

EUROPEAN LANGUAGE GRID

European Language Grid: Overview of the project and demo of the platform

Andrejs Vasiljevs (Tilde) – ELG Consortium

21-09-2020 ELG – ELG Workshop – co-located with Baltic HLT (virtual meeting)
<http://www.european-language-grid.eu>

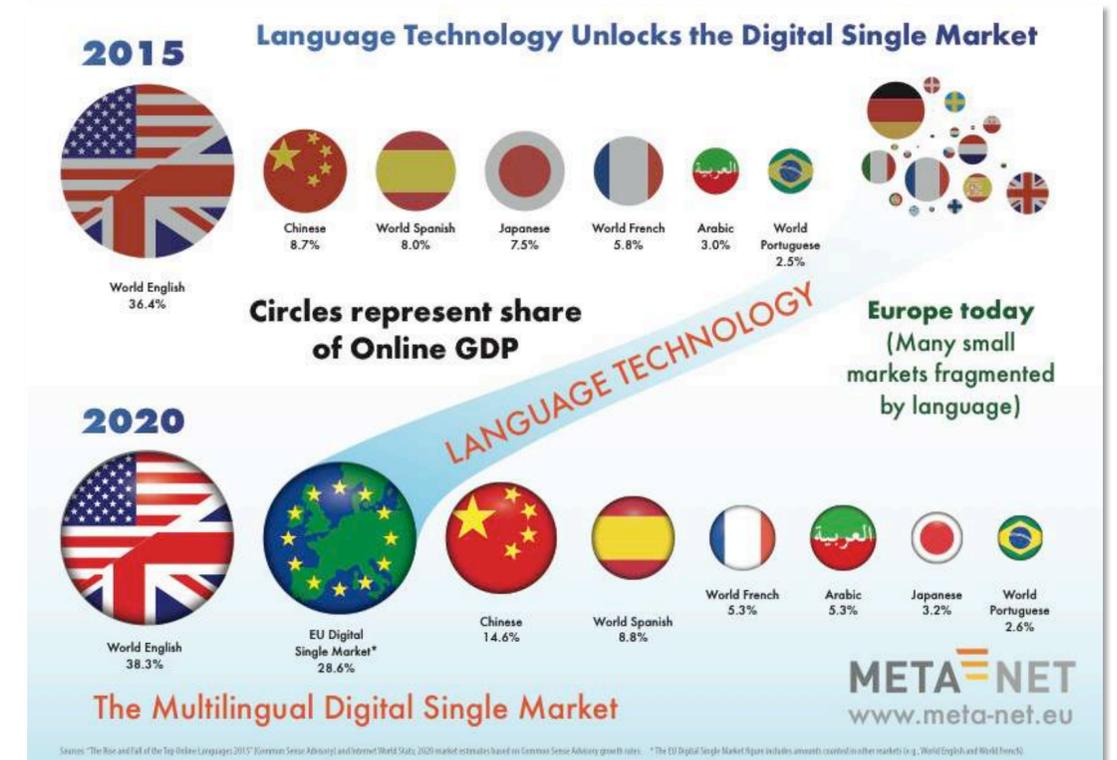


European Language Grid

- ELG project: Introduction
- ELG and the European LT Industry
- Demo of the platform
- Q&A

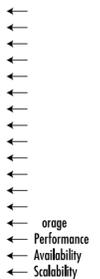
Point of Departure: Multilingualism in Europe

- Multilingualism is at the heart of the European idea
- 24 official EU languages – they all have the same status
- Dozens of 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
 - There's no LT platform for Europe! (yet)

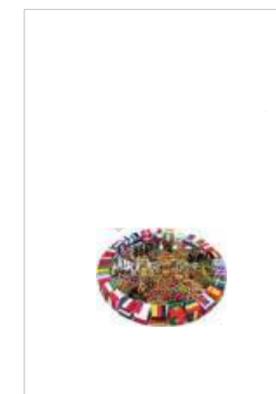


Motivation and Context

- The European LT landscape and market are very fragmented (cf. the CEF market study report).
- The European LT community has been demanding a **European Language Technology platform** for several years.
- **META^{NET}** vision: **European Service Platform for Language Technologies.**
- First mention in the **Strategic Research Agenda for Multilingual Europe 2020** (published in early 2013).



The concept was later refined in the three LT SRIAs (2015, 2016, 2017).



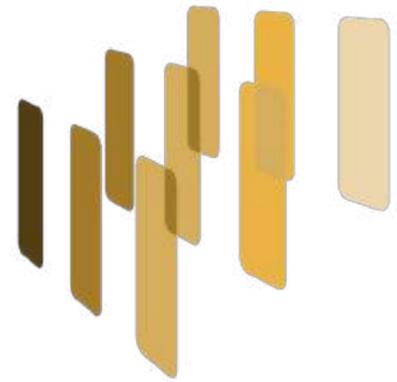
META FORUM 2015



META FORUM 2016



META FORUM 2017



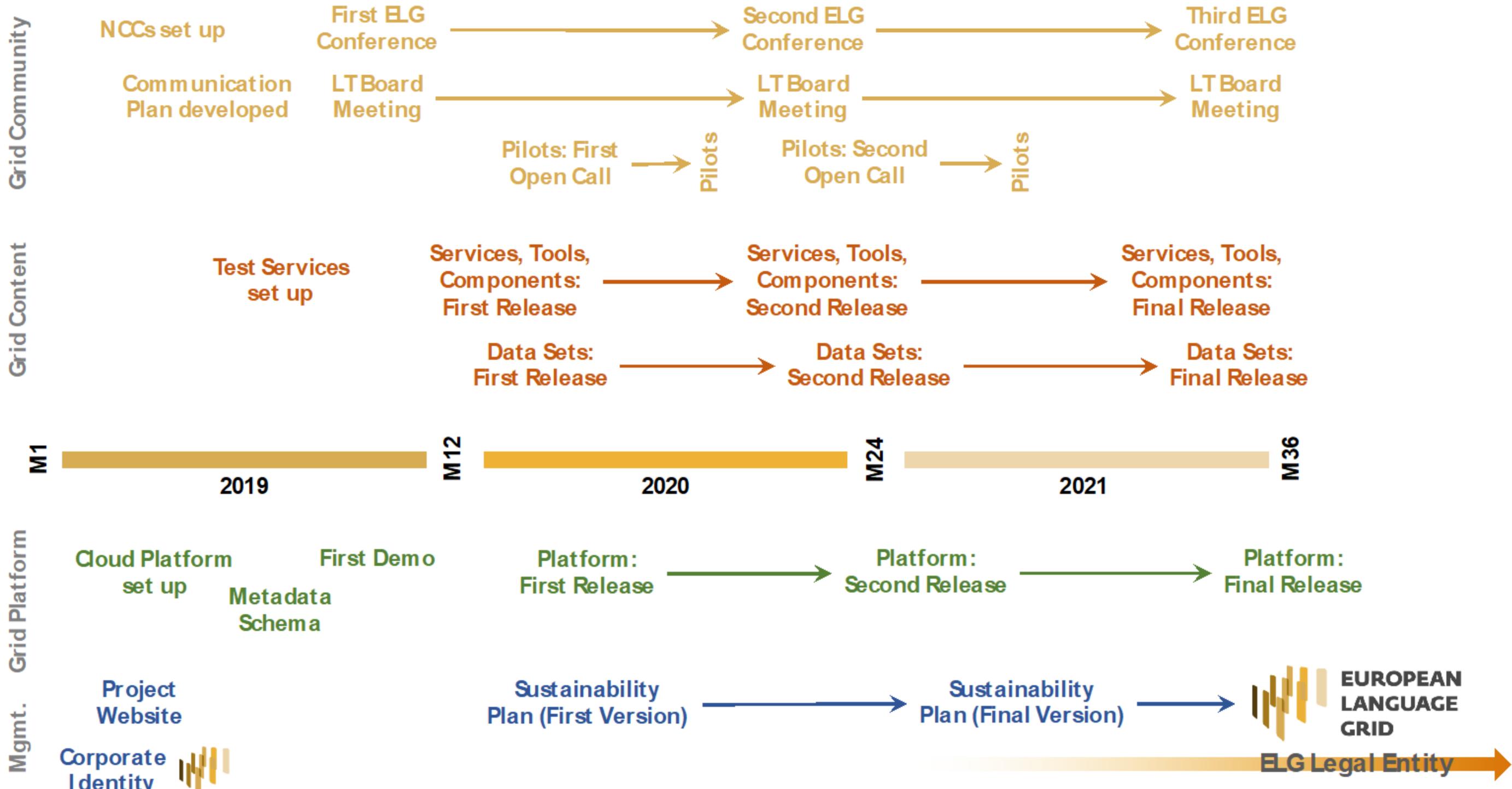
EUROPEAN LANGUAGE GRID

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, both functional and non-functional.
3. Enable the European LT community to upload services and data sets into the ELG, 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



ELG Platform

- Infrastructure, backend, frontend.
- Built with robust, scalable, reliable, widely used technologies that are constantly developed further.
- Ability to scale with the growing demand and supply of resources.
- Interactive, modern web interface.
- Provides the base technology for a catalogue or repository of
 - Functional services, data sets, tools, technologies, models etc.
 - LT companies, research organisations, research projects, service and application types, languages etc. (“yellow pages” of European LT).

ELG Content

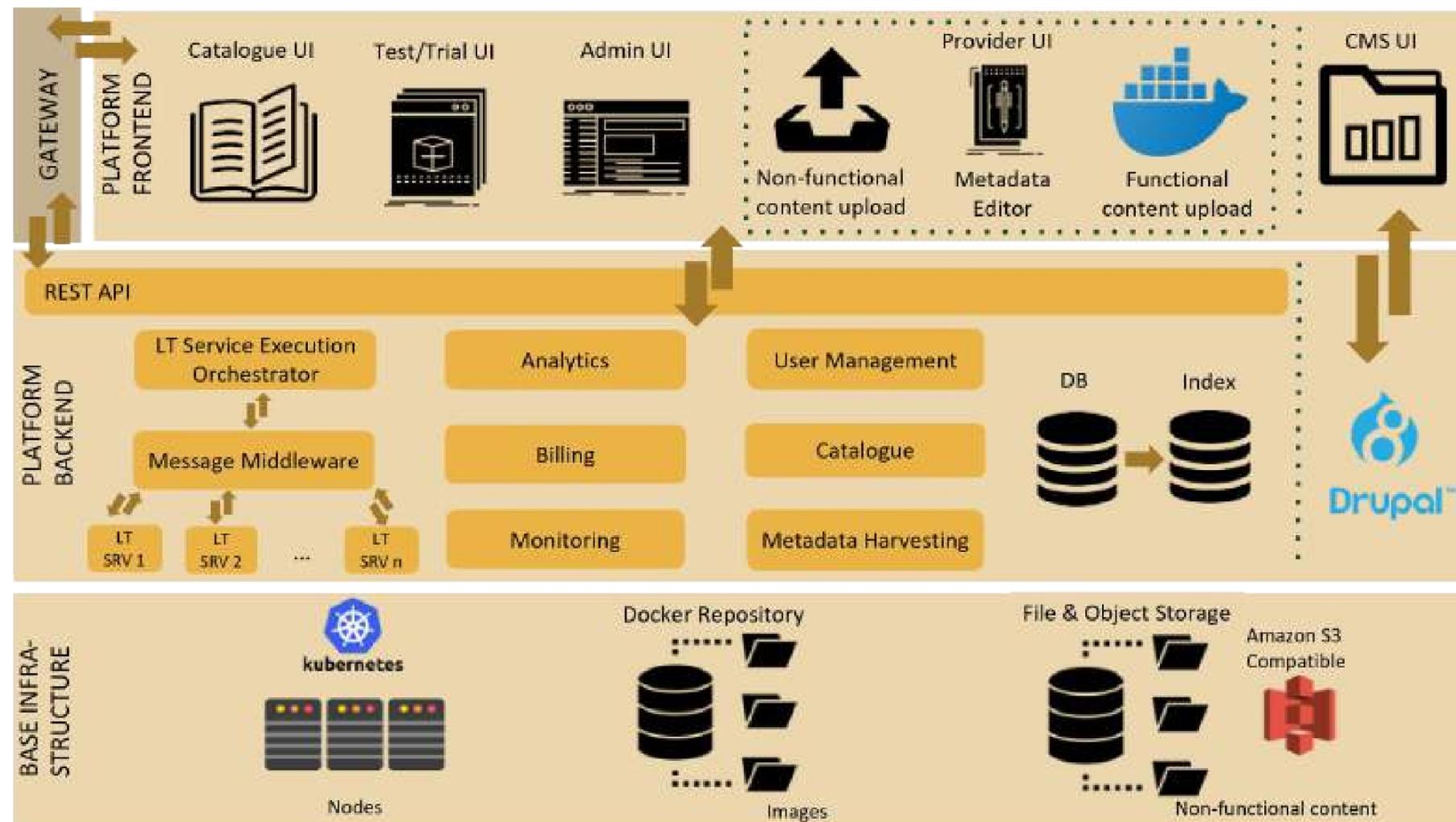
- Content: services, LRs, data sets, tools, catalogue/directory content.
 - Functional Content: running LT services, can be integrated into other systems.
 - Non-functional Content: LRs and data sets but also records of companies, research groups etc.
 - Stakeholders are able to upload and make available their own services or LRs.
- Functional services realised by
 - Containerising an LT service (dockerisation).
 - Easy and efficient for LT providers to create and upload containers.

