

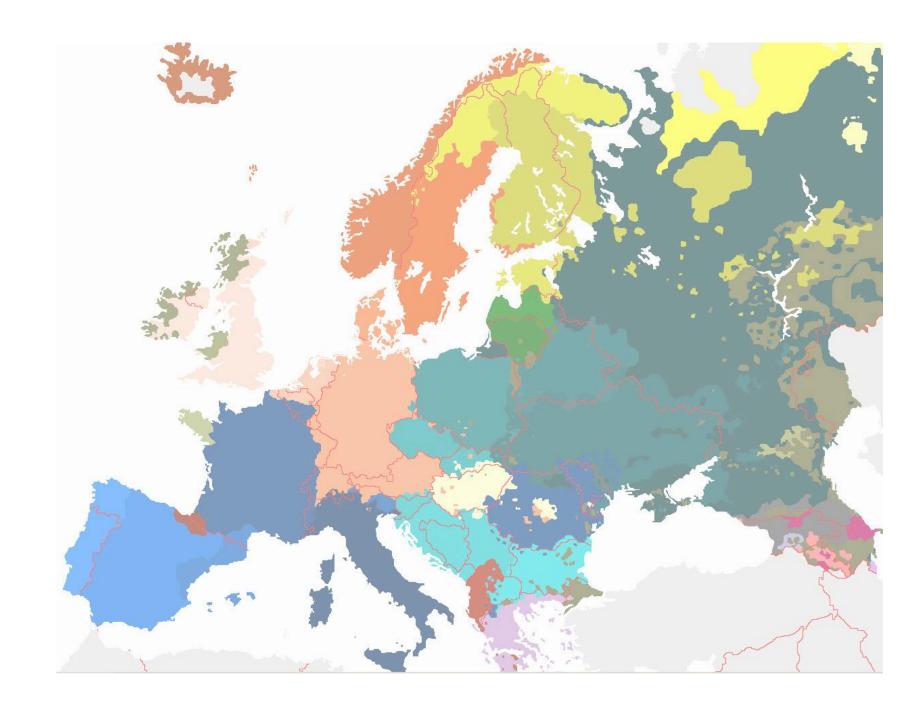
# The European Language Grid and European Language Equality Projects

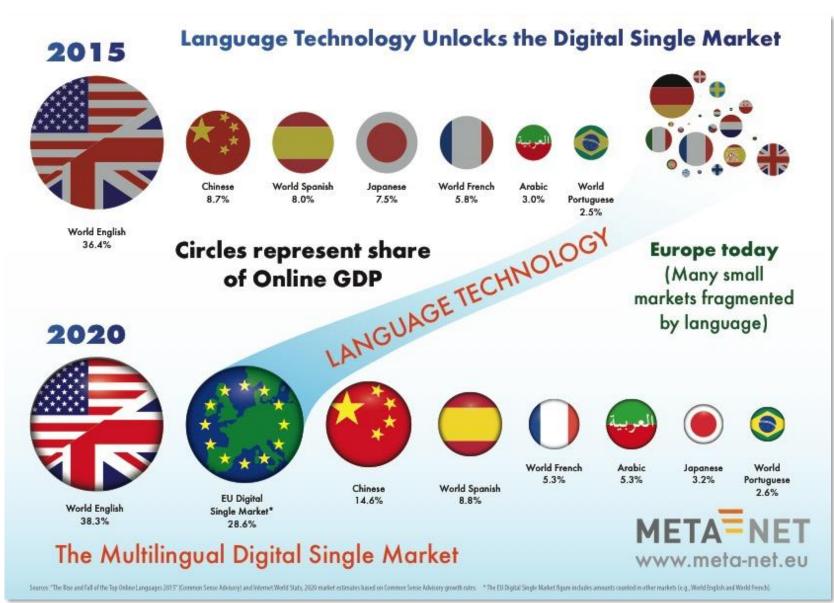
Penny Labropoulou, Athena RC/ILSP 27-05-2022 ELG NCC Workshop Slovenia

http://www.european-language-grid.eu http://www.european-language-equality.eu

### Multilingualism in Europe

- Multilingualism is at the heart of the European idea
- 24 EU languages they all have the same status
- Dozens of 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 (including availability of LTs and LRs)











#### T4ME

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



2010

### **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 2010**

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



### 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)



### **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)





### **EUROPEAN LANGUAGE EQUALITY**

2021-2022

2018

### **EUROPEAN LANGUAGE GRID**

2019-2022

#### 2017 2013 2015

### 2012

#### **META-NORD**

EU-funded (ICT PSP)

Project to establish an open linguistic infrastructure in the Baltic and Nordic countries (2011 – 2013)

2011



META NORD

#### **METANET4U**

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



#### **CESAR**

EU-funded project (ICT PSP) 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)



### 2014

### "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**

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**

(Version 0.9)

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









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





















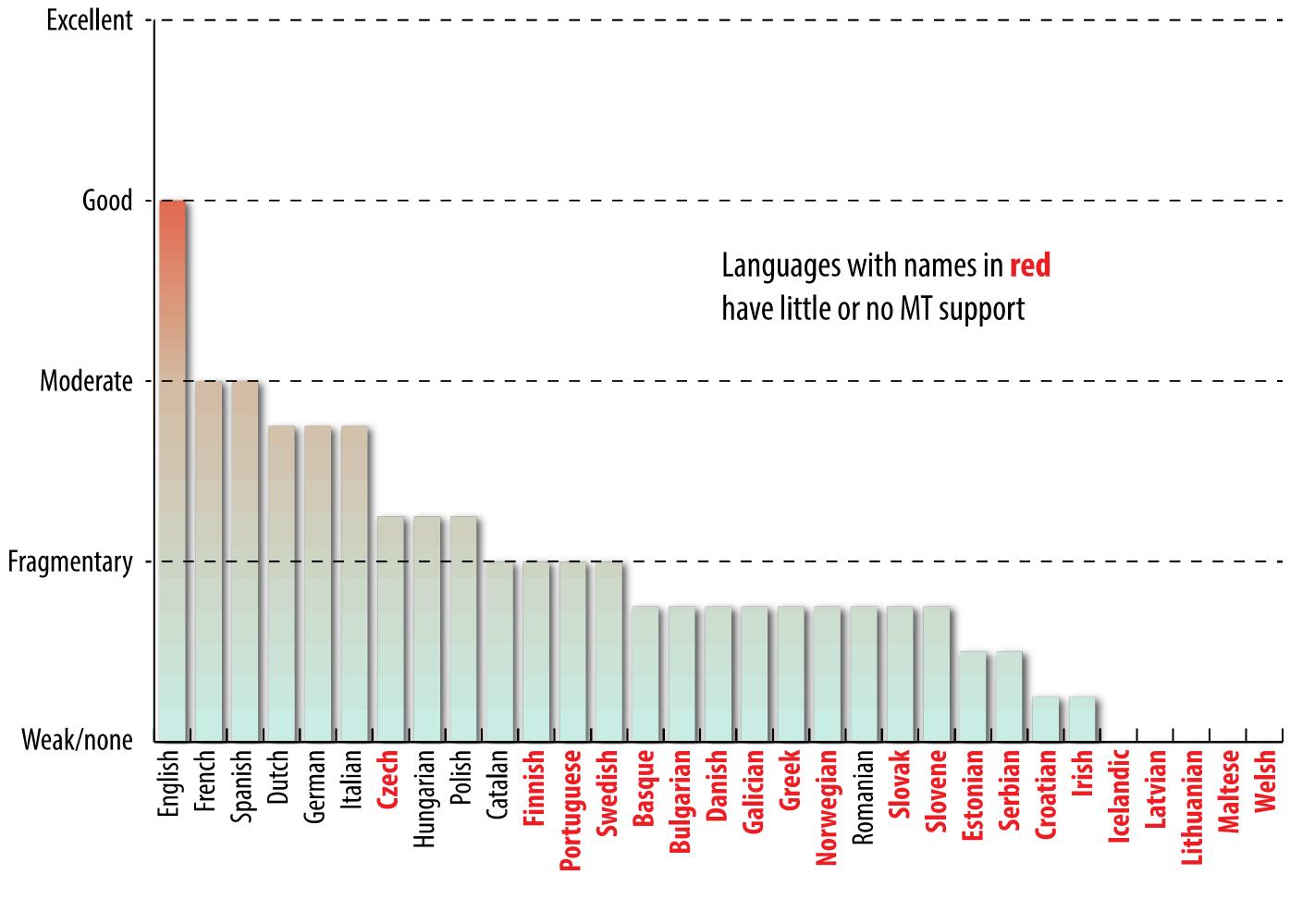














Source: META-NET White Paper Series: Europe's Languages in the Digital Age. Springer, Heidelberg, New York, Dordrecht, London, September 2012. Georg Rehm and Hans Uszkoreit (series editors)

## "Language Equality" EP 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 Al
- 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
- 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

OJ L 345, 31.12.2003, p. 90.

OJ L 175, 27.6.2013, p. 1.

OJ L 318, 4.12.2015, p. 1.
 OJ C 320, 16.12.2008, p. 1.

