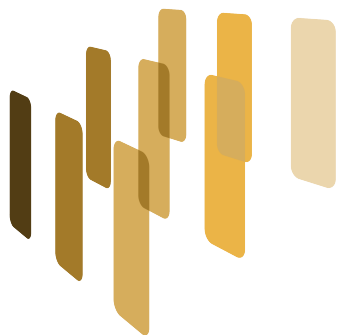




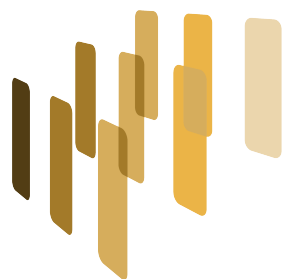
EUROPEAN LANGUAGE GRID

META^{NET}
META^{FORUM} 2020

European Language Grid: Broad Overview – Services and Resources – Platform

Georg Rehm (DFKI), Kalina Bontcheva (University of Sheffield), Stelios Piperidis (ILSP/R.C. Athena)
georg.rehm@dfki.de, k.bontcheva@sheffield.ac.uk, spip@athenarc.gr

01/02/03-12-2020 META-FORUM 2020 – Piloting the European Language Grid (virtual conference)
<http://www.european-language-grid.eu>



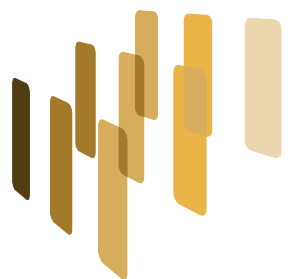
EUROPEAN LANGUAGE GRID



Kick-off meeting, 22/23 January 2019

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.



EUROPEAN LANGUAGE GRID



Consortium

Very good working relationship with all relevant initiatives

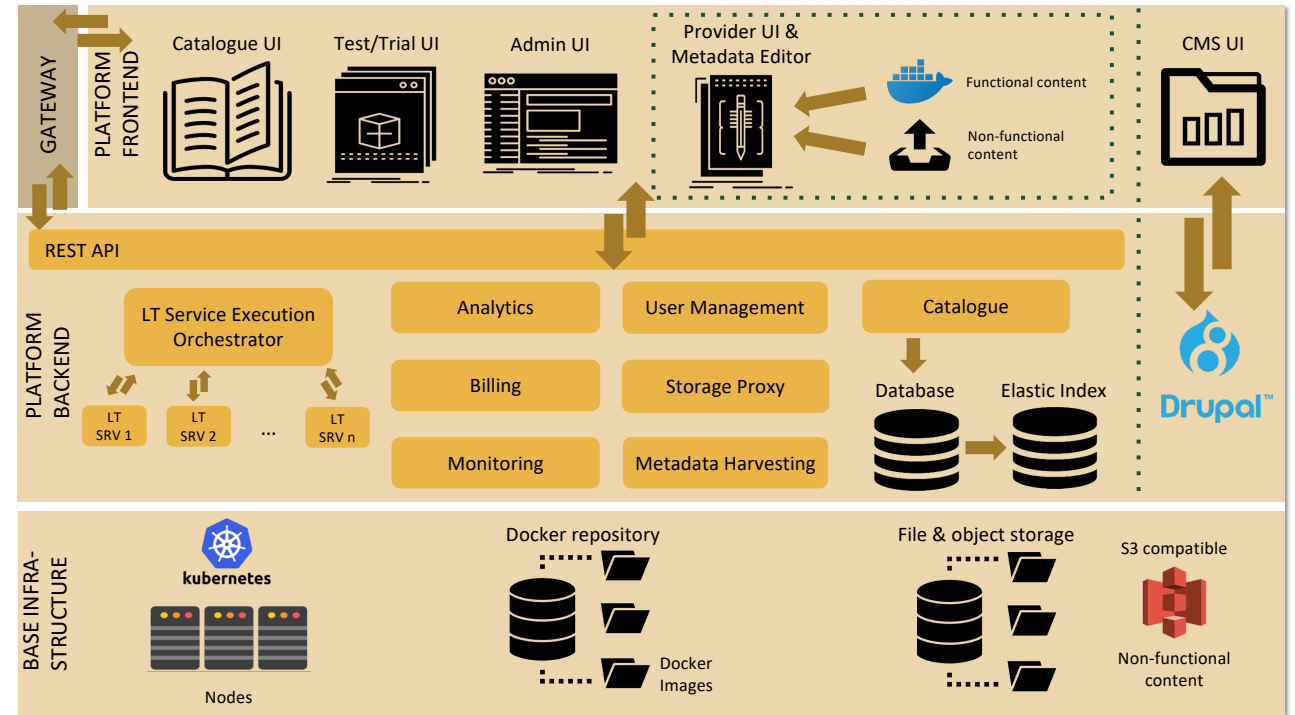
- META-NET and META
- LT Innovate
- AI4EU and its various satellite projects (HumanE AI Net, VISION etc.)
- CLAIRE, emerging AI PPP and other European AI initiatives
- ELRC, CEF AT, e-translation
- CLARIN
- Big Data Value Association
- EOSC *and others*