ELG Community

- Stakeholders: LT provider and buyer companies, research centres, universities, administrations etc.
- Plus: 15-20 pilot projects, the six ICT-29b) RIA projects, AI4EU, META-NET, LT Innovate and other initiatives
- 32 NCCs as a strong international network that act as national bridges to identify content, interest companies in making available their services.
- European LT Council (LTC) as an international, pan-European body, in which LT-related matters can be discussed and coordinated.
- Conferences, training events, presentations, social media, blog posts etc.

European Language Grid – Release 1 (September 2020)

- User registration, authentication, authorisation
- User categories, respective rights and policies
- LT metadata upload
- Metadata conversion and harvesting from ELRC-SHARE, META-SHARE, ELRA
- LT service registration, integration
- LT service try out and execution
- LT data browsing, searching, downloading
- Online documentation (APIs, basic platform services, metadata schema etc.)
- 840 entries in the ELG catalogue:
 - 161 functional services and tools
 - 556 data sets and language resources



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

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

Stakeholders and Collaborators

ELG Open Calls: 15-20 pilot projects

ICT-29b research and innovation projects

AI4EU – European AI on demand platform

Other related projects: Elexis, Lynx, MeMAD, Fandango, CEF INEA projects (MAPA, NTEU) etc.

Core networks: META-NET, LT Innovate

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



Selected Collaborations

ICT-29b) projects

- Various conference calls, participation of all projects in META-FORUM 2019
- Commitments of several projects to contribute services and/or data sets

AI4EU

- Various conference calls and face-to-face meetings
- Topics: platform architecture and technical building blocks; publication processes and QA; open calls; metadata and semantics; platform interoperability; data set and service exchange; making resources visible and searchable to one another; ELG contributed to AI4EU SRIA; AI4EU contributed to META-FORUM 2019
- Two joint papers (LREC 2020, IWLTP 2020)



Global Under Resourced Media Translation



COMPRISE



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 ROEVA

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



BØLETTE SANDFORD PEDERSEN

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

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 ways 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
- First concept prepared (Business Model Canvas).
- Q3/2021: Establish legal entity (probably with a soft start)



Open Calls for Pilot Projects

Two open calls for pilot projects

- **Open Call #1:** 03/04 2020
- **Open Call #2:** 09/10 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€**

Project duration: **9-12 months**

Eligibility: **SMEs, research orgs.**

Open Call #1 – Statistics

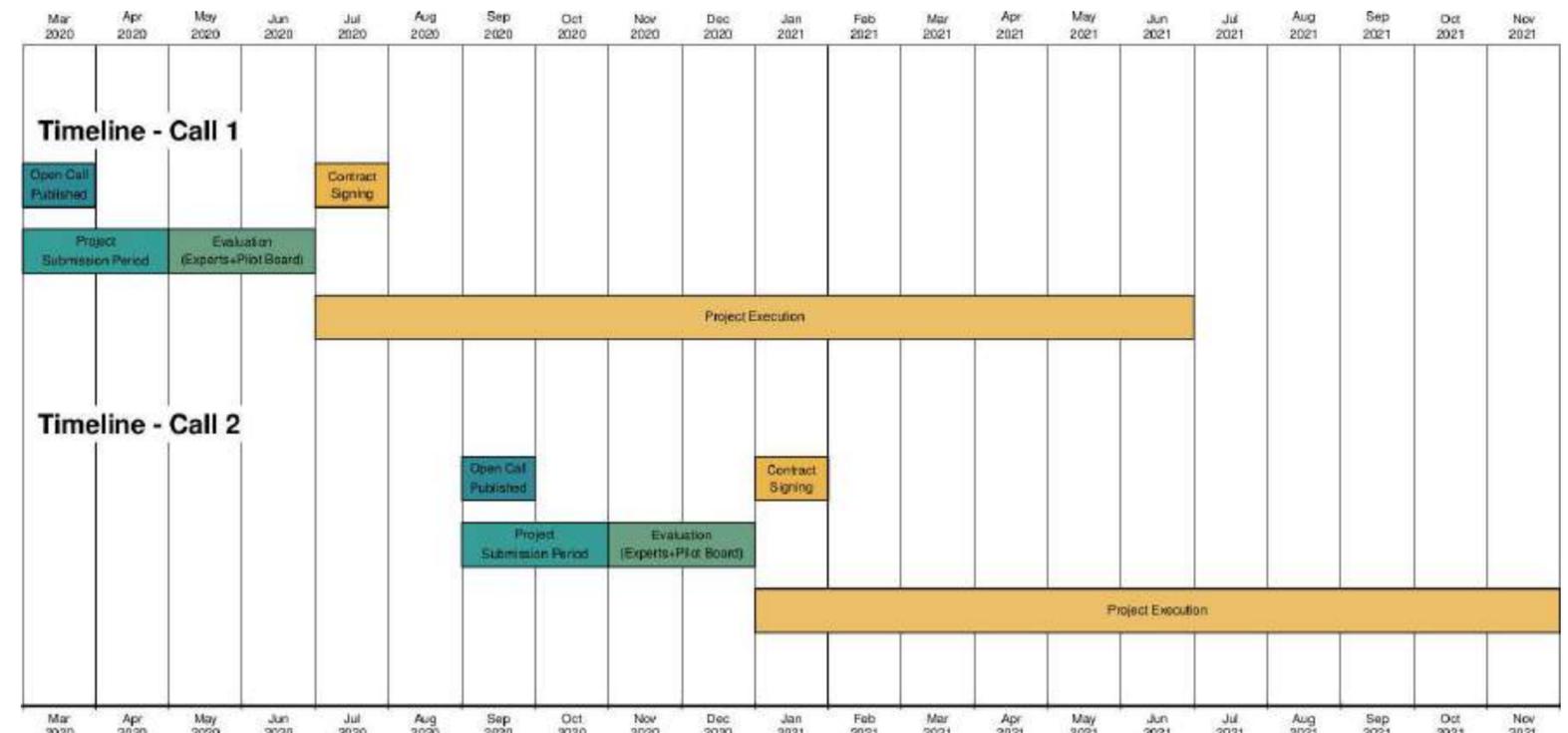
121 project proposals submitted

110 project proposals evaluated

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

Total amount requested: 16.9M€

10 projects selected on 29 June 2020



European Language Grid: An Overview

Georg Rehm¹, Maria Berger¹, Ela Elsholz¹, Stefanie Hegeler¹, 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 Arranz⁶, Andrejs Vasiljevs⁶, Orians Anwar⁶, Andis Lagzdins⁶, Jūlija Meņņika⁶, Gerhard Backfried⁷, Eric Dikič⁷, Miroslav Janoušek⁷, Katja Prinz⁷, Christoph Prinz⁷, Severin Stampfer⁷, Dorothea Thomas-Aniola⁷, José Manuel Gómez Pérez⁸, Andrés García Silva⁸, Cristian Berrio⁹, Ulrich Germann⁹, Steve Renals⁹, Ondřej Klejčt⁹

¹DFKI GmbH, Germany • ²ILSP/Athens 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

With 24 official EU and many additional languages, multilingualism is enabled through Language Technologies (LTs). European LT business is diverse and world-class, with technologies that outperform the global players. It spans states, languages, verticals and sectors, significantly holding back its fragmentation by establishing the ELG as the primary platform for LT in an easy-to-integrate way, access to hundreds of commercial and non-commercial tools and services as well as data sets and resources. Once fully operational, the LT community to deposit and upload their technologies and data sets into other resources. The ELG will boost the Multilingual Digital Single Market jobs and opportunities. Furthermore, the ELG project organises two open Competence Centres (NCCs) and the European LT Council (ETC) for out-

Keywords: LR Infrastructures and Architectures, LR National/International

1. Introduction

With 24 official EU languages and many additional ones, multilingualism, cross-lingual and cross-cultural communication in Europe as well as an inclusive EU Digital Single Market can only be enabled and firmly established through Language Technologies (LTs) (Rehm, 2016). The European LT industry is dominated by hundreds of SMEs and a few large players. Many are world-class, with technologies that outperform the global players. However, European LT business is also fragmented by nation states, languages, domains and sectors (Vasiljevs et al., 2019), significantly holding back its impact. In addition, many European languages are severely under-resourced and, thus, in danger of digital language extinction (Kohm and Uszkoreit, 2012; Korman, 2013; Rehm et al., 2014; Rehm et al., 2016a), which is why there is an enormous need for a European LT platform as a unifying umbrella (Rehm and Uszkoreit, 2013; Rehm et al., 2016b; STOA, 2017; Rehm, 2017; Rehm and Hegeler, 2018; European Parliament, 2018). The project European Language Grid (ELG; 2019-2021) addresses this fragmentation by establishing the ELG as the primary platform and marketplace for the European LT community, both industry and research.¹ 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 sets and

¹<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 Rehm¹, Katrin Marheinecke¹, Stefanie Hegeler¹, Stelios Piperidis², Kalina Bontcheva³, Jan Hajič⁴, Khalid Choukri⁵, Andrejs Vasiljevs⁶, Gerhard Backfried⁷, Christoph Prinz⁷, José Manuel Gómez Pérez⁸, Luc Meertens⁹, Paul Lukowicz⁹, Josef van Genabith⁹, Andrea Lösch⁹, Philipp Slusallek⁹, Morten Irgens¹⁰, Patrick Gattelier¹¹, Joachim Köhler¹², Laure Le Bars¹³, Dimitra Anastasiou¹⁴, Albina Auksoyriūtė¹⁵, Núria Bel¹⁶, António Branco¹⁷, Gerhard Budin¹⁸, Walter Daelemans¹⁹, Koenraad De Smedt²⁰, Radovan Garabik²¹, Maria Gavrilidou²², Dagmar Gromann²³, Svetla Koeva²⁴, Simon Krek²⁵, Cvetana Krstev²⁶, Kristin Lindén²⁷, Bernardo Magnini²⁸, Jan Odijk²⁹, Macej Ogrodniczuk³⁰, Eirikur Rögnvaldsson³¹, Mike Rosner³², Bolette Sandford Pedersen³³, Inguna Skadiņa³⁴, Marko Tadić³⁵, Dan Tufiş³⁶, Tamás Váradi³⁷, Kadri Vider³⁸, Andy Way³⁹, François Yvon⁴⁰

¹DFKI GmbH, Germany • ²ILSP/Athens 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 •

¹²Germany • ¹³SAP, Germany • ¹⁴Luxembourg Institute of Science and Technology, Luxembourg • ¹⁵Lithuanian Language, Lithuania • ¹⁶University Pompeu Fabra, Spain • ¹⁷University of Vienna, Austria • ¹⁸University of Antwerp, Belgium • ¹⁹University of Sciences, Slovakia • ²⁰Bulgarian Academy of Sciences, Bulgaria • ²¹City of Belgrade, Serbia • ²²University of Helsinki, Finland • ²³FBK, Italy • ²⁴ICS, Polish Academy of Sciences, Poland • ²⁵University of Iceland, Iceland • ²⁶University of Copenhagen, Denmark • ²⁷University of Latvia, Latvia • ²⁸Romanian Academy of Sciences, Romania • ²⁹MTA Research Institute for Linguistics, Estonia • ³⁰Dublin City University and ADAPT Centre, Ireland • ³¹CNRS, France • ³²DFKI GmbH, Germany • ³³DFKI GmbH, Germany • ³⁴DFKI GmbH, Germany • ³⁵DFKI GmbH, Germany • ³⁶DFKI GmbH, Germany • ³⁷DFKI GmbH, Germany • ³⁸DFKI GmbH, Germany • ³⁹DFKI GmbH, Germany • ⁴⁰DFKI GmbH, Germany

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

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.