## European LT Market – CEF Study

### **Approach and Observations**

- Background: SMART 2016/0103
   contract: contribute to CEF AT as
   "multilingualism enabler" for CEF DSIs.
- European LT vendors grouped per type of tech: Translation, Speech, Search, NLU, Analytics
- EU market approx. 1B€ in 2020 disrupted by dominant global players
- SMEs: 70% of EU LT vendors up to 50 employees
- Revenue per company is growing
- Market is highly fragmented: hundreds of SMEs, many address very specific niches, sectors and languages

### Recommendations

- Europe is strong in R&I, but not successful to scale innovations and capture the market
- Europe needs European alternatives to fill the gaps and deficiencies and to avoid reliance on monopolies
- Multilingual DSM should be developed on its own infrastructure
- Public procurement can be the major driver for European LT industry to avoid dependence on monopolies
- Plans needed to avoid brain drain
- A platform is needed to connect demand and supply as well as industry and research
- **CEF** study estimated the European LT Market to reach 1B€ in 2020.
- ➤ Numbers obsolete, market is going through the roof thanks to A!!















The Global Natural Language Processing Market size is expected to reach \$29.5 billion by 2025, rising at a market growth of 20.5% CAGR during the forecast period

Globe Newswire FOLLOW

January 14, 2020 4:26am Comments

>Truly incredible market forecast!

(Source: Global Newswire)