Kick-off meeting, 22/23 January 2019

European Language Grid – Release 1 (June/July 2020)

- 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
- Online documentation
- ELG Release 1 (June/July 2020):
 - 539 entries in the ELG catalogue
 - 161 functional services and tools
 - 358 data sets and language resources



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

→ Kalina Bontcheva: ELG LT services and resources

→ Stelios Piperidis: ELG platform and demonstration

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

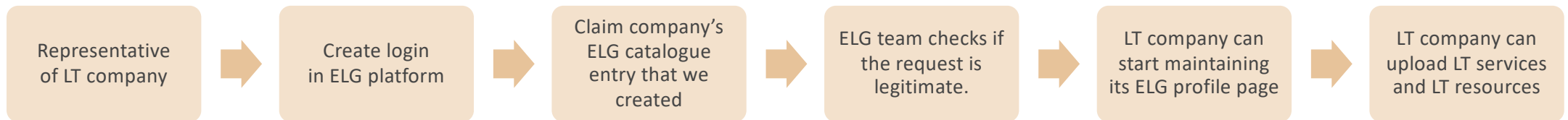


META-FORUM 2019 (8/9 October) – Brussels, Belgium

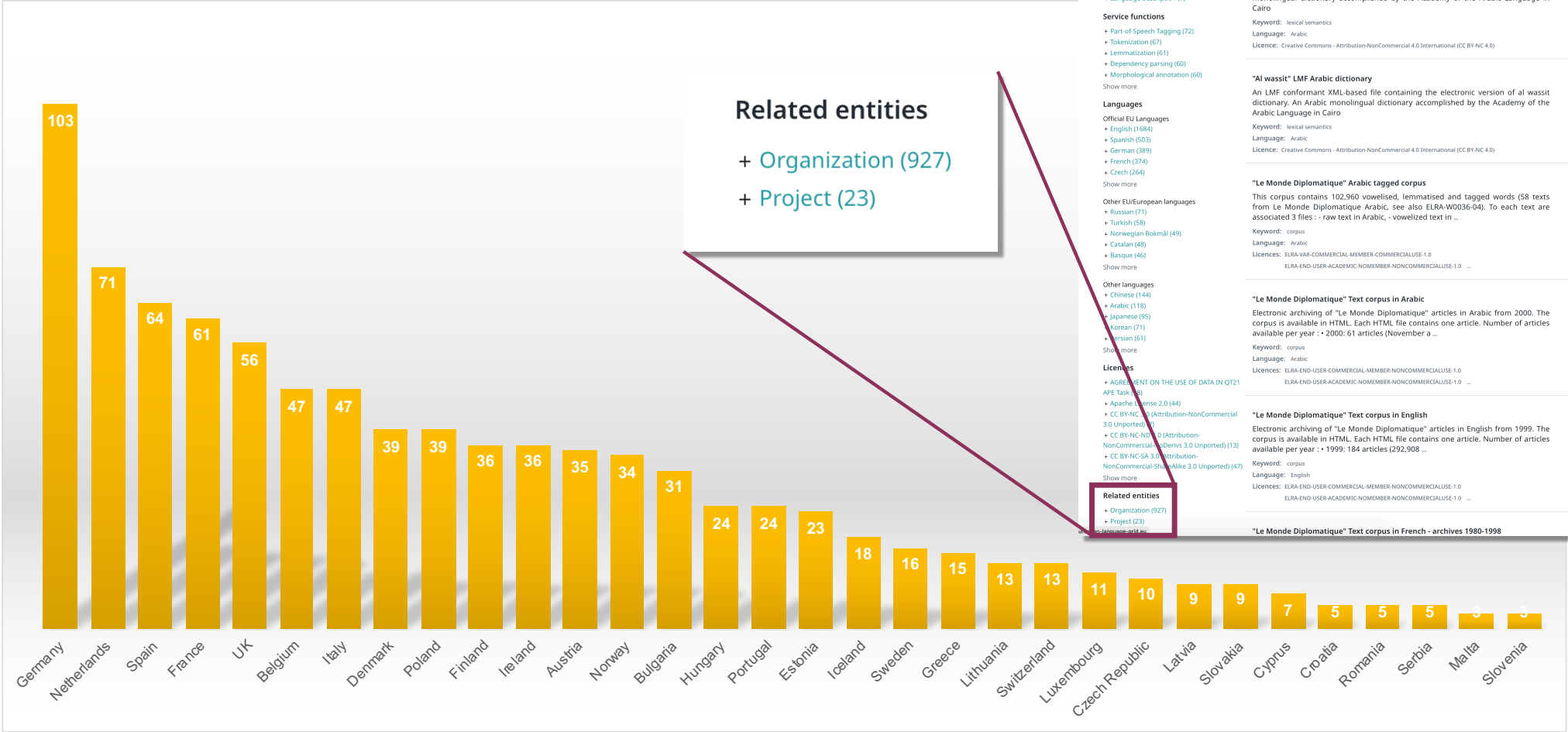
European LT Industry in the ELG



- European LT industry is big and fragmented.
- Many companies operate in niches – based on application type, language or region.
- We need to bring European LT industry together fully to exploit its potential.
- ELG – with a lot of help from the 32 National Competence Centres – has compiled lists of companies and academic organisations that develop LT in all European countries.
- Thanks are also due to Yulia Akhulkova (Nimdzi) for her input from the *Nimdzi LT Atlas*!
- Today we make available the first result of this combined effort:
A comprehensive list of 900+ European LT companies in the ELG.
- Why is this important? To bootstrap the ELG catalogue and the eventual ELG community.



European LT Industry in the ELG



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)

Show more

Other European languages

- + Russian (71)
- + Turkish (58)
- + Norwegian Bokmål (49)
- + Catalan (48)
- + Basque (46)

Show more

Other languages

- + Chinese (144)
- + Arabic (118)
- + Japanese (95)
- + Korean (71)
- + Persian (61)

Show more

Licences

- + AGREEMENT ON THE USE OF DATA IN QT21 APE Task (8)
- + Apache License 2.0 (44)
- + CC BY-NC-ND (Attribution-NonCommercial 3.0 Unported) (1)
- + CC BY-NC-ND (Attribution-NonCommercial-ShareAlike 3.0 Unported) (13)
- + CC BY-NC-SA 3.0 (Attribution-NonCommercial-ShareAlike 3.0 Unported) (47)

Show more

Related entities

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

3710 search results

"Al wassit" Arabic dictionary

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)

Lexical/Conceptual resource
114 views

"Al wassit" LMF Arabic dictionary