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, Mickael Rigault³, Victoria Arranz³, Khalid Choukri³, Gerhard Backfried⁴, José Manuel Gómez Pérez⁵, Andrés García Silva⁵

¹ILSP/Athens 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, 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 Hegeler, 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).

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 Wolf³, Ricardo Usbeck³, Joachim Köhler³, Miltos Deligiannis², Katerina Gkirtzou², Johannes Fischer⁴, Christian Chiaros⁵, Niels Feldhus⁵, Julián Moreno-Schneider⁶, Florian Kintzel⁶, Elena Mantiel⁶, Víctor Rodríguez Donco⁶, John P. McCrae⁷, David Laqua⁷, Irina Patricia Theile⁸, Christian Dittmar⁸, Kalina Bontcheva⁹, Ian Roberts⁹, Andrejs Vasiljevs⁹, Andis Lagzdins⁹

¹DFKI GmbH, Germany • ²ILSP/Athens 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

In this paper, we concentrate on two core aspects: (1) cross-platform search and platform service workflows. We devise five different levels (of increasing implementation in a wider federation of AI/IT platforms). We illustrate the QURATOR, LYNX and SPEAKER.

International Projects, Tools, Systems, Applications, Web Services

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

From their respective unique offerings. Only by discussing and agreeing upon standards as well as technical and operational concepts for AI/IT platform interoperability, can we benefit from the highly fragmented landscape and its specialised platforms. This paper takes a few initial steps, which we demonstrate primarily using the two platforms AI4EU and ELG (European Language Grid) but also including QURATOR, LYNX and SPEAKER. These platforms are introduced in Section 2, where we also compare their architectures. Section 3 introduces requirements and prerequisites for platform interoperability, including shared semantics as well as legal and operational interoperability, followed by a description of five levels of platform interoperability that exhibit an increasing level of conceptual complexity. Section 4 summarises the paper and presents next steps. We contribute to the challenge of platform interoperability by identifying this topic as a crucial common development target and by suggesting a roadmap for the implementation of different levels of interoperability.

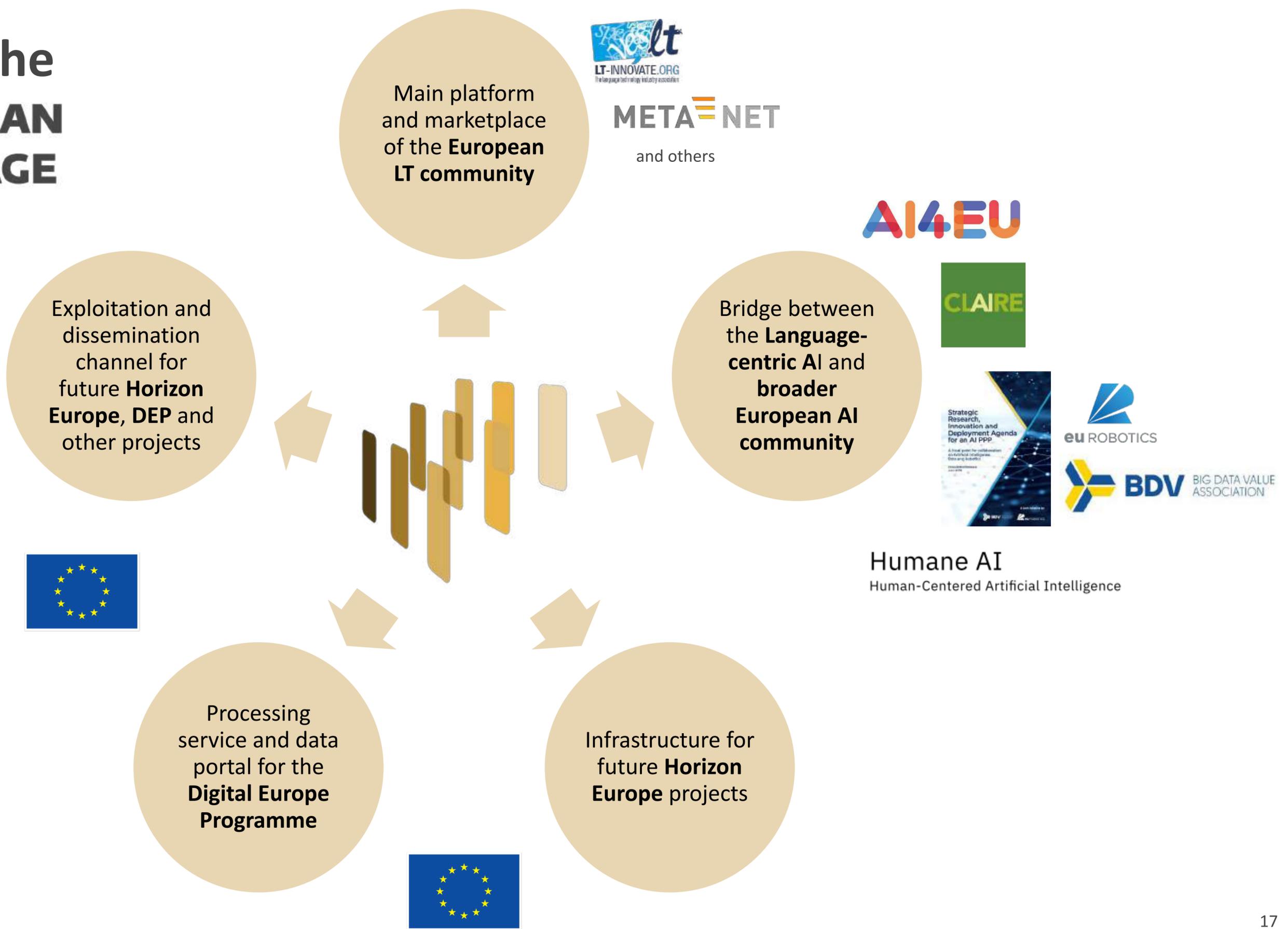
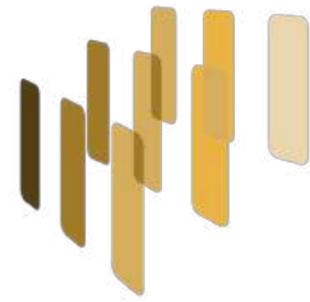
2. The Platforms

In the following, we describe the platforms AI4EU (Section 2.1), ELG (Section 2.2), QURATOR (Section 2.3), LYNX (Section 2.4) and SPEAKER (Section 2.5).

2.1. AI4EU

In January 2019, the AI4EU consortium with more than 80 partners started its work to build the first European AI on-demand platform. The main goals are: the creation and support of a large European ecosystem to facilitate collaboration between all European AI actors (scientists, entrepreneurs, SMEs, industries, funding agencies, citizens etc.), the design of a European AI on-demand platform to share AI resources produced in European projects, including high-level services, expertise in research and innovation, components and data sets, high-powered computing resources and access to seed funding for innovative

Future roles of the EUROPEAN LANGUAGE GRID





Session 6 – ELG and the European LT Industry

08/09-10-2019, Brussels – META-FORUM 2019

ELG and the European LT Industry

Language Technology: Industry vs. Research

- Language Technology has always been extremely research-intensive.
- If a company wants to stay competitive, it must stay in touch with state-of-the-art research.
- Not only Google, Facebook, Amazon but also Bosch, EML, Tilde and many, many others.
 - Industry & research: successful EU LT companies that *don't* do research are *very* rare.
 - Research & industry: many European LT research organisations establish spin-offs.
- Time from lab to market is getting shorter and shorter – ELG wants to reduce it even more.
- Our common goal is for the European LT industry to thrive, to become as successful as possible and to participate as a key player in the enormous \$29.5B LT market by 2025.
- How? Position ELG as the main hub for the most recent industry-relevant research technologies and data sets so that industry can pick them up in order to integrate them.



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

➤ **CEF study estimates the European LT Market to reach 1B€ in 2020.**

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*



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)

European LT Market – CEF Study

Approach and Observations

- Background: SMART 2016/0103 contract supporting the EC in CEF AT: contribute to CEF AT as “multilingualism enabler” for CEF DSIs.

- of tech: Translation, Speech, Search, NLP, Analytics
- EU market approx. 10B – disrupted by dominant global players
- DSIs: 70% of EU LT vendors up to 50 employees
- Revenue per company is growing
- Market is highly fragmented: hundreds of DSIs, many address specific niches, sectors and languages

Recommendations

- Market is strong in research and innovation, but not successful to commercialise and capture the value
- Market needs European alternatives to fill the gaps and deficiencies and to avoid reliance on monopolies
- Multilingual DSI 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

Final study report on CEF Automated Translation value proposition in the context of the European LT market/ecosystem

FINAL REPORT
A study prepared for the European Commission
DG Communications Networks, Content & Technology
By:

CROS SLAND
YLBI
eDA
IDC

Digital Single Market

Country	2015	2016	2017
Germany	10000	10000	10000
France	10000	10000	10000
Netherlands	10000	10000	10000
Rest of EU 28	20000	27000	20000
European LT market	50000	67000	50000
Non-European LT market	10000	10000	10000



European LT Market – CEF Study

Approach and Observations

- Background: SMART 2016/01000 contract supporting the EC in CEF AT, contribute to CEF AT as "multilingual enabler" for CEF DNs.
- European LT vendors grouped per type of tech: Translation, Speech, Search, etc.