### 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.
- 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.











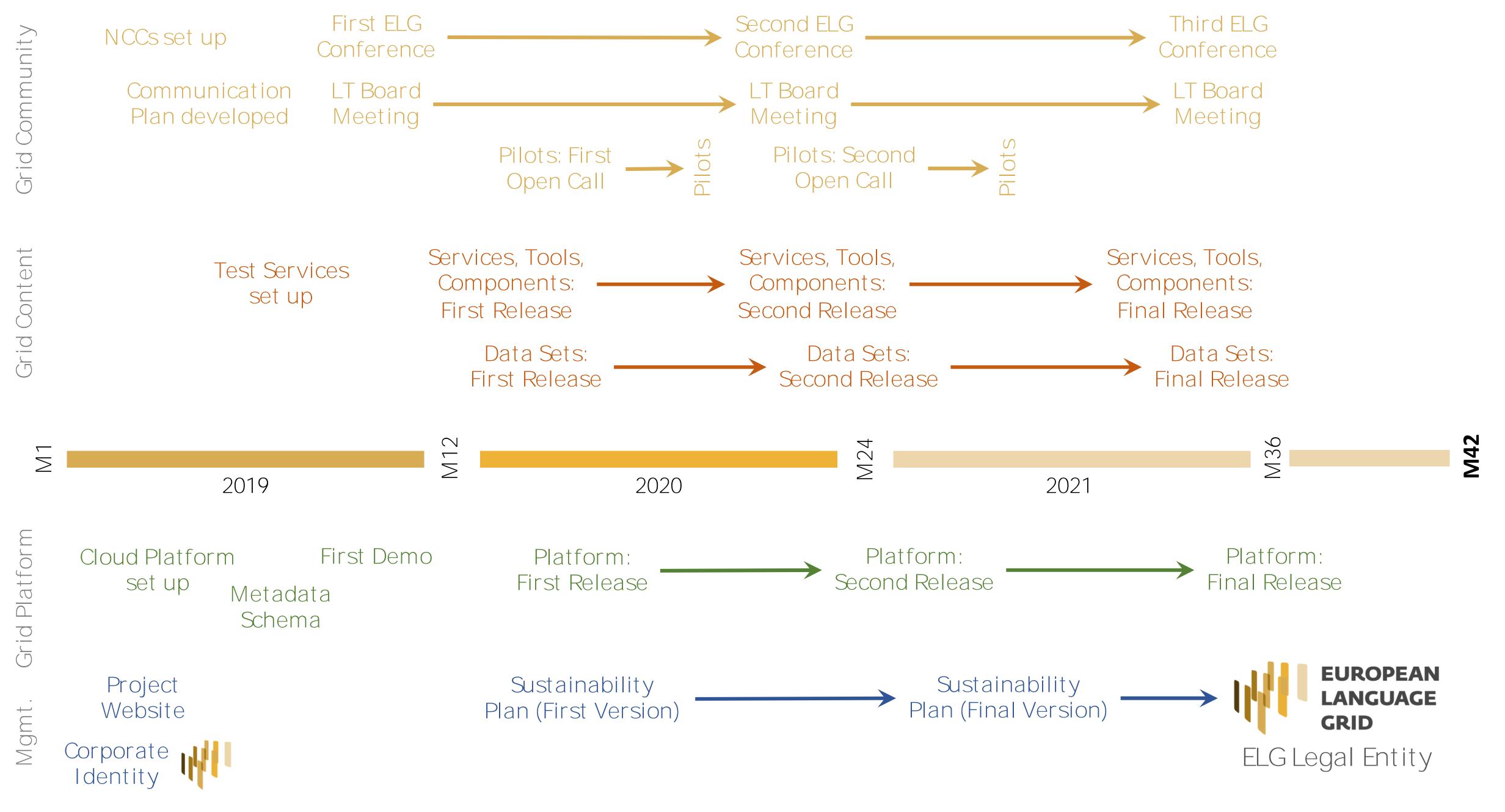




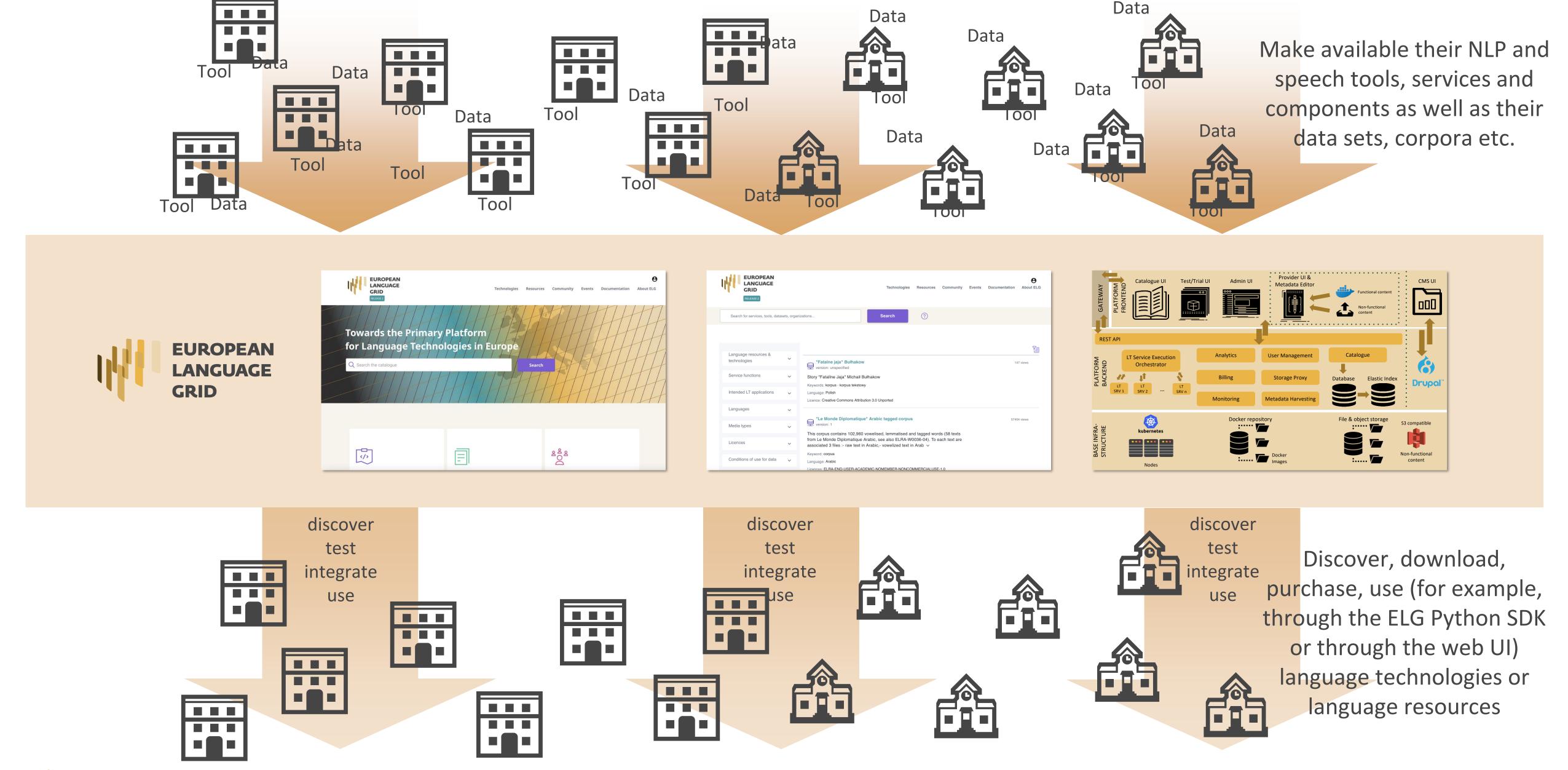


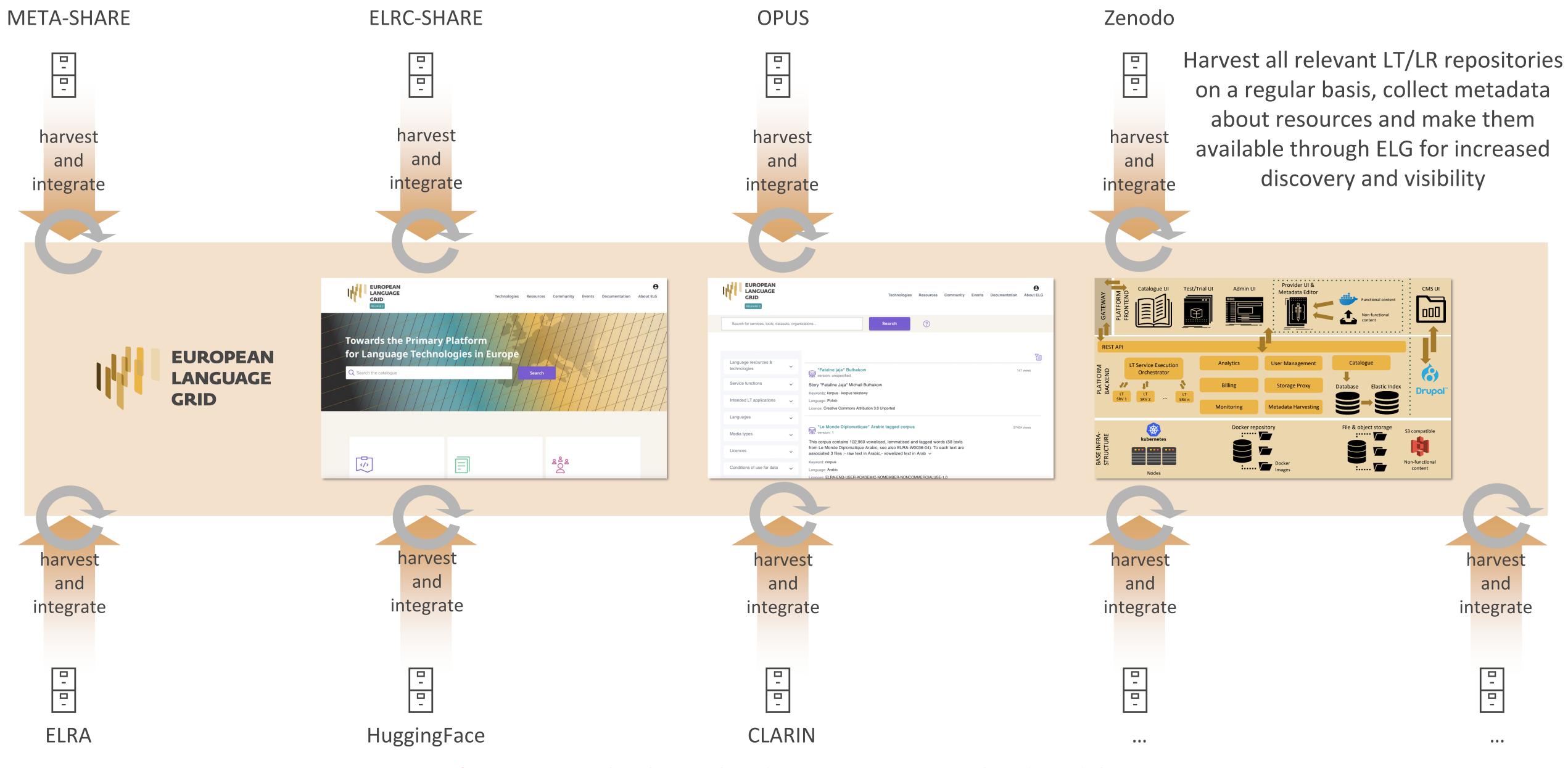






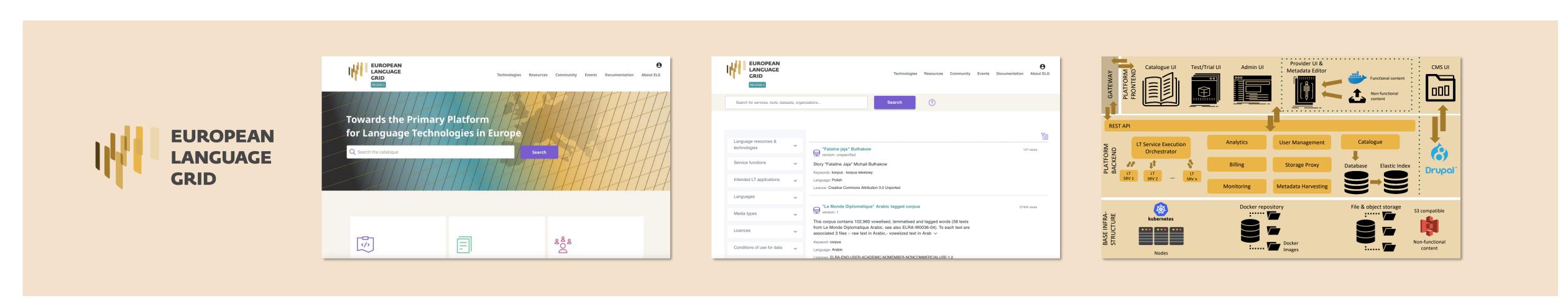
### Developers of Language Technologies: Companies, Universities, Research Centres (approx. 1800 organisations in total)





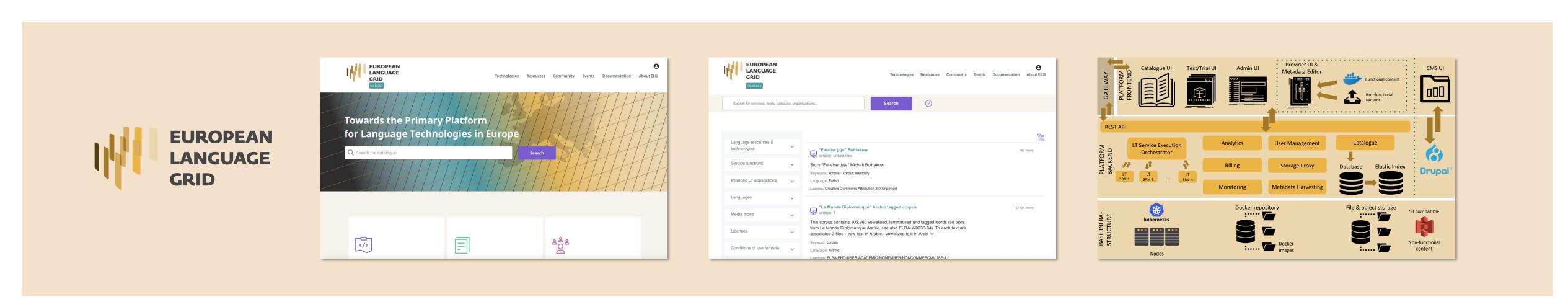
Harvesting of Language Technology-related Resource, Data, Tool and Model Repositories

₩ ELG



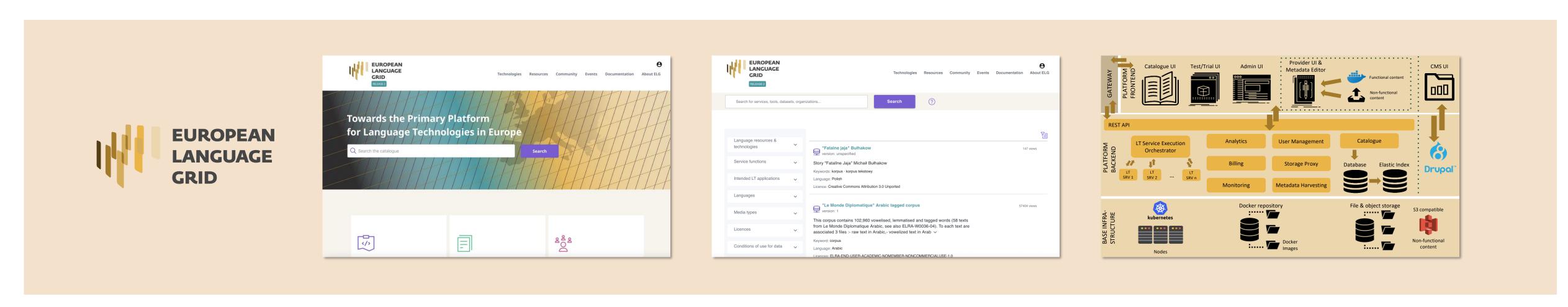
European Language Grid is a joint technology platform for the whole European Language Technology community (approx. 1800 organisations)





