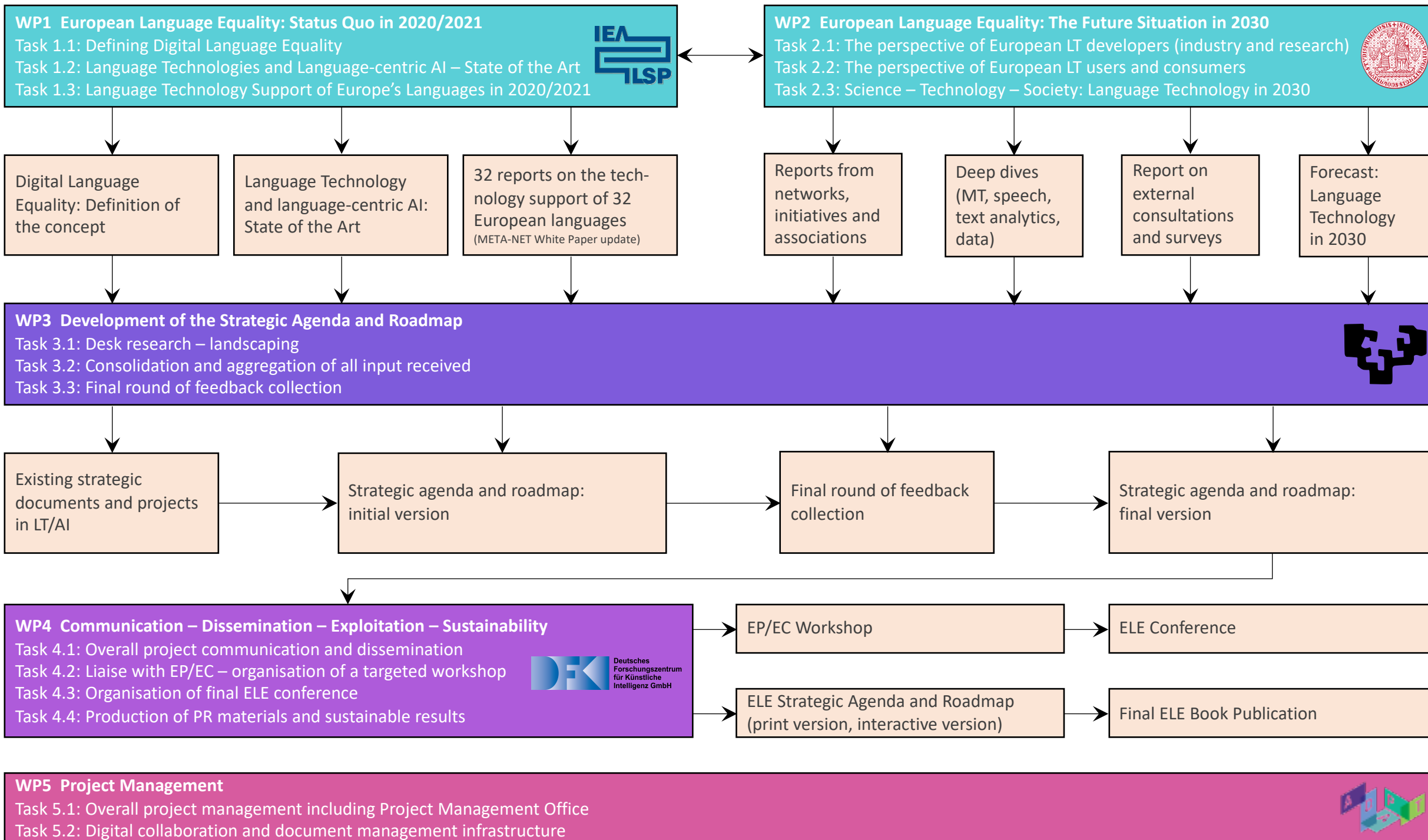
**Runtime:** 18 months – ELE and ELG will both finish up in June 2022 with a joint META-FORUM 2022

**Started on 1 January 2021**

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

We invite all language communities, all LT developing companies and all LT/NLP/Speech research teams to participate!  
Please participate in our two surveys – one for *LT developers*, one for *LT users* (URLs on the ELE website).

# Work Packages and main Deliverables



# Developing an agenda and a roadmap for achieving full digital language equality in Europe by 2030

**ELE**

Online survey for LT developers

Shape the future of Language Equality in Europe and participate now!

Online survey for LT Users and Consumers

Shape the future of Language Equality in Europe and participate now!