EU market approx. 1B€

SMEs: 70% of EU LT vendors up to 50 employees

Market is highly fragmented: hundreds of SMEs, many address specific niches, sectors and languages

Recommendations

- Europe is strong in research and innovation, 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 non-European
- Multilingual DNs should be developed on its own contract

	2018	2019	2020
Germany	197M€	217M€	240M€
UK	189M€	209M€	232M€
France	88M€	96M€	105M€
Netherlands	55M€	60M€	66M€
Rest of EU 28	249M€	277M€	305M€
European LT Market Forecast 2018-2020	778M€	859M€	948M€



European LT Market – CEF Study

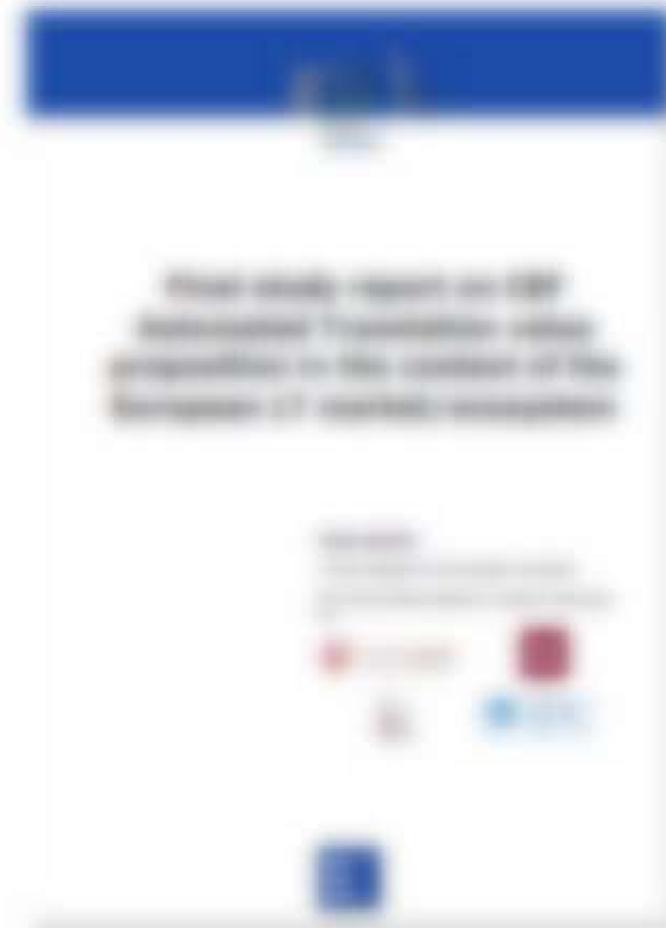
Approach and Observations

- Background: SMART 2016/01000 contract supporting the EC in CEF AT; contribute to CEF AT as "multilingual enabler" for CEF DNs.
- European LT vendors grouped per type of tech: Translation, Speech, Search, NLP, Analytics
- EU market approx. 10B – disrupted by dominant global players
- Startups: 70% of EU LT vendors up to 50 employees
- Revenue per company is growing
- Market is highly fragmented: hundreds of startups, many address specific niches, sectors and languages

Recommendations

Europe is strong in research and innovation, but not successful to scale innovations and capture the market

- to fill the gaps and deficiencies and to avoid reliance on monopolies
- Multilingual DNN 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



	2018	2019	2020
Germany	20700	21700	20000
UK	20000	20000	21200
France	9000	9000	20700
Netherlands	1000	2000	9000
Rest of EU 28	20000	21700	20700
European LT market	70000	80000	90000
European 2020 goal			

European LT Market – CEF Study

Approach and Observations

- Background: SMART 2016/01000 contract supporting the EC in CEF AT, contribute to CEF AT as "multilingualism enabler" for CEF DNs.
- European LT vendors grouped per type of tech: Translation, Speech, Search, NLP, Analytics
- EU market approx. 10% of global market, disrupted by dominant global players
- Service: 70% of EU LT vendors up to 50 employees
- Revenue per company is growing
- Market is highly fragmented: hundreds of SMEs, many address specific niches, sectors and languages

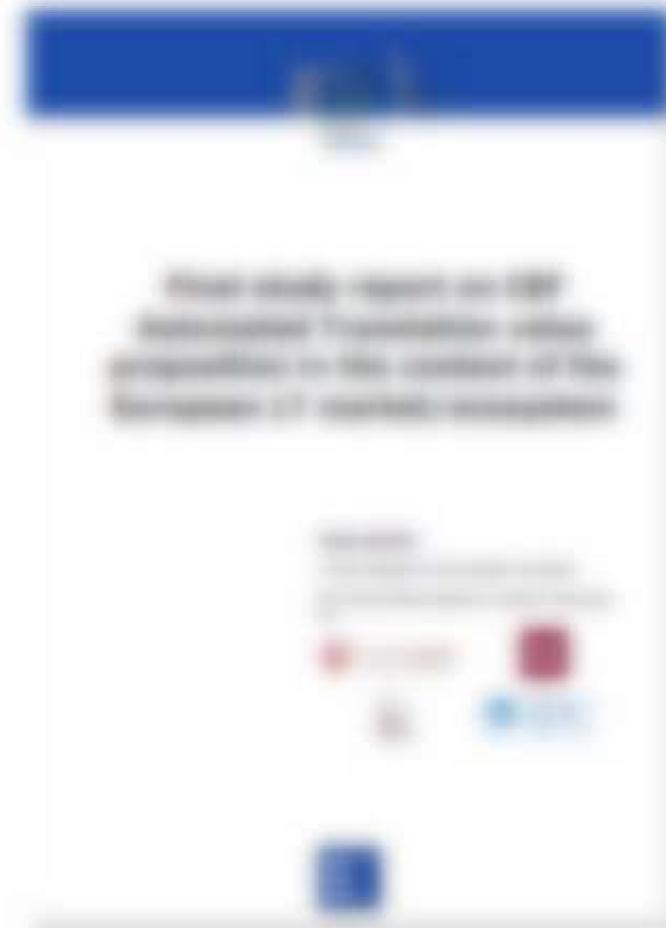
Recommendations

- Europe is strong in research and innovation, 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

to avoid dependence on monopolies

to avoid dependence on monopolies



	2018	2019	2020
Germany	20700	21700	22000
France	10000	10000	11200
Spain	6000	6000	10700
Italy	1000	1000	1000
Rest of EU 28	2000	2700	2000
Germany LT market	17000	18700	19000
France LT market	10000	10000	10000

European LT Market – CEF Study

Approach and Observations

- Background: SMART 2016/01000 contract supporting the EC in CEF AT, contribute to CEF AT as "multilingual enabler" for CEF DSMs.
- European LT vendors grouped per type of tech: Translation, Speech, Search, NLP, Analytics
- EU market approx. 10M – disrupted by dominant global players
- DSMs: 70% of EU LT vendors up to 50 employees
- Revenue per company is growing
- Market is highly fragmented: hundreds of DSMs, many address specific niches, sectors and languages

Recommendations

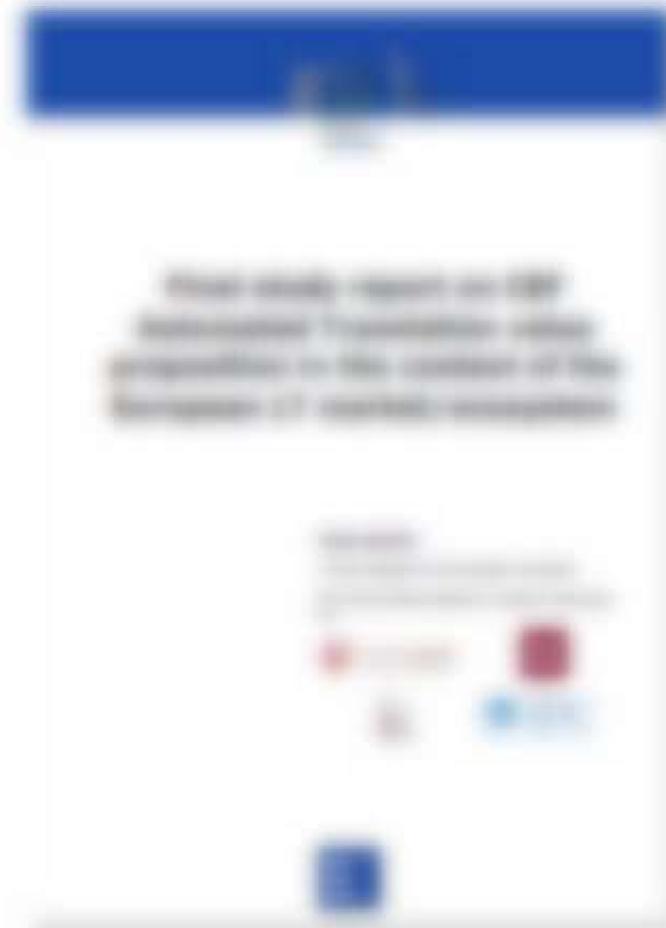
- Europe is strong in research and innovation, but not successful to scale innovations and capture the market

Europe needs European alternatives to fill the gaps and deficiencies and

Multilingual DSM should be developed on its own infrastructure

Public procurement can be the major driver for European LT industry

- A platform is needed to connect demand and supply as well as industry and research

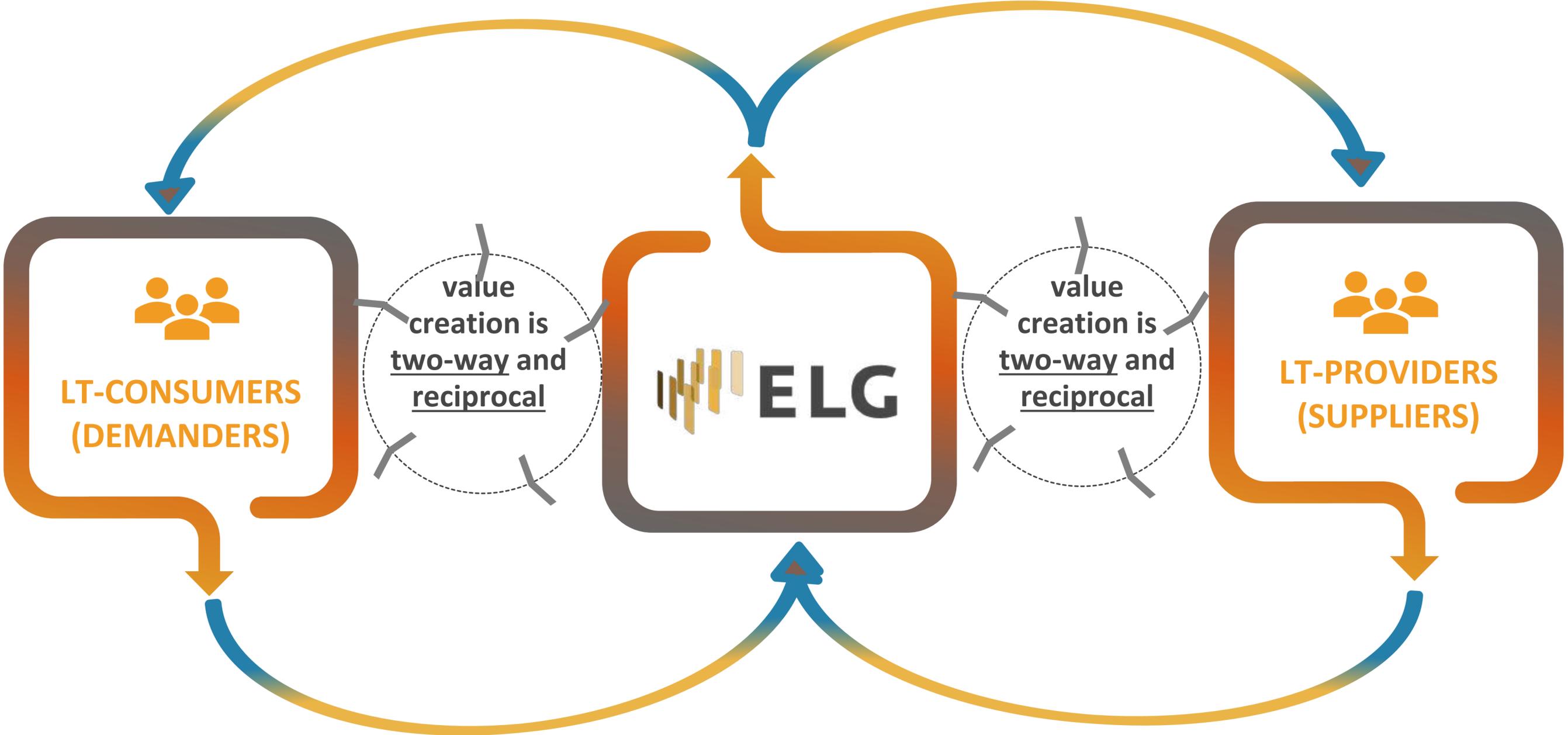


	2015	2016	2018
Translation	21700	21700	22000
Speech	20000	20000	21200
Search	10000	10000	10500
NLP	1000	1000	1000
Total of EU LT	24700	22700	23500
Revenue of EU LT	17000	18000	19000