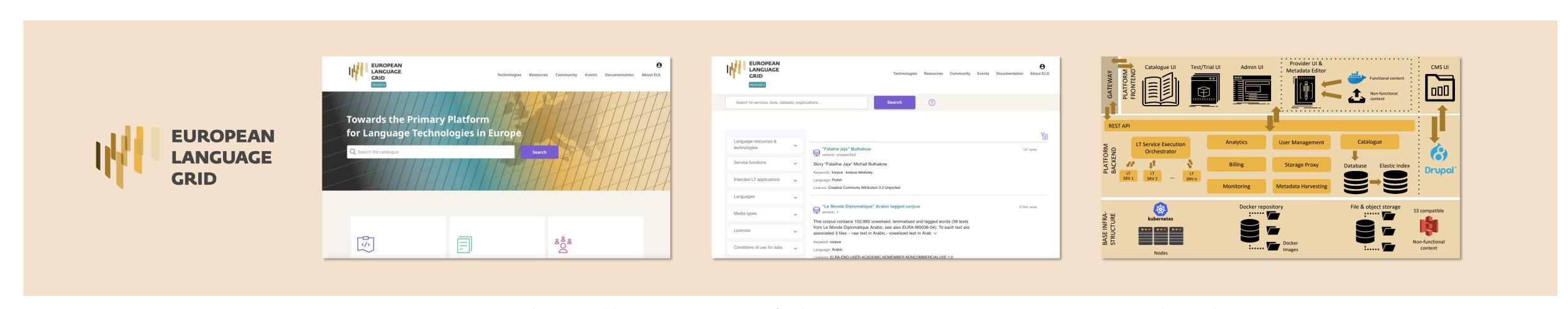
European Language Grid is a joint data, tool and resource sharing platform





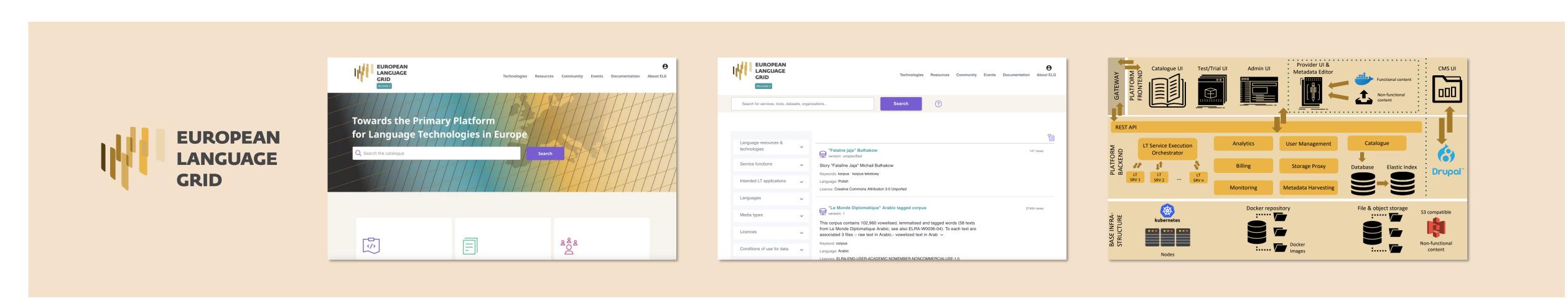
European Language Grid is a marketplace for the whole European Language Technology community





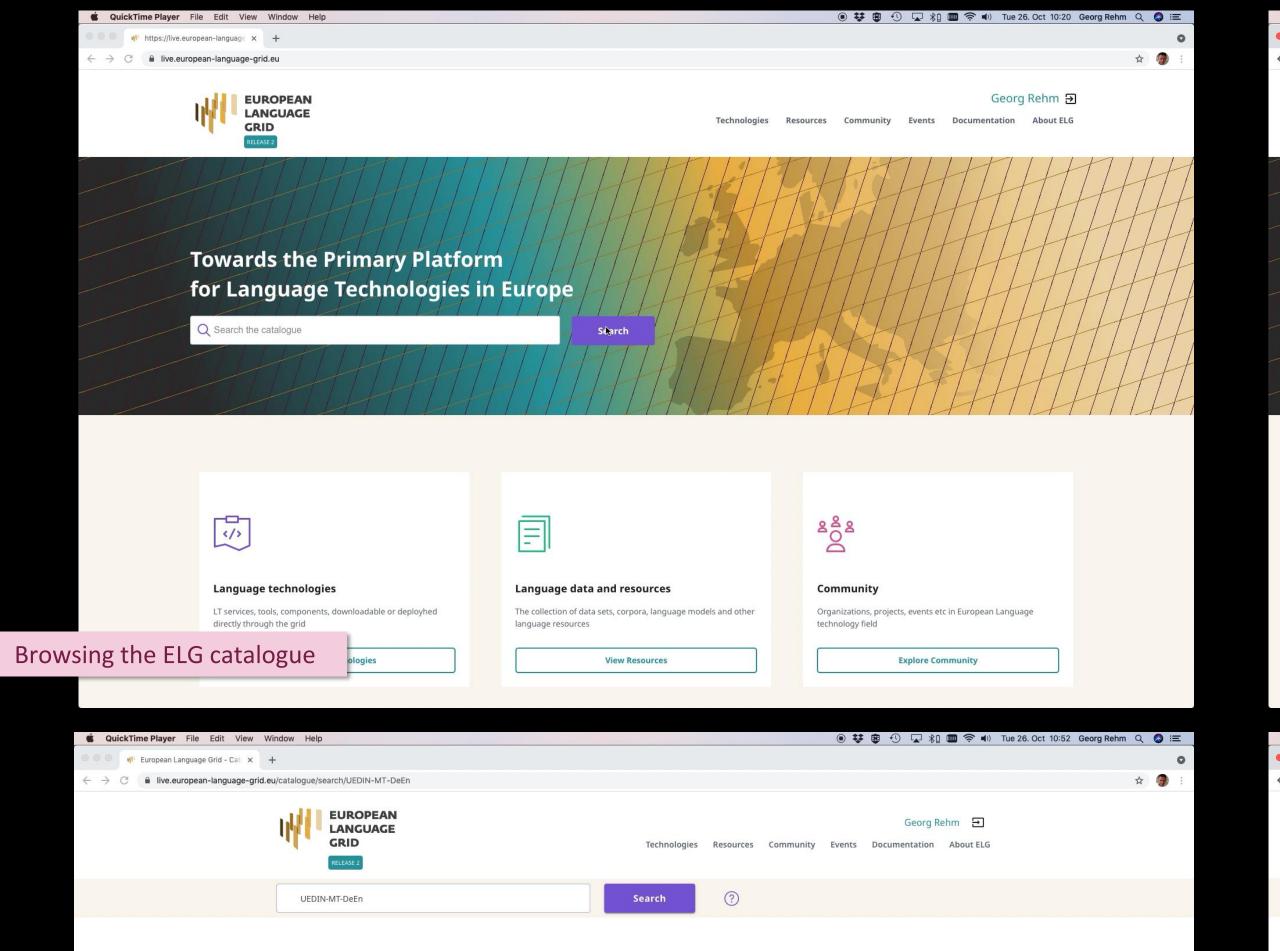
European Language Grid is the yellow pages of the European Language Technology community