## Overview

With a large and all-encompassing consortium consisting of 52 partners covering all European countries, research and industry and all major pan-European initiatives, the European Language Equality (ELE) project develops a strategic research, innovation and implementation agenda as well as a roadmap for achieving full digital language equality in Europe by 2030. The

## Results

The main tangible results to be produced by the ELE project are as follows:

- The research partners will prepare updates of the 32 META-NET White Papers, which will also include the results of an enormous and exhaustive data collection procedure that will provide a detailed, empirical and dynamic map of technology

## Motivation

The ELE project is a direct response to the resolution “Language equality in the digital age”, which was passed by the European Parliament in a landslide vote – 592 votes in favour and only 45 against – in September 2018. The report provides more than 40 recommendations, including:

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



# 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**).
- **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; organise **META-FORUM 2021 and 2022**; develop **ELG Release 3** (early 2022); establish the **ELG legal entity** (late 2021).



## European Language Grid: An Overview

Georg Rehm<sup>1</sup>, Maria Berger<sup>1</sup>, Ela Elsholz<sup>1</sup>, Stefanie Hegele<sup>1</sup>, Florian Kintzel<sup>1</sup>, Katrin Marheinecke<sup>1</sup>, Stelios Piperidis<sup>2</sup>, Miltos Deligiannis<sup>2</sup>, Dimitrios Galanis<sup>2</sup>, Katerina Gkirtzou<sup>2</sup>, Penny Labropoulou<sup>2</sup>, Kalina Bontcheva<sup>3</sup>, David Jones<sup>3</sup>, Ian Roberts<sup>3</sup>, Jan Hajič<sup>4</sup>, Jana Hamrlová<sup>4</sup>, Lukáš Kačena<sup>4</sup>, Khalid Choukri<sup>5</sup>, Victoria Arranz<sup>5</sup>, Andrejs Vasiljevs<sup>6</sup>, Orians Anvari<sup>6</sup>, Andis Lagzdins<sup>6</sup>, Jūlija Melņika<sup>6</sup>, Gerhard Backfried<sup>7</sup>, Erinc Dikici<sup>7</sup>, Miroslav Janosik<sup>7</sup>, Katja Prinz<sup>7</sup>, Christoph Prinz<sup>7</sup>, Severin Stampler<sup>7</sup>, Dorothea Thomas-Aniola<sup>7</sup>, José Manuel Gómez Pérez<sup>8</sup>, Andres Garcia Silva<sup>8</sup>, Cristian Berrio<sup>8</sup>, Ulrich Germann<sup>9</sup>, Steve Renals<sup>9</sup>, Ondrej Klejch<sup>9</sup>

<sup>1</sup>DFKI GmbH, Germany • <sup>2</sup>ILSP/Athens RC, Greece • <sup>3</sup>University of Sheffield, UK • <sup>4</sup>Charles University, Czech Republic • <sup>5</sup>Evaluations and Language Resources Distribution Agency (ELDA), France • <sup>6</sup>Tilde, Latvia • <sup>7</sup>SAIL LABS Technology GmbH, Austria • <sup>8</sup>Expert System Iberia SL, Spain • <sup>9</sup>University 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 Labropoulou<sup>1</sup>, Katerina Gkirtzou<sup>1</sup>, Maria Gavrilidou<sup>1</sup>, Miltos Deligiannis<sup>1</sup>, Dimitrios Galanis<sup>1</sup>, Stelios Piperidis<sup>1</sup>, Georg Rehm<sup>2</sup>, Maria Berger<sup>2</sup>, Valérie Mapelli<sup>2</sup>, Mickaël Rigault<sup>2</sup>, Victoria Arranz<sup>2</sup>, Khalid Choukri<sup>2</sup>, Gerhard Backfried<sup>2</sup>, José Manuel Gómez Pérez<sup>2</sup>, Andres Garcia Silva<sup>2</sup>

<sup>1</sup>ILSP/Athens RC, Greece, <sup>2</sup>DFKI GmbH, Germany, <sup>3</sup>Evaluations and Language Resources Distribution Agency (ELDA), France, <sup>4</sup>SAIL LABS Technology GmbH, Austria, <sup>5</sup>Expert System Iberia SL, Spain  
{penny.gkirtzou, maria.mdcl, galanisd, 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 Cloud<sup>1</sup> or the Google dataset search service<sup>2</sup>, 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 Market<sup>3</sup> 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) project<sup>4</sup> 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

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 platform<sup>5</sup> 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

<sup>5</sup>The 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).