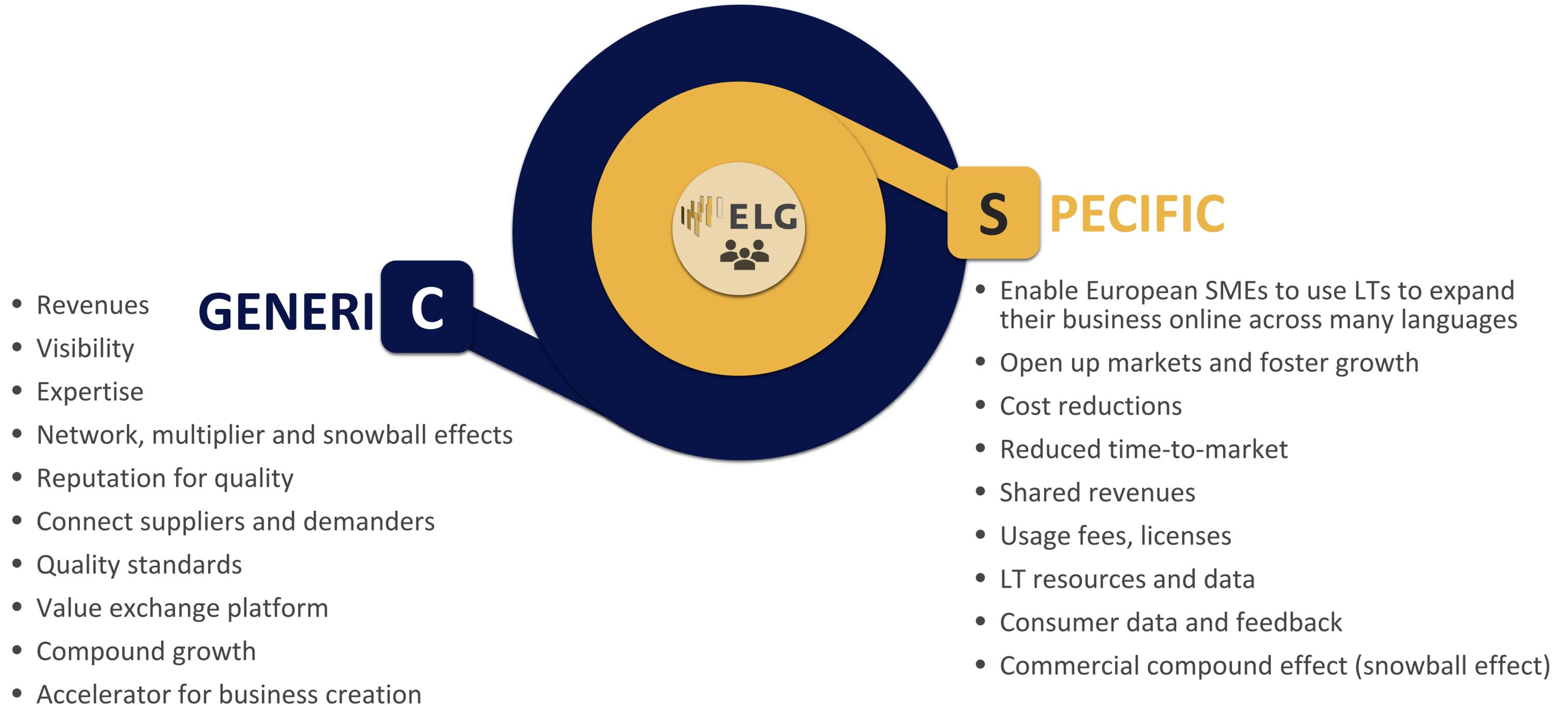
Platform-Driven Business Model of the ELG

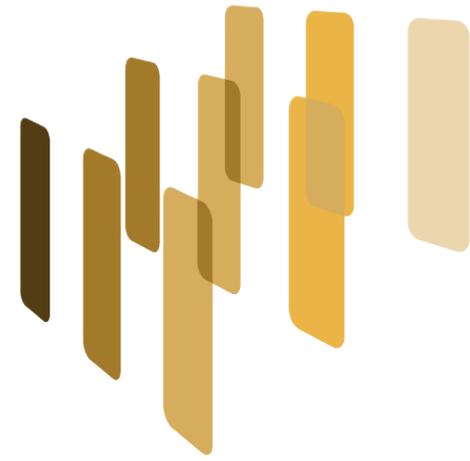
- The world's largest taxi firm, Uber, owns no cars.
- The world's most popular media company, Facebook, creates no content.
- The world's most valuable retailer, Alibaba, carries no stock.
- The world's largest accommodation provider, Airbnb, owns no property.

Platform-Driven Business Model of the ELG



Industry Value Creation (generic >> specific)