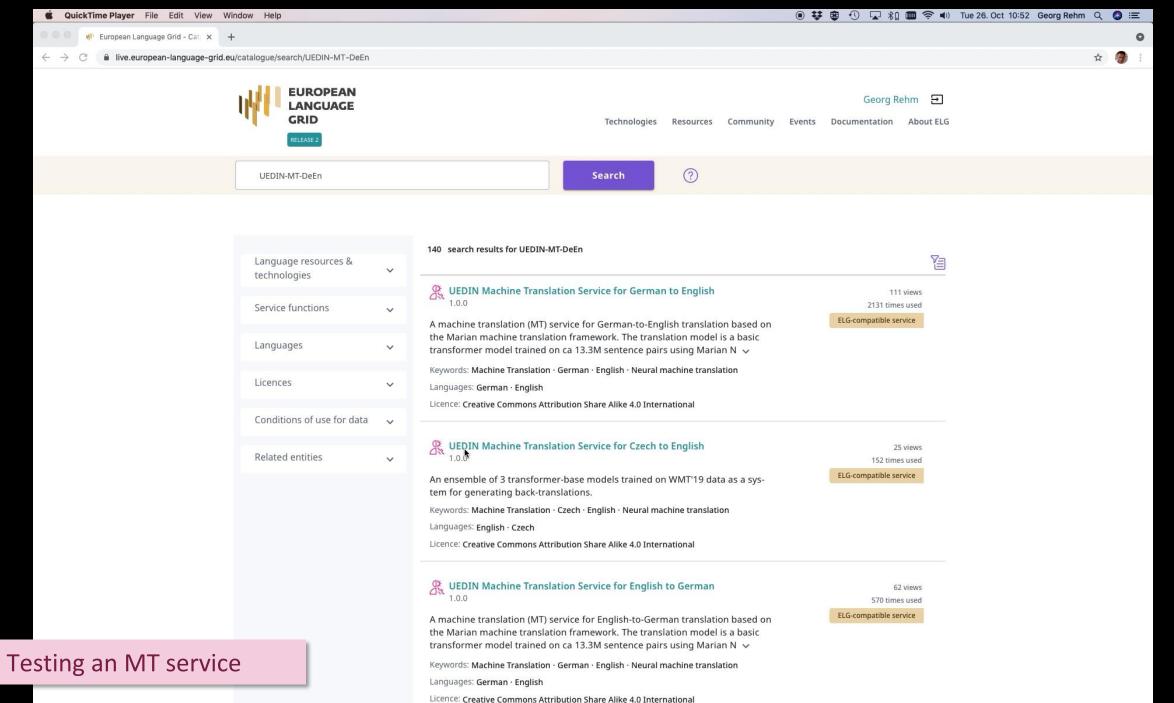


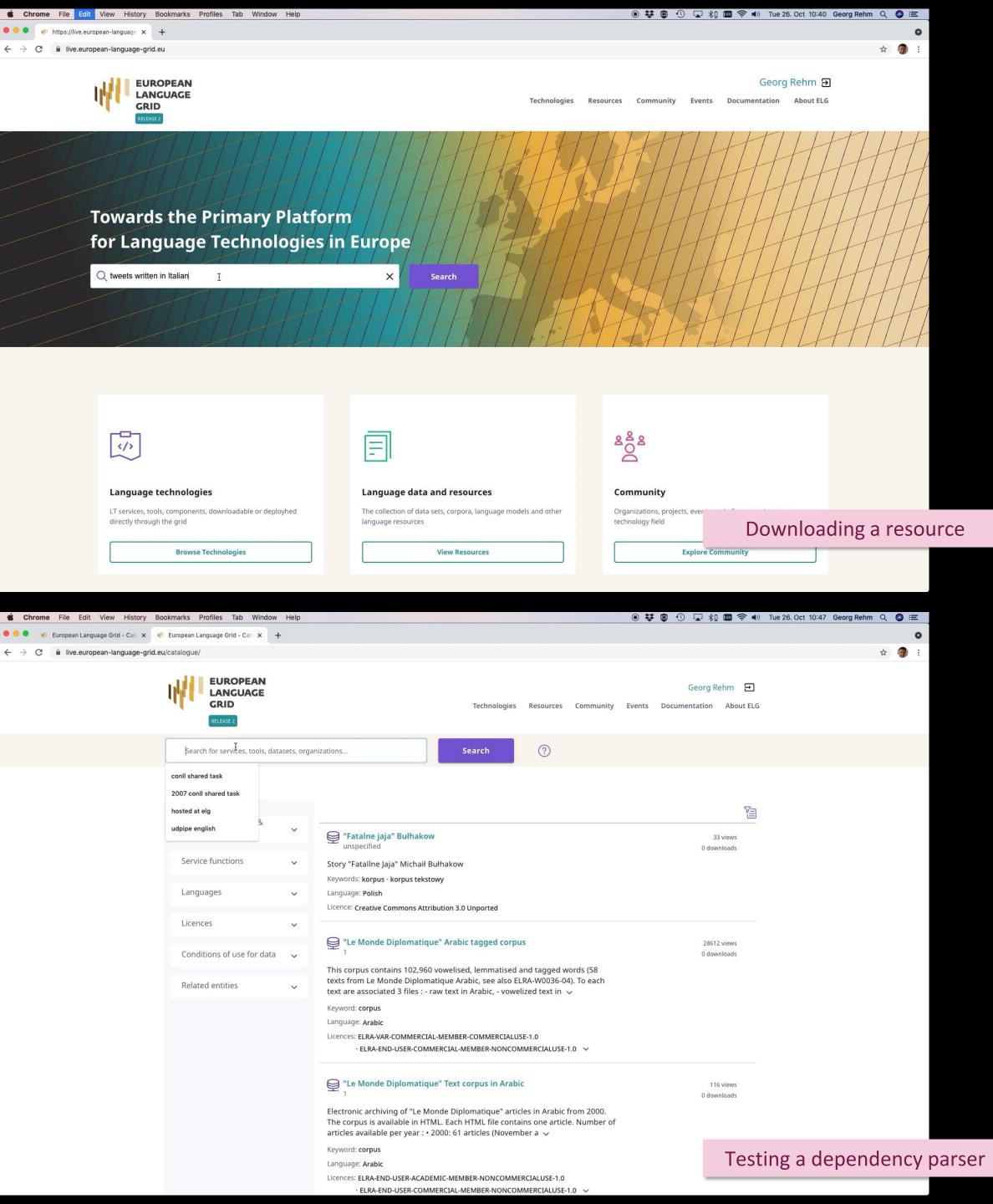


We want to make available all European Language Technologies and Language Resources in a one-stop-shop