<sup>1</sup><https://www.eosc-portal.eu>

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

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

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

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

Georg Rehm<sup>1</sup>, Katrin Marheinecke<sup>1</sup>, Stefanie Hegele<sup>1</sup>, Stelios Piperidis<sup>2</sup>, Kalina Bontcheva<sup>3</sup>, Jan Hajič<sup>4</sup>, Khalid Choukri<sup>5</sup>, Andrejs Vasiljevs<sup>6</sup>, Gerhard Backfried<sup>7</sup>, Christoph Prinz<sup>7</sup>, José Manuel Gómez Pérez<sup>8</sup>, Luc Meertens<sup>9</sup>, Paul Lukowicz<sup>9</sup>, Josef van Genabith<sup>9</sup>, Andrea Lösch<sup>9</sup>, Philipp Sussallek<sup>9</sup>, Morten Irgens<sup>10</sup>, Patrick Gatellier<sup>11</sup>, Joachim Kühler<sup>12</sup>, Laure Le Bars<sup>13</sup>, Dimitra Anastasiou<sup>14</sup>, Albina Auksoirūtė<sup>15</sup>, Núria Bel<sup>16</sup>, António Branco<sup>17</sup>, Gerhard Budin<sup>18</sup>, Walter Daelemans<sup>19</sup>, Koenraad De Smedt<sup>20</sup>, Radovan Garabik<sup>21</sup>, Maria Gavrilidou<sup>22</sup>, Dagmar Gromann<sup>23</sup>, Svetla Koeva<sup>22</sup>, Simon Krek<sup>24</sup>, Cvetana Krsteva<sup>24</sup>, Krister Lindén<sup>25</sup>, Bernardo Magnin<sup>26</sup>, Jan Odijk<sup>27</sup>, Maciej Ogródniczak<sup>28</sup>, Eirikur Rignvaldsson<sup>29</sup>, Mike Rosner<sup>30</sup>, Bolette Sandford Pedersen<sup>31</sup>, Inguna Skadiņa<sup>32</sup>, Marko Tadić<sup>33</sup>, Dan Tufiş<sup>34</sup>, Tamás Váradi<sup>35</sup>, Kadri Vider<sup>36</sup>, Andy Way<sup>37</sup>, François Yvon<sup>38</sup>

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

### European Language Grid: A Joint Platform for the European Language Technology Community

Georg Rehm<sup>1</sup>, Stelios Piperidis<sup>2</sup>, Kalina Bontcheva<sup>3</sup>, Jan Hajič<sup>4</sup>, Victoria Arranz<sup>5</sup>, Andrejs Vasiljevs<sup>6</sup>, Gerhard Backfried<sup>7</sup>, José Manuel Gómez-Pérez<sup>8</sup>, Ulrich Germann<sup>9</sup>, Rémi Calizzano<sup>1</sup>, Nils Feldhus<sup>1</sup>, Stefanie Hegele<sup>1</sup>, Florian Kintzel<sup>1</sup>, Katrin Marheinecke<sup>1</sup>, Julián Moreno-Schneider<sup>1</sup>, Dimitrios Galanis<sup>2</sup>, Penny Labropoulou<sup>2</sup>, Miltos Deligiannis<sup>2</sup>, Katerina Gkirtzou<sup>2</sup>, Athanasia Kolovou<sup>2</sup>, Dimitris Gkoumas<sup>2</sup>, Leon Voukoutis<sup>2</sup>, Ian Roberts<sup>3</sup>, Jana Hamrlová<sup>4</sup>, Dušan Variš<sup>4</sup>, Lukáš Kačena<sup>4</sup>, Khalid Choukri<sup>5</sup>, Valérie Mapelli<sup>5</sup>, Mickaël Rigault<sup>5</sup>, Jūlija Melņika<sup>6</sup>, Miroslav Jánošík<sup>7</sup>, Katja Prinz<sup>7</sup>, Andrés García-Silva<sup>8</sup>, Cristian Berrio<sup>8</sup>, Ondrej Klejch<sup>9</sup>, Steve Renals<sup>9</sup>

<sup>1</sup>DFKI GmbH, Berlin, Germany; <sup>2</sup>R. C. Athena, ILSP, Greece; <sup>3</sup>University of Sheffield, UK; <sup>4</sup>Charles University, Czechia; <sup>5</sup>ELDA, France; <sup>6</sup>Tilde, Latvia; <sup>7</sup>HENSOLDT Analytics, Austria; <sup>8</sup>Expert System, Spain; <sup>9</sup>University 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.<sup>1</sup> This scalable system is targeted to evolve into the primary