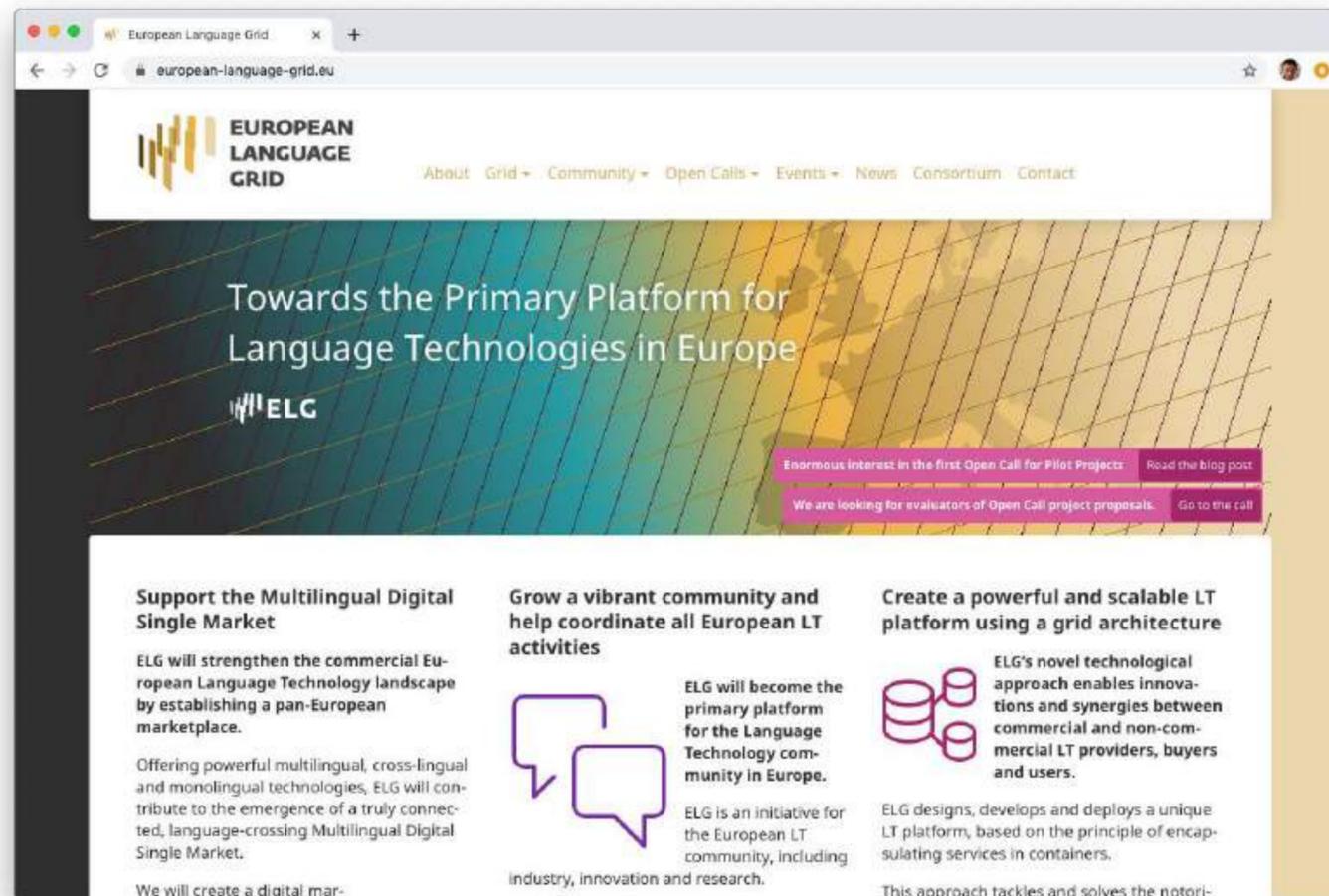




European Language Grid

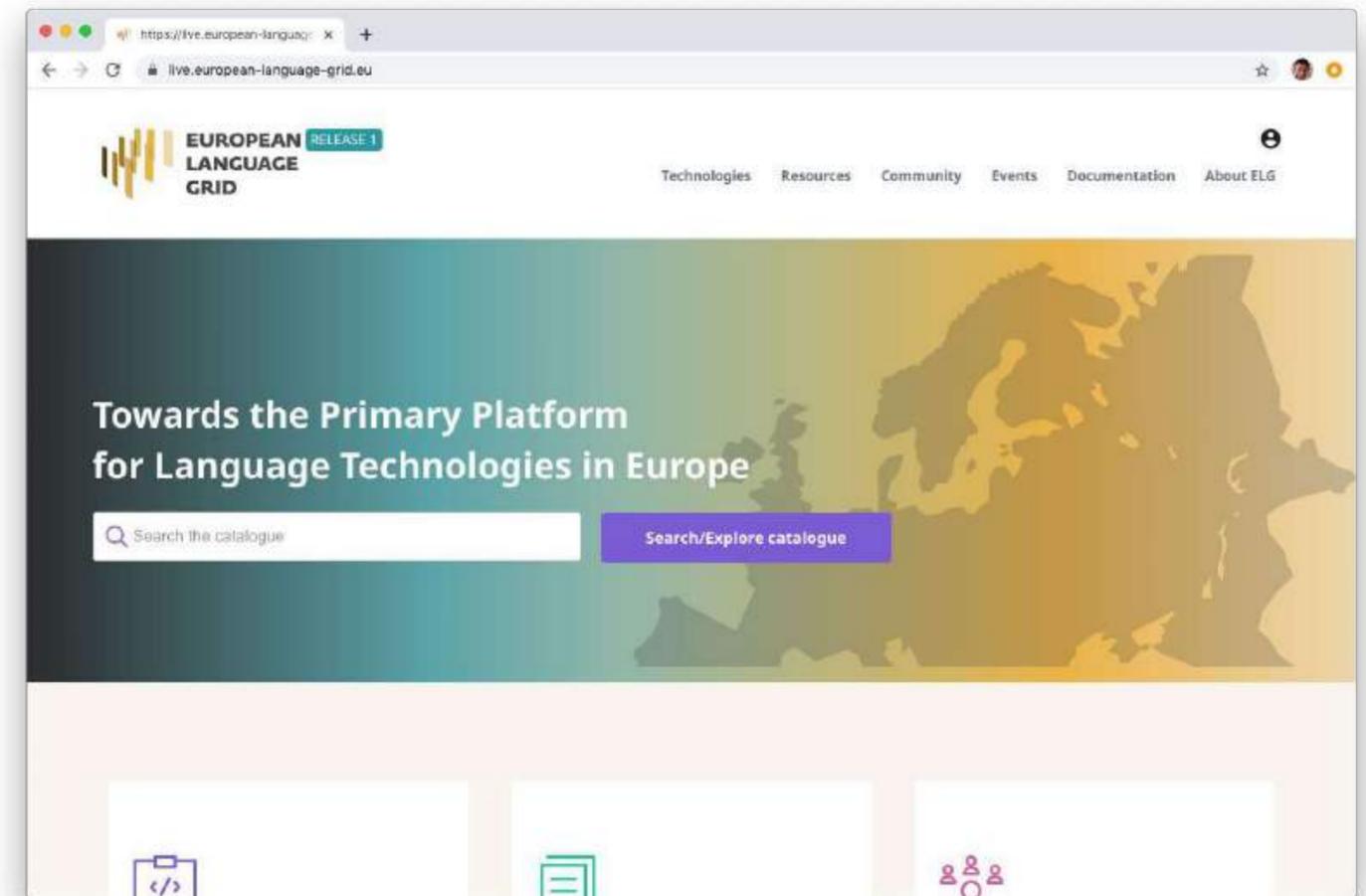
Demo of the Platform

European Language Grid



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

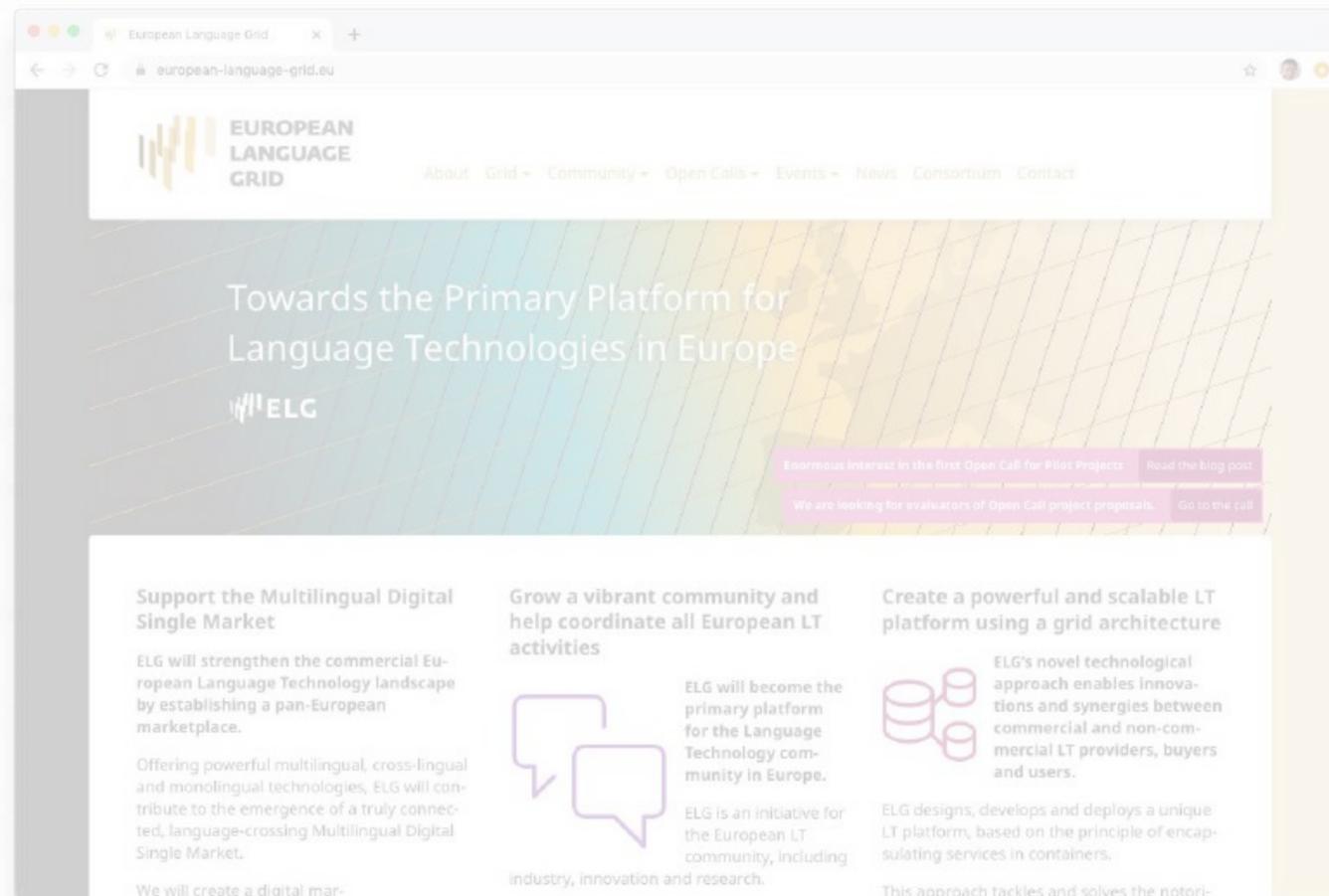
- ELG project website (“Wordpress website”) – set up and managed by WP8
- Typical EU project information plus a lot of additional content (events, open calls, community, NCCs etc.)
- This website needed to be rolled out soon after the start of the project (M3)
- The whole content of the project website will be integrated into the European Language Grid website in early 2021 (“website migration”)



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

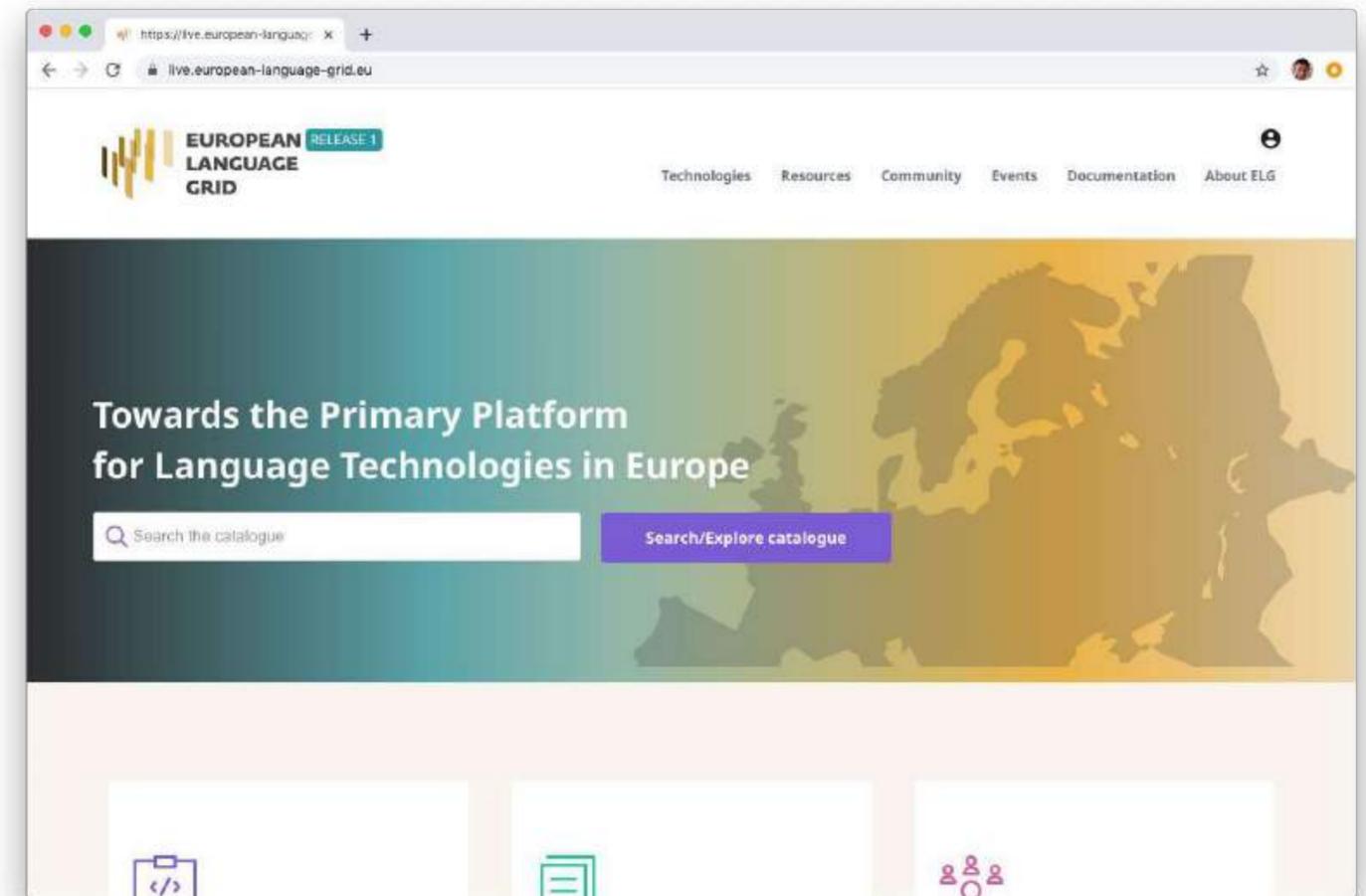
- The actual European Language Grid platform and website
- The next revision of the ELG project website (left) will include dynamic content that will be pulled from the ELG repository (right), e.g., languages, number of services, categories of services etc.
- Developed by WP1, WP2, WP3, WP4, WP5 with additional input from WP6 and WP7

European Language Grid



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

- ELG project website (“Wordpress website”) – set up and managed by WP8
- Typical EU project information plus a lot of additional content (events, open calls, community, NCCs etc.)
- This website needed to be rolled out soon after the start of the project (M3)
- The whole content of the project website will be integrated into the European Language Grid website in early 2021 (“website migration”)



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

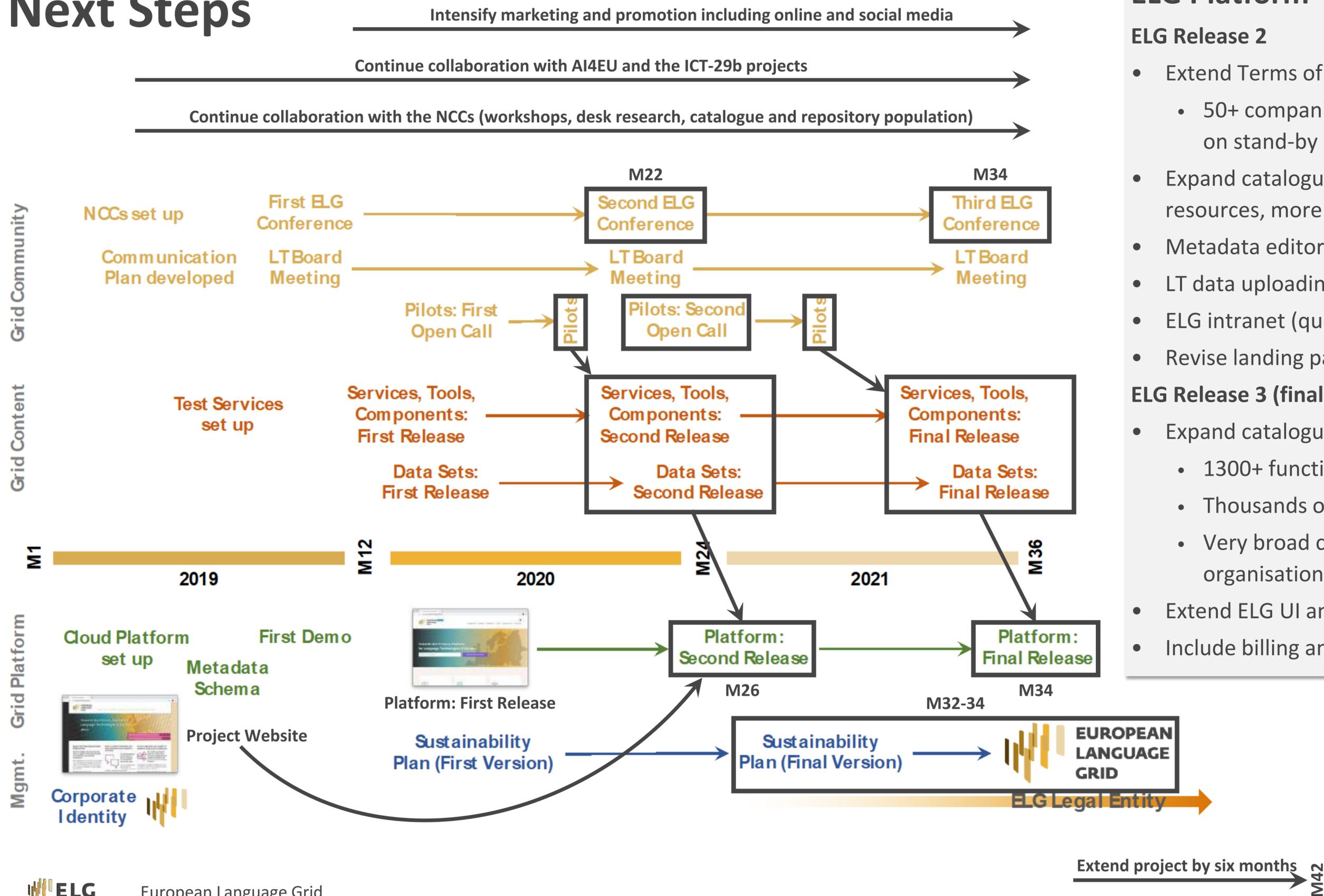
- The actual European Language Grid platform and website
- The next revision of the ELG project website (left) will include dynamic content that will be pulled from the ELG repository (right), e.g., languages, number of services, categories of services etc.
- Developed by WP1, WP2, WP3, WP4, WP5 with additional input from WP6 and WP7