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)

Lexical/Conceptual resource
21 views

"Le Monde Diplomatique" Arabic tagged corpus

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, - vowelised text in ...

Keyword: corpus

Language: Arabic

Licences: ELRA-VAR-COMMERCIAL-MEMBER-COMMERCIALUSE-1.0

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

Corpus
1 views

"Le Monde Diplomatique" Text corpus in Arabic

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 ...

Corpus
1 views

"Le Monde Diplomatique" Text corpus in English

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 ...

Corpus
1 views

"Le Monde Diplomatique" Text corpus in French - archives 1980-1998

Corpus

Stakeholders and Users and Collaborators

ELG Open Calls: approx. 15 pilot projects

ICT-29b research and innovation projects

ELE – European Language Equality

AI4EU – European AI on demand platform

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

Core networks: META-NET, LT Innovate

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



EUROPEAN
LANGUAGE
EQUALITY

Prêt-à-LLOD

Gourmet

Global Under Resourced Media Translation



Community: NCCs and LTC

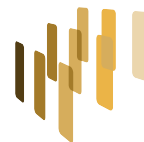
32 National Competence Centres (NCCs)

- 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.*

→ NCC session (tomorrow, 09:30-10:45)



Austria



DAGMAR GROMANN

Centre for Translation Studies, University of Vienna

Belgium



WALTER DAELEMANS

Computational Linguistics and Psycholinguistics, University of Antwerp

Bulgaria



SVETLA KOEVA

Institute for Bulgarian Language, Bulgarian Academy of Sciences

Croatia



MARKO TADIĆ

Department of Linguistics, University of Zagreb

Cyprus



DORA LOIZIDOU

University of Cyprus

Czech Republic



JAN HAJIČ

Institute of Formal and Applied Linguistics, Charles University Prague

Denmark



BOLETTE SANDFORD PEDERSEN

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

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
- Concept prepared (Business Model Canvas).
- Q1/2022: Establish legal entity (probably with a soft start)

→ Sustainability session (Thursday, 11:15-11:45)



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 orgs.**

➔ Open Call and Pilot Project session (today, 10:45-12:00)

Open Call #1 – Statistics

110 project proposals evaluated