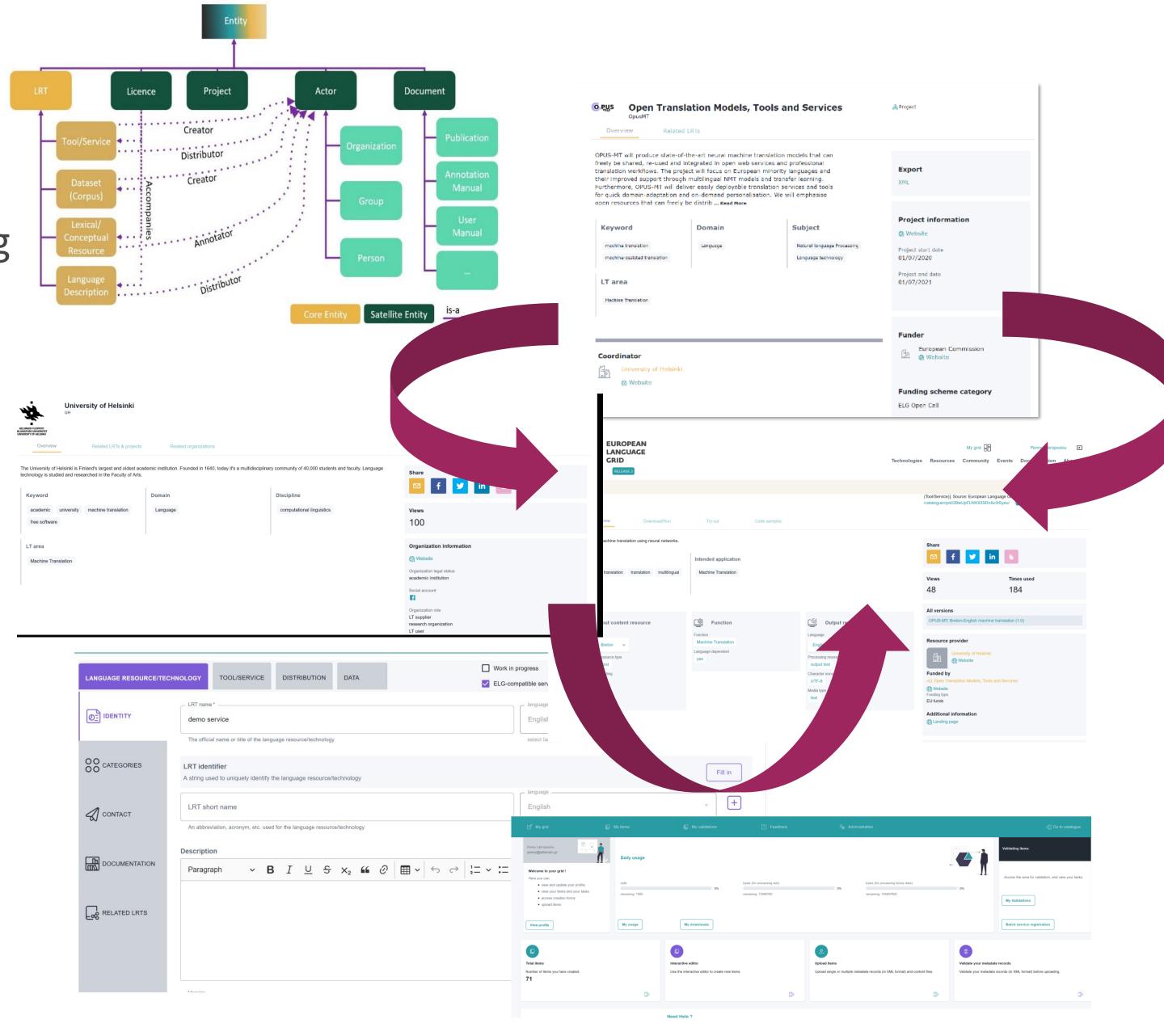




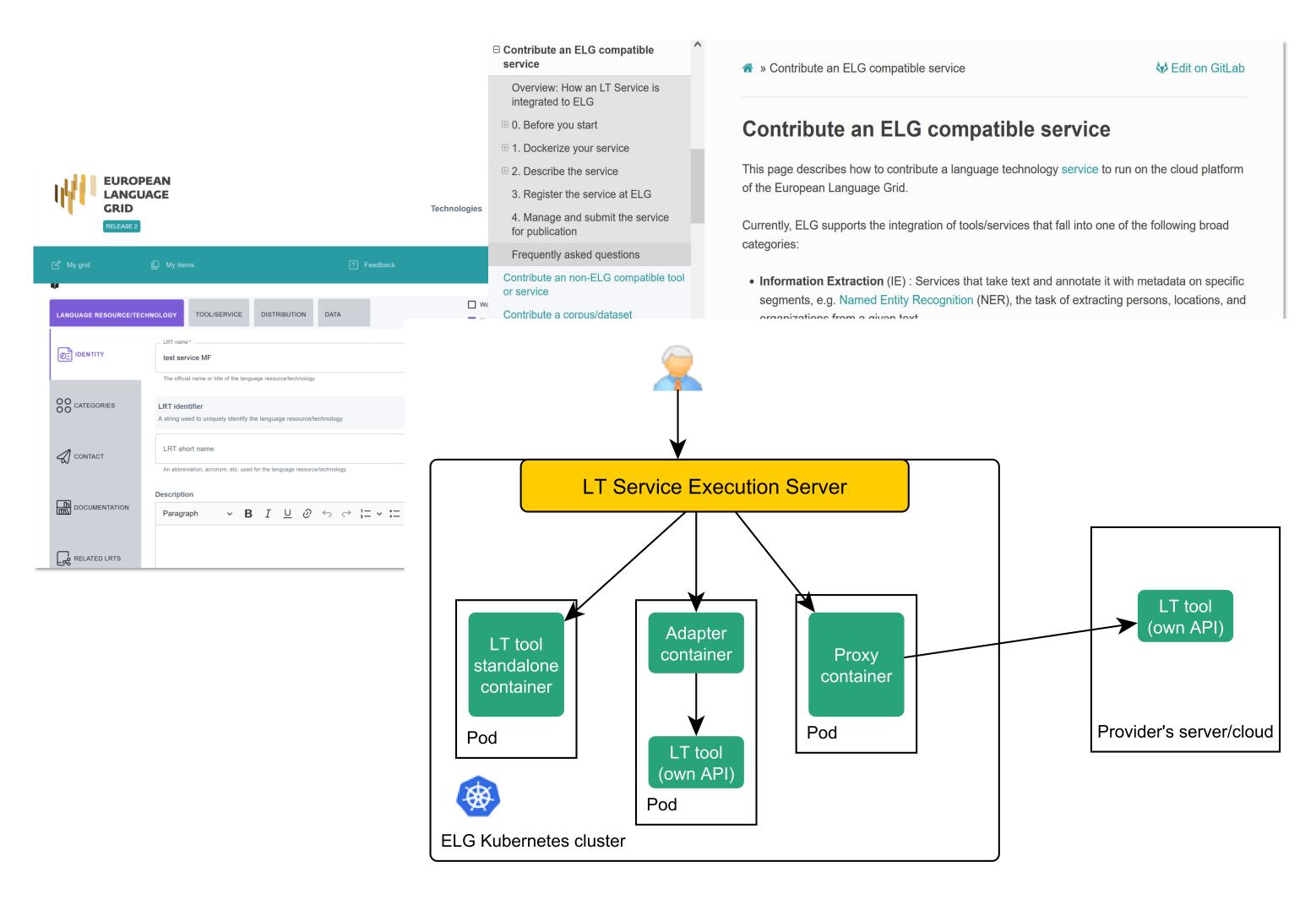


On the provider side

- LT community members
  - can contribute by providing formal descriptions compliant with a dedicated metadata schema
    - describe organisations
       (parents or divisions) &
       projects
    - describe & provide
       access (remote or
       integrated in ELG) LRTs
    - linked with each other
  - must register and get authenticated as providers



## A few details on service provision



- LT service providers need to provide a **Docker image** with their LT tool or service
- Docker images have to be uploaded to a **Docker registry** (e.g., GitLab, DockerHub)
- Three different options:
  - LT tool packaged in one Docker image exposing an ELGcompliant endpoint
  - LT tool running outside the ELG infrastructure – proxy image with one or more ELG-compliant endpoints
  - LT tool requiring an adapter adapter image exposes an ELGcompliant endpoint as proxy

18

**WELG** 

European Language Grid



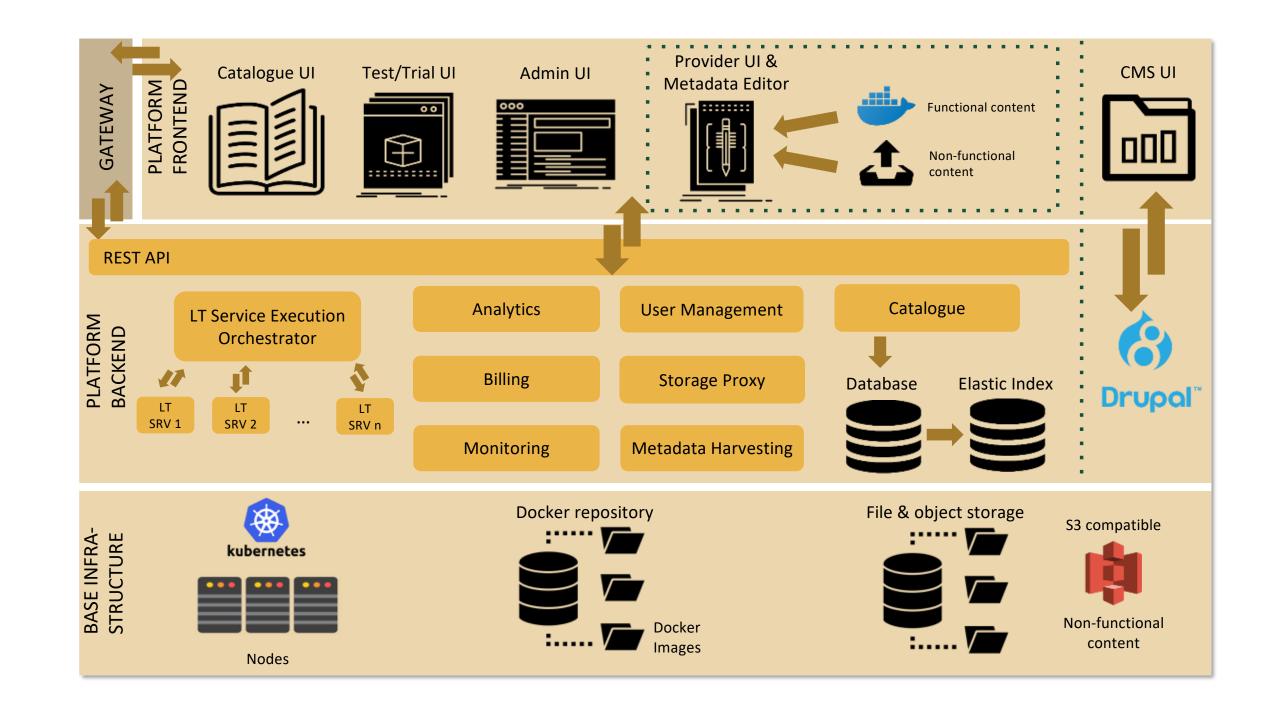
# Current State of Play (May 2022) Release 2

- 12,000+ Resources
  - 6271 corpora and data sets
  - 3102 services and tools
    - 772 services integrated and running in ELG
  - 2241 lexical/conceptual resources
  - 487 models and grammars
- 1778 LT organisations (companies, research organisations)

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

Sources	^
+ ELG / ELE	(8785)
+ ELRC-SHARE	(1299)
+ ELRA Catalogue of Language Resources	(1180)
+ Zenodo	(492)
+ Hugging Face	(385)
+ LINDAT/CLARIAH-CZ	(353)

+ Quantum Stat Datasets	(261)	
+ CLARIN-PL	(254)	
+ CLARIN.SI	(218)	
+ LREC Shared LRs (ELRA)	(144)	
+ META-SHARE/ILSP	(69)	
+ META-SHARE/DFKI	(2)	
+ TUdatalib	(1)	
+ labsai/eddi	(1)	