European Language Grid

Next steps

Next Steps



ELG Platform – Selected Next Steps

ELG Release 2

- Extend Terms of Use for 3rd-party services and resources
 - 50+ companies and research organisations on stand-by and eager to contribute
- Expand catalogue and repository: more services, more resources, more organisations, more projects
- Metadata editor and associated features (“claim page”)
- LT data uploading and storage
- ELG intranet (queue for resource ingestion and QA)
- Revise landing page, UI features and fixes

ELG Release 3 (final version during project runtime)

- Expand catalogue and repository – projected final state:
 - 1300+ functional services and tools
 - Thousands of data sets and resources
 - Very broad coverage in terms of languages, organisations (companies, research) and projects
- Extend ELG UI and feature set (including the intranet)
- Include billing and payment

Summary



- Establish ELG as the primary platform and market place for LT in Europe.
- ELG is an initiative *from* the European LT community *for* the European LT community.
- AI and LT are not separate categories. LT **is** AI – and always has been (“Language-centric AI”).
- Global market size by 2025 is enormous: we want the European LT community to be a key player.
- As the marketplace for the European LT business space, ELG will strengthen Europe’s position in this field.
- European LT landscape is extremely fragmented: ELG aims to provide just the right umbrella platform.
- ELG caters for industry and research – because it *has* to in order to reach its goals.
- We want to improve the visibility and reach of all members of the European LT landscape.
- Enable potential buyers (*from all over Europe*) to find the right suppliers (*from all over Europe*).
- Make it very easy for *potential buyers* to experiment with LT, converting them to *actual buyers*.
- ELG is an ambitious, long-term initiative. We will establish a legal entity for sustainability.
- Our field is quite small. We should concentrate on making this initiative a joint success for all of us.
- *At the half-way point we’re happy to report that ELG is running exceptionally well and according to plan.*



European Language Grid

Discussion – Q&A



European Language Grid

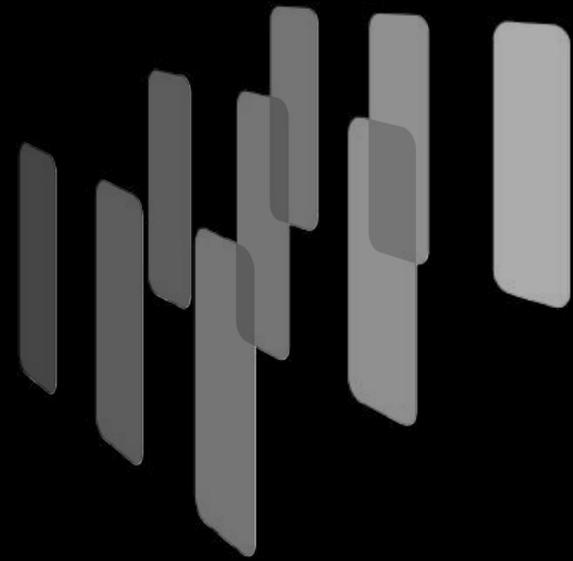
Thank you!



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

Andrejs Vasiljevs (Tilde) – ELG Consortium

21-09-2020 ELG – ELG Workshop – co-located with Baltic HLT (virtual meeting)
<http://www.european-language-grid.eu>



EUROPEAN LANGUAGE GRID

Backup Slides



European Language Grid

- META-NET and META-NET White Papers
- Technical Details
- Gallery
- Pilot Project Call #1: Details
- Sustainability Plan: Details
- ELG Publications



META-NET
(T4ME)

CESAR

META-NORD

METANET4U

META-NET

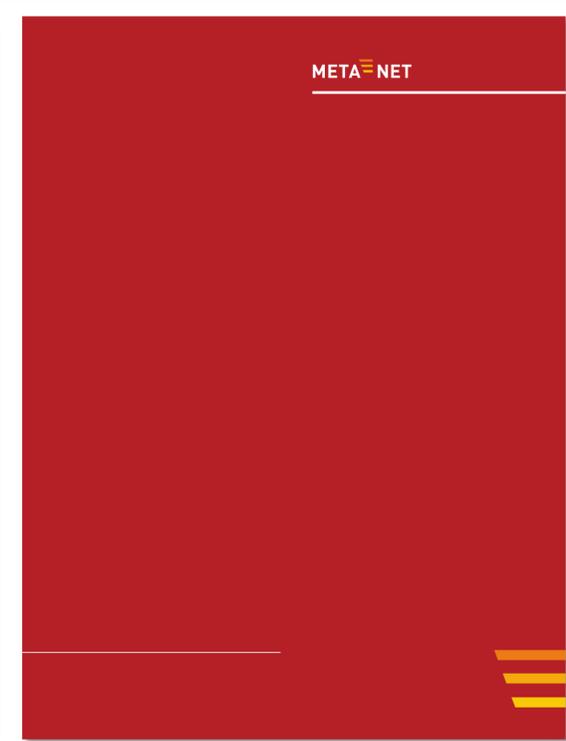
META

- 60 research centres in 34 countries, founded in 2010
- Chair of Executive Board: Jan Hajic (CUNI)
- Deputies: Josef van Genabith (DFKI), Andrejs Vasiljevs (Tilde)
- General Secretary: Georg Rehm (DFKI)

- Multilingual Europe Technology Alliance



Published in 2013



31 volumes, published in 2012



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

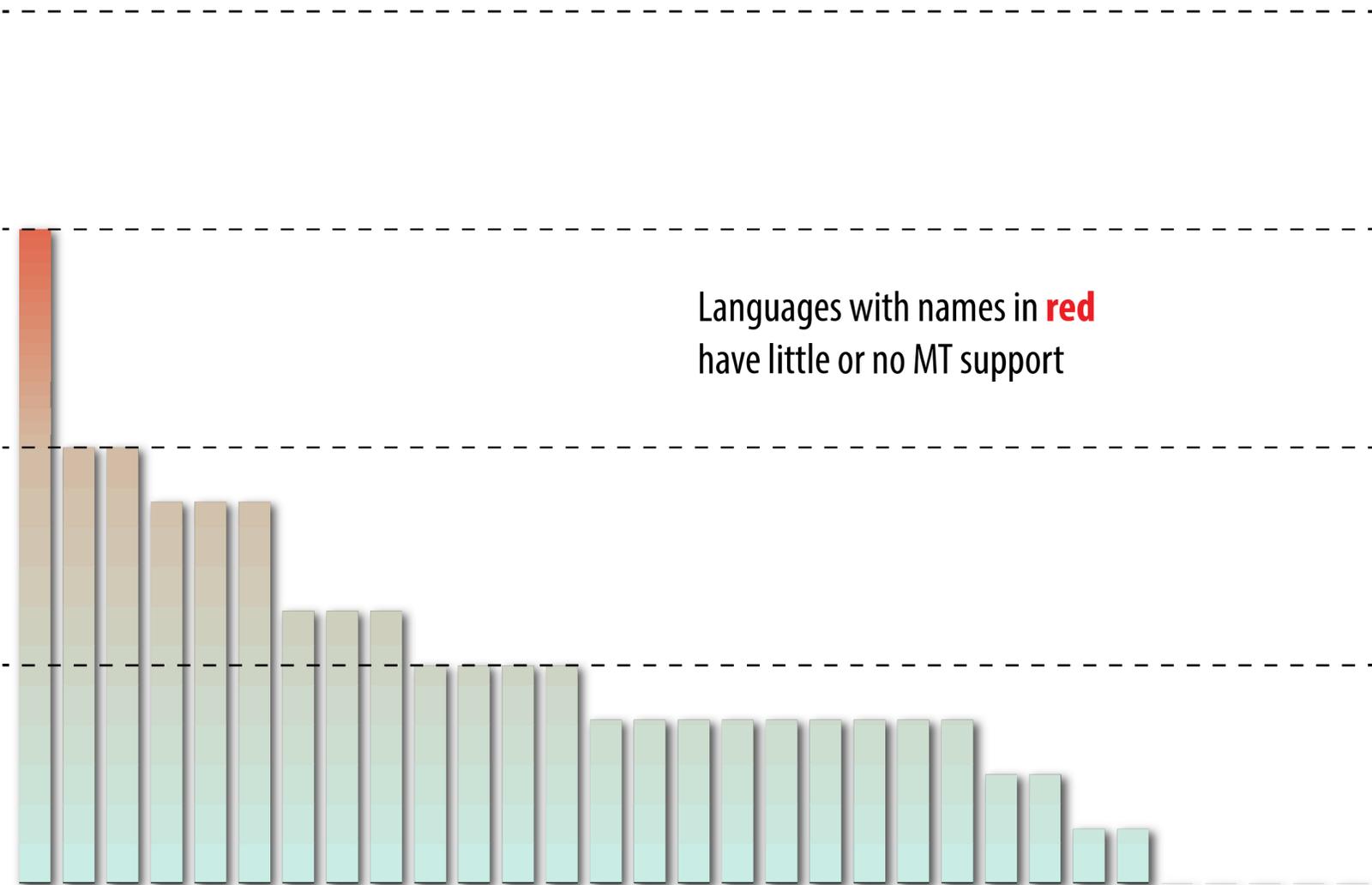


Machine Translation **Text Analytics** **Speech** **Resources**

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



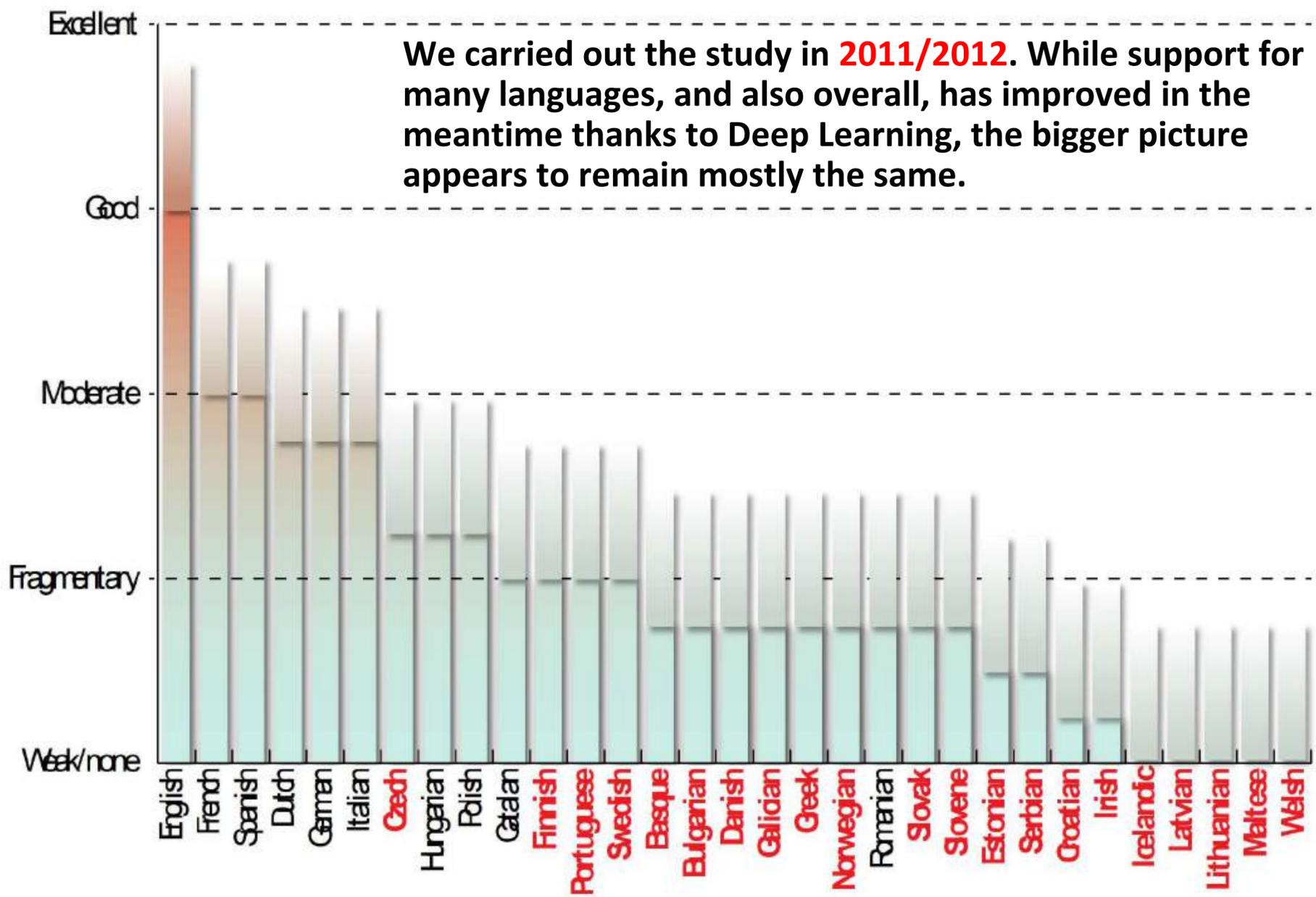
ort



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.