<sup>1</sup><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 ELG (2019-2022), which will establish a legal entity in early 2022, the platform will provide to approx. 1300 commercial and non-commercial tools and 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 to deploy them through the grid. This is also meant to support *digital language equality* 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 all 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 Rehm<sup>1</sup>, Dimitrios Galanis<sup>2</sup>, Penny Labropoulou<sup>2</sup>, Stelios Piperidis<sup>2</sup>, Martin Wolf<sup>3</sup>, Ricardo Usbeck<sup>3</sup>, Joachim Köhler<sup>3</sup>, Miltos Deligiannis<sup>2</sup>, Katerina Gkirtzou<sup>2</sup>, Johannes Fischer<sup>4</sup>, Christian Chiarcos<sup>5</sup>, Nils Feldhus<sup>1</sup>, Julián Moreno-Schneider<sup>1</sup>, Florian Kintzel<sup>1</sup>, Elena Montiel<sup>6</sup>, Victor Rodríguez Doncel<sup>6</sup>, John P. McCrae<sup>7</sup>, David Laguna<sup>8</sup>, Irina Patricia Theille<sup>8</sup>, Christian Dittmar<sup>8</sup>, Kalina Bontcheva<sup>8</sup>, Ian Roberts<sup>8</sup>, Andrejs Vasiljevs<sup>8</sup>, Andis Lagzdins<sup>9</sup>

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

Coming up:

- European Language Grid book (2022)
- European Language Equality book (2022)

Plus: ELG Online Documentation

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

**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**  
@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

[Edit profile](#)

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>

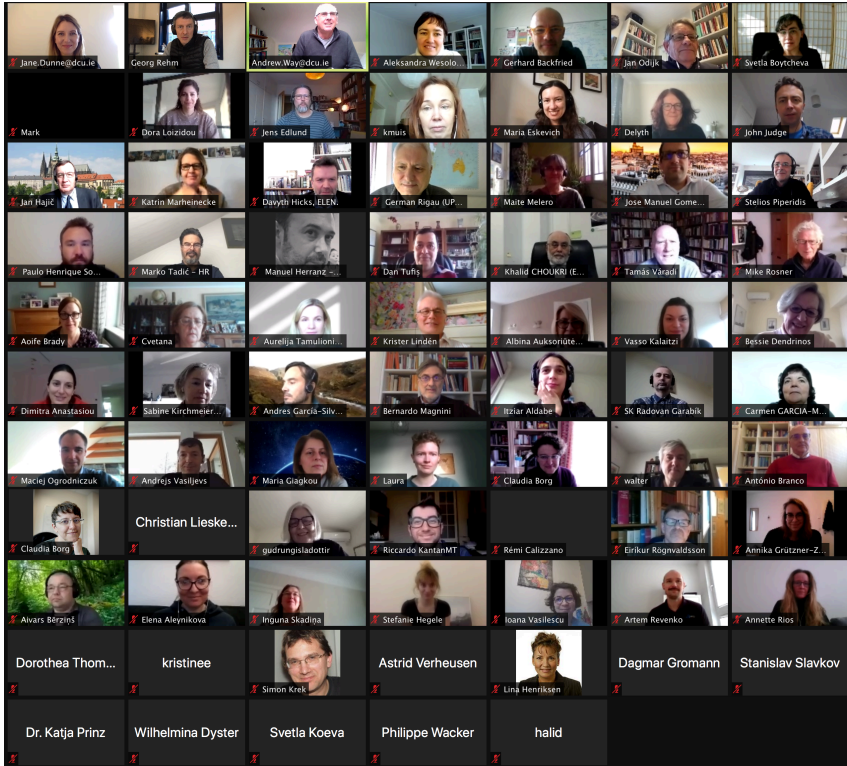
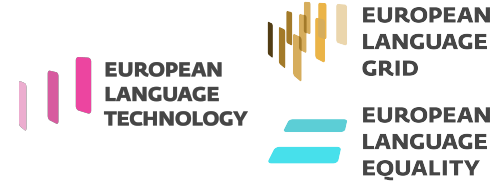
<https://www.european-language-technology.eu>



Not covered at all in this presentation: how to provide services or resources – slides can be provided, a YouTube tutorial video is currently being produced.

With many thanks to 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 № 825627 (ELG).

Prof. Dr. Georg Rehm (DFKI GmbH, Germany)
Coordinator ELG, Co-coordinator ELE
18-10-2021 ELG NCC Workshop Czech Republic & Slovakia
http://www.european-language-grid.eu
http://www.european-language-equality.eu