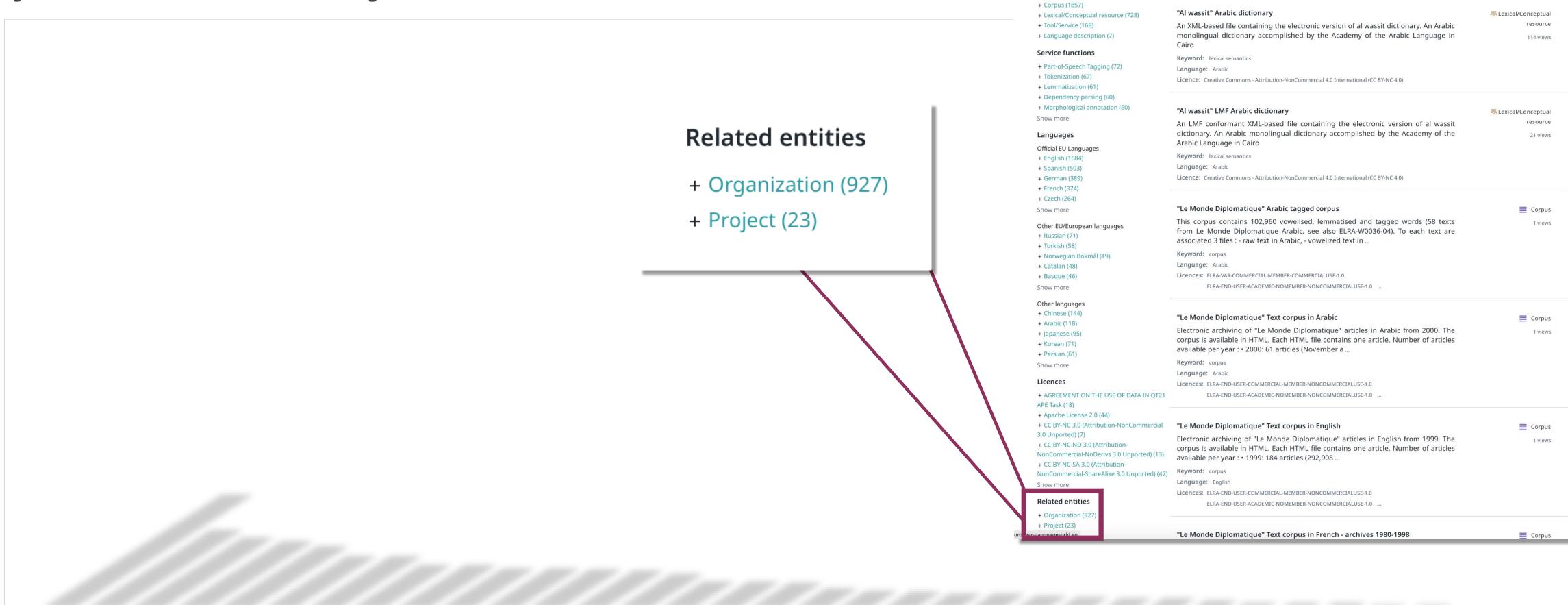


Language resources & technologies

3710 search results

Clear all filters (x)

## **European LT Industry in the ELG**



**WELG** 

European Language Grid

### Stakeholders and Users and Collaborators





Sourmet

**Global Under Resourced MEdia Translation** 









ELG Open Calls: 15 pilot projects

ICT-29b research and innovation projects

ELE – European Language Equality

Al4EU – European Al 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 hody i**EUROPEAN**tegic LT-related matters can be discusse copylinated
- Main goal: represent and support European LT community.



- 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
- 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 Resear
- 9. Finland: Krister Linden, Department of Modern Languages, University of Helsinki
- 10. France: François Yvon, CNRS-LIMSI
- Germany: Georg Rehm, Speech and Language Technology Lab, DFKI
- 12.1 Greece: Maria Gavriilidou, ILSP, R.C. "Athena"

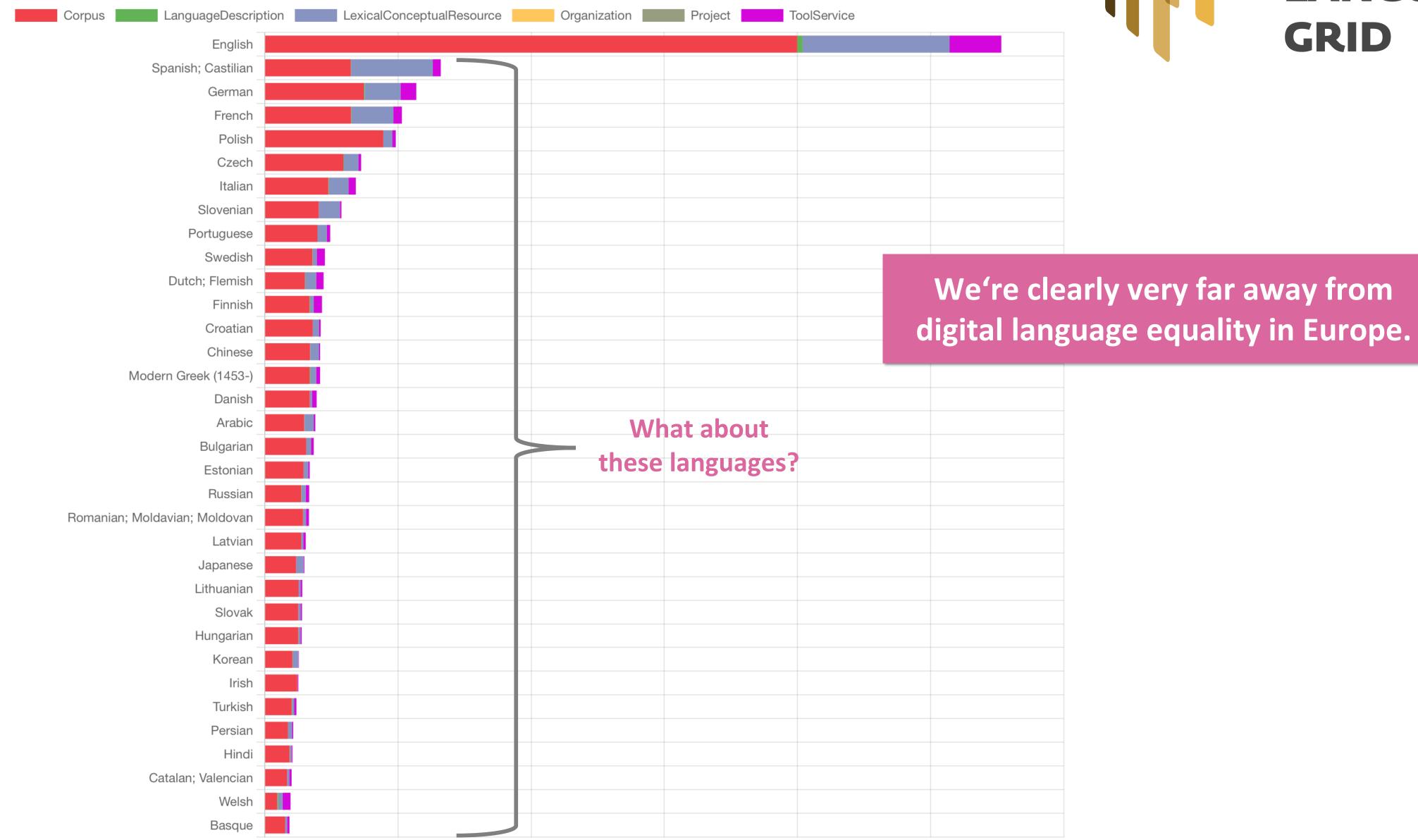
- 13. Hungary: Tamás Várárdi, Research Institute for Linguistics, Hungarian Academy of Sciences
- 14. **Iceland:** Eirikur Rögnvaldsson, School of Humanities, University of Iceland
- 15. Ireland: Andy Way, ADAPT Centre and School of Computing, Dublin City University
- 16. Italy: Bernardo Magnini, Human Language Technology, Fondazione Bruno Kessler (FBK)
- 17. Latvia: Inguna Skadina, Institute of Mathematics and Computer Science, University of Latvia
- 18. Lithuania: Albina Auksoriūtė, 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: Vincent Vandeghinste, Dutch Language Institute, Centre for Computational Linguistics, University of Leuven
- 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
- **Romania:** Dan Tufis, Research Institute for Artificial Intelligence, Romanian Academy of Sciences
- **Serbia:** Cvetana Krstev, Faculty of Mathematics, Belgrade University (UBG)
- 27. Slovakia: Radovan Garabik, Ludovit 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 Bourlard, Idiap Research Institute
- **UK:** Kalina Bontcheva, Department of Computer Science, University of Sheffield



## Languages of the Resources





₩ ELG

European Language Grid

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





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

## Approach



DLE: 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

State of the art of the LT field as a whole

State of the art for at least 31

European languages

Forward looking: projections into the future in different LT areas

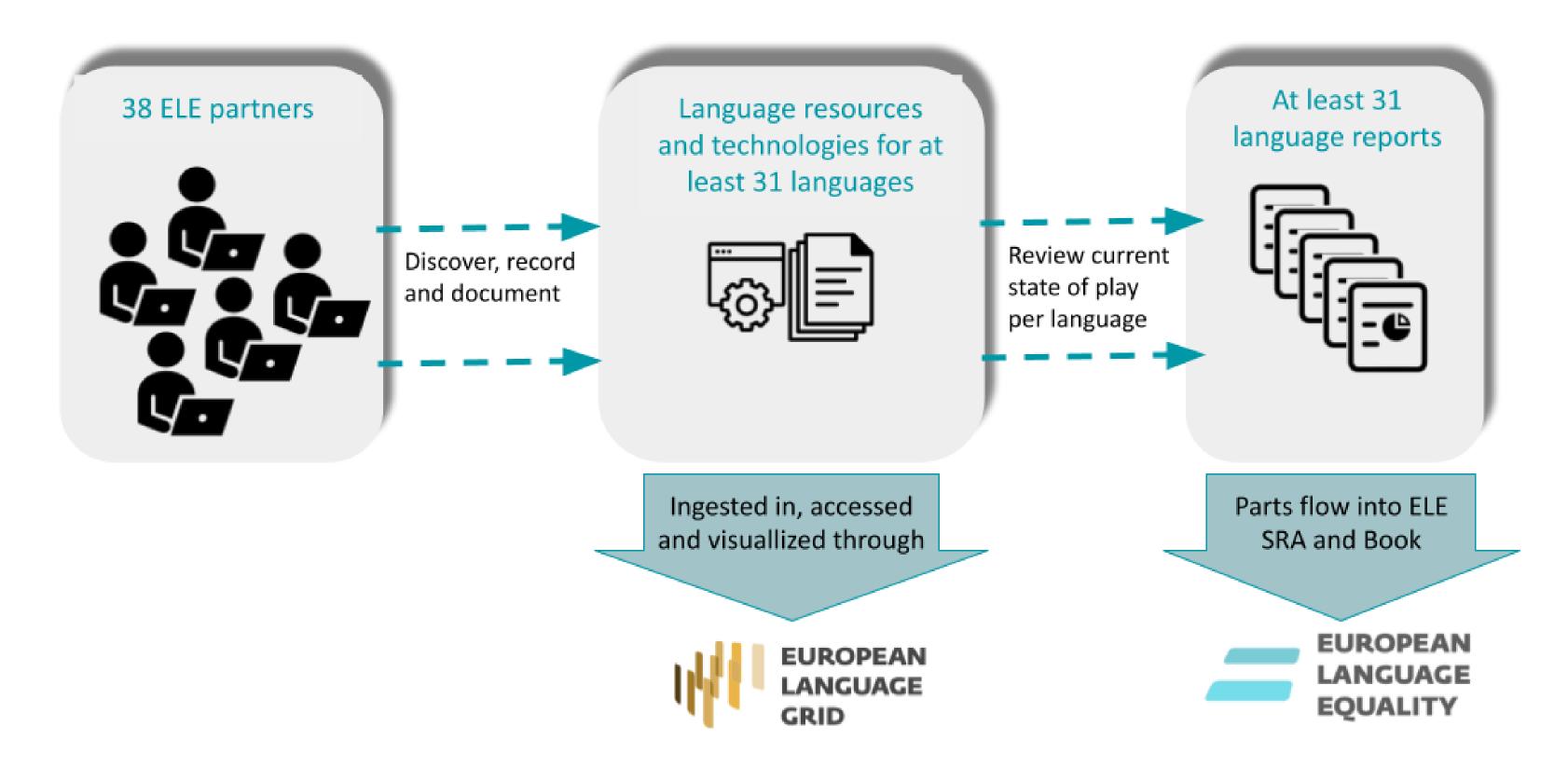
Definition of
Digital Language
Equality and
development of
DLE metric

Strategic agenda and roadmap

Consultations with stakeholders

## Campaign for the documentation of LRTs





List of all languages investigated by the ELE consortium:

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



European Language Grid

## Language reports



Tradition of META-NET White papers

Quantitative and qualitative evaluations of the level of support each language receives

through technology:

brief introduction to the LT field

general facts about the language

availability of LRTs

D1.27	Report on Polish	IPIPAN	February 2022	24 pages
D1.28	Report on Portuguese	FCULisbon	February 2022	21 pages
D1.29	Report on Romanian	ICIA	February 2022	22 pages
D1.30	Report on Slovak	JULS	February 2022	20 pages
D1.31	Report on Slovenian	JSI	February 2022	24 pages
D1.32	Report on Spanish	BSC	February 2022	25 pages
D1.33	Report on Swedish	KTH	April 2022	27 pages
D1.34	Report on Welsh	BNGR	February 2022	29 pages

- support it receives through dedicated funding programs and projects
- participation in research infrastructures

•

- Language reports available at: <a href="https://european-language-equality.eu/deliverables/">https://european-language-equality.eu/deliverables/</a>
- Parts included in the final ELE agenda and book.

### **DLE Metric**



### **Technological Factors**

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



### **Contextual Factors**

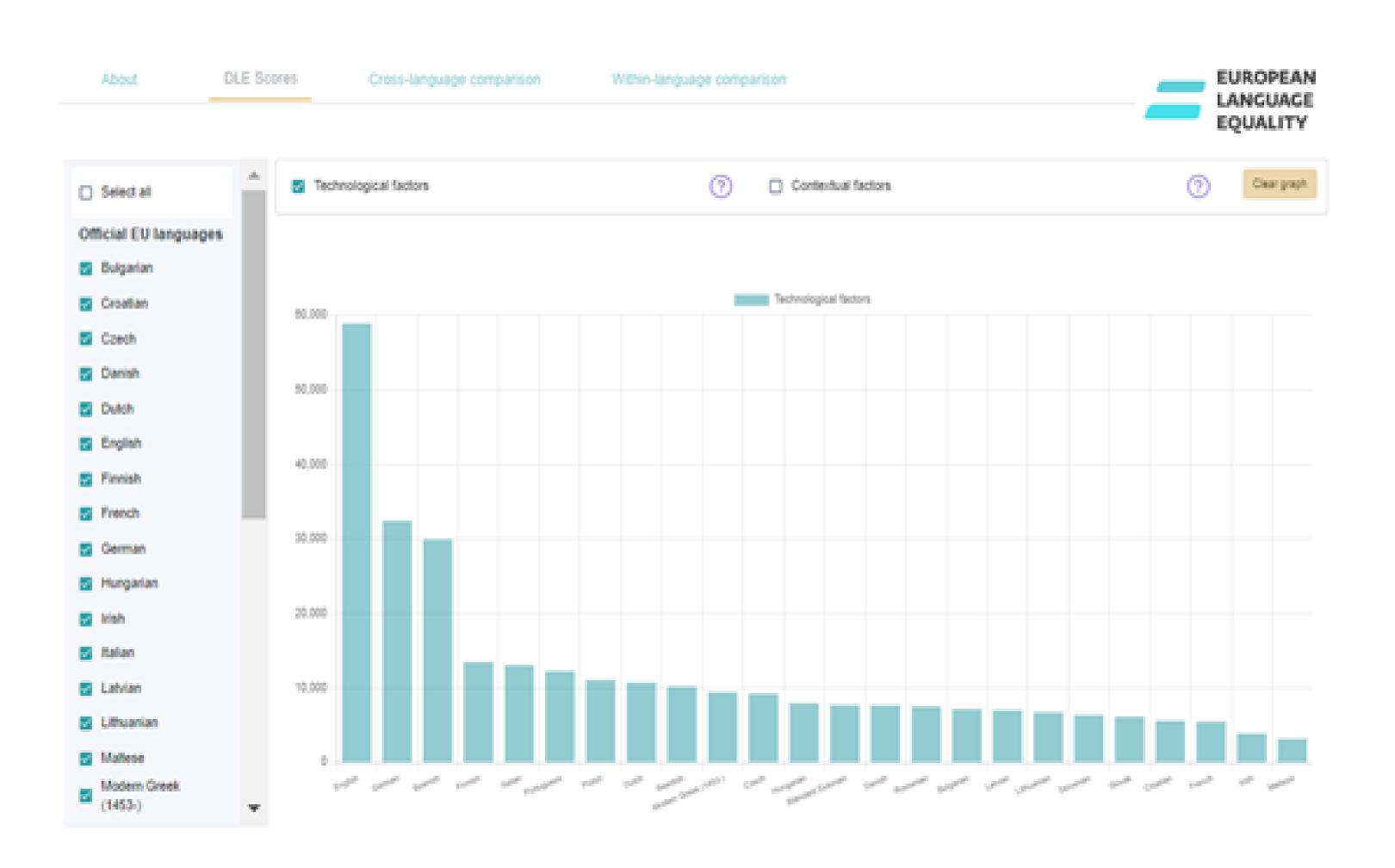
- Industry
- Online
- Research & Development & Innovation
- Technology

- Economy
- Education
- Funding
- Law
- Media
- Policy
- Public administration
- Society



## **ELG/ELE** dashboard





- A web-based interactive dashboard developed by ELE as part of the ELG Platform which exposes the contents of the ELG catalogue as interactive visuals dynamically created by user queries
- Accessed at: <a href="https://live.european-language-grid.eu/catalogue/dashboard">https://live.european-language-grid.eu/catalogue/dashboard</a>
- Why a dashboard? It provides
  - an easy-to-use mechanism for exposing and monitoring the technological and contextual factors that contribute to DLE
  - an updatable and comparable measurement of the level of LT support and provision that each European language enjoys.

European Language Grid

### **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



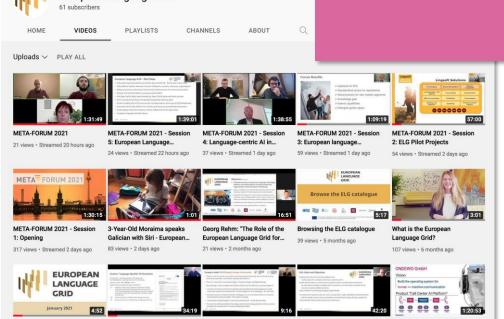
JOINING THE EUROPEAN LANGUAGE GRID:

Together Towards Digital Language Equality

8/9 June 2022

Brussels, Belgium

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





**European Language Grid** 

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





## Thank you!

Penny Labropoulou, Athena RC/ILSP



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

25-05-2022 ELG NCC Workshop Bulgaria http://www.european-language-grid.eu http://www.european-language-equality.eu