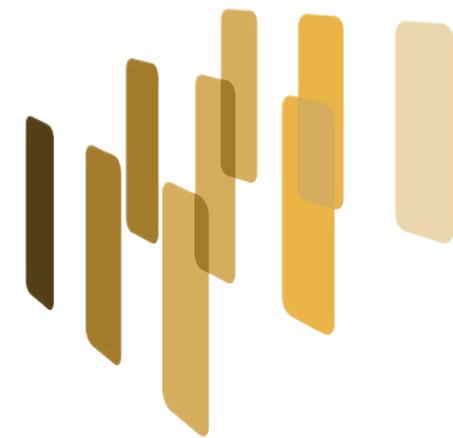
Source: META-NET White Paper Series: Europe's Languages in the Digital Age. Springer, Heidelberg, New York, Dordrecht, London, September 2012. Georg Rehm and Hans Uszkoreit (series editors)

ELG Infrastructure & Backend

- Basic infrastructure: Kubernetes
- Metadata model completed, database model under development
- Catalogue backend implemented (Django)
- All functionalities offered as REST services (JSON)
- PostgreSQL for metadata storage, Elasticsearch engine for indexing
- Authentication and authorisation (Keycloak/Gatekeeper)
- ELG LT service APIs (REST, JSON messages)
- Generic APIs for all LT service classes
- Metadata conversion and harvesting

ELG Frontend

- GUIs for different ELG users, LT providers, LT consumers (AngularJS, React)
- GUI uses ELG catalogue REST services
- LT service registration
- Browse, search incl. faceted search
- Try out UIs
- API documentation and code samples
- CMS with ELG-related content (Drupal)







First ELG hackathon week (DFKI Berlin) – end of May 2019



Second ELG hackathon week (University of Sheffield) – end of August 2019

ELG Open Call #1 – Statistics

121 project proposals submitted

- Formal requirements pre-checked
- 11 proposals rejected (duplicates)
- 110 projects for evaluation

Applicants from 29 different countries

- Outside EU: Iceland, Israel, Norway, Serbia, South Africa, Switzerland, Turkey, United Kingdom

Type A proposals vs. Type B proposals

- 79 project proposals type A (contribute resources, services, tools or data sets to the ELG)
- 31 proposals type B (develop applications using LRs/LTs available in the ELG)

103 applicants

- 62 proposals by SMEs (A: 36; B: 26); 57 unique SMEs
- 48 project proposals by research organizations (A: 43; B: 5); 46 unique research organisations
- 7 applicants submitted two projects (A: 1; B: 1): five SMEs, two research organisations

Total amount requested

- EUR 16.9 million
- Approx. EUR 1.3 million of funding available for this call

Average size of the project

- EUR 153k – *10 projects selected on 29 June 2020*

Sustainability 1/2

Establish ELG as the primary platform and market place for commercial and non-commercial LTs. Enable the European LT community to upload services and data sets, to make them available, to deploy them and also to connect with, and make use of resources made available by others.

To achieve the required scale, we need, among others: high availability and performance; SLAs for (paid) services that are applied in real industrial settings; billing; support etc.

These necessary characteristics of the platform and initiative create various non-trivial costs:

- Cloud hosting (CPU, GPU, RAM, SSD) and bandwidth
- Team: Operations, Development, Accounting, Marketing, Support, Management
- Legal (SLAs, contracts, GDPR etc.)
- Misc. (rent, hardware, electricity etc.)

A sustainable and long-term operational model is required.

Sustainability 2/2

To make ELG successful, we need to identify a way to cover the incurred costs on a long-term basis. We want to find a consensus for a *sustainable operational model*.

One viable option: *establish a legal entity including business and operations plan*.

Basic options: a) for-profit company, b) non-profit company, c) association, d) foundation

Current state of play and timeline:

- Collection and analysis of 10+ potential blueprint/template organisations.
- Business Model Canvas approach: first consolidated BMC is in progress.
- Face-to-face meeting of ELG partners dedicated to sustainability (Q3/2020)
- Testing and further refinement of consolidated BMC with external experts (Q4/2020)
- Arrive at a consensus by approx. Q1/2021.
- Establish legal entity in approx. Q3/2021.

***Potential* Ingredients of Sustainability Plan 1/2**



Regular ads (for companies, services, conferences etc.)

Sponsored content (first result of search/catalogue, marked as “sponsored”)

- Sponsored services
- Sponsored data sets
- Sponsored companies
- Other sponsored catalogue entries

Training events, tutorials, webinars (for a fee for commercial players, free for academia)

Conferences (event registration fees; sponsorship packages for companies)

General consulting services around ELG and language-centric AI (esp. for companies)

Membership fees (if we decide to establish an association or similar)

Project grants (international as well as national projects)

***Potential* Ingredients of Sustainability Plan 2/2**



LTaaS – LT services: hosting of services, models, data sets

(hosting of services or large models creates costs; publicly funded research results to be made available OA; difference between industry and research; costs need to be covered)

- For companies that develop LT – as a primary or secondary dissemination channel, to extend their reach, to open up new markets etc.
- Brokering of commercial LT services (for a fee, split between service owner and ELG)
- For R&D project consortia – as an *additional* dissemination and exploitation channel for industry-relevant research results

PaaS – LT platform functionality: combine ELG services into workflows

RaaS – Hosting service for whole repositories

LT marketplace: matchmaking fee for connecting a buyer with the right supplier

Voices from Industry

- “ELG should enable and foster the development of an LT/AI platform ecosystem.”
- “Don’t duplicate anything.”
- “The ELG platform needs a crystal-clear list of obligations and benefits that is simple and easy to follow.”
- “Platform interoperability: ELG needs to complement existing or emerging clouds.”
- “By combining our offers with those of other European players we can reach new business opportunities that we can’t reach alone.”
- “Create synergies with the AI4EU platform: ELG as the language branch of the AI4EU platform.”
- “Provide a distribution and dissemination channel to the EU Member States so that the results of their funding programmes can be disseminated internationally efficiently.”
- “ELG as an additional distribution channel to existing channels.”
- “ELG as a channel to get and to make business.”
- “ELG should foster continuous activity and involvement with LT in Europe.”
- “Emphasise the combination aspect, combine different services, data sets, tools, from different companies.”
- “Combination of services and data sets from *multiple* vendors: combine these services into flexible workflows.”

Selected Publications



- Georg Rehm, Maria Berger, Ela Elsholz, Stefanie Hegele, Florian Kintzel, Katrin Marheinecke, Stelios Piperidis, Miltos Deligiannis, Dimitris Galanis, Katerina Gkirtzou, Penny Labropoulou, Kalina Bontcheva, David Jones, Ian Roberts, Jan Hajic, Jana Hamrlová, Lukáš Kačena, Khalid Choukri, Victoria Arranz, Andrejs Vasiljevs, Oriens Anvari, Andis Lagzdiņš, Jūlija Meļņika, Gerhard Backfried, Erinç Dikici, Miroslav Janosik, Katja Prinz, Christoph Prinz, Severin Stampfer, Dorothea Thomas-Aniola, José Manuel Gómez Pérez, Andres Garcia Silva, Christian Berrío, Ulrich Germann, Steve Renals, and Ondrej Klejch. **European Language Grid: An Overview**. In Nicoletta Calzolari et al., editors, *Proc. of the 12th Language Resources and Evaluation Conf. (LREC 2020)*, pages 3359-3373, Marseille, France, 2020.
- Penny Labropoulou, Katerina Gkirtzou, Maria Gavriilidou, Miltos Deligiannis, Dimitris Galanis, Stelios Piperidis, Georg Rehm, Maria Berger, Valérie Mapelli, Michael Rigault, Victoria Arranz, Khalid Choukri, Gerhard Backfried, José Manuel Gómez Pérez, and Andres Garcia-Silva. **Making Metadata Fit for Next Generation Language Technology Platforms: The Metadata Schema of the European Language Grid**. In Nicoletta Calzolari et al. editors, *Proc. of the 12th Language Resources and Evaluation Conf. (LREC 2020)*, pages 3421-3430, Marseille, France, 2020.
- Georg Rehm, Katrin Marheinecke, Stefanie Hegele, Stelios Piperidis, Kalina Bontcheva, Jan Hajic, Khalid Choukri, Andrejs Vasiljevs, Gerhard Backfried, Christoph Prinz, José Manuel Gómez Pérez, Luc Meertens, Paul Lukowicz, Josef van Genabith, Andrea Lösch, Philipp Slusallek, Morten Irgens, Patrick Gatellier, Joachim Köhler, Laure Le Bars, Dimitra Anastasiou, Albina Auksoriūtė, Núria Bel, António Branco, Gerhard Budin, Walter Daelemans, Koenraad De Smedt, Radovan Garabík, Maria Gavriilidou, Dagmar Gromann, Svetla Koeva, Simon Krek, Cvetana Krstev, Krister Lindén, Bernardo Magnini, Jan Odijk, Maciej Ogrodniczuk, Eiríkur Rögnvaldsson, Mike Rosner, Bolette Pedersen, Inguna Skadina, Marko Tadić, Dan Tufiş, Tamás Váradi, Kadri Vider, Andy Way, and François Yvon. **The European Language Technology Landscape in 2020: Language-Centric and Human-Centric AI for Cross-Cultural Communication in Multilingual Europe**. In Nicoletta Calzolari et al., editors, *Proc. of the 12th Language Resources and Evaluation Conf. (LREC 2020)*, pages 3315-3325, Marseille, France, 2020.
- Georg Rehm, Dimitrios Galanis, Penny Labropoulou, Stelios Piperidis, Martin Weiß, Ricardo Usbeck, Joachim Köhler, Miltos Deligiannis, Katerina Gkirtzou, Johannes Fischer, Christian Chiarcos, Nils Feldhus, Julián Moreno-Schneider, Florian Kintzel, Elena Montiel, Víctor Rodríguez Doncel, John P. McCrae, David Laqua, Irina Patricia Theile, Christian Dittmar, Kalina Bontcheva, Ian Roberts, Andrejs Vasiljevs, and Andis Lagzdiņš. **Towards an Interoperable Ecosystem of AI and LT Platforms: A Roadmap for the Implementation of Different Levels of Interoperability**. In Georg Rehm, Kalina Bontcheva, Khalid Choukri, Jan Hajic, Stelios Piperidis, and Andrejs Vasiljevs, editors, *Proc. of the 1st Int. Workshop on LT Platforms (IWLTP 2020)*, pages 96-107, Marseille, France, 2020. 16 May 2020.
- Georg Rehm, Kalina Bontcheva, Khalid Choukri, Jan Hajic, Stelios Piperidis, and Andrejs Vasiljevs, editors. **Proceedings of the 1st International Workshop on Language Technology Platforms (IWLTP 2020, co-located with LREC 2020)**, Marseille, France, 2020. 16 May 2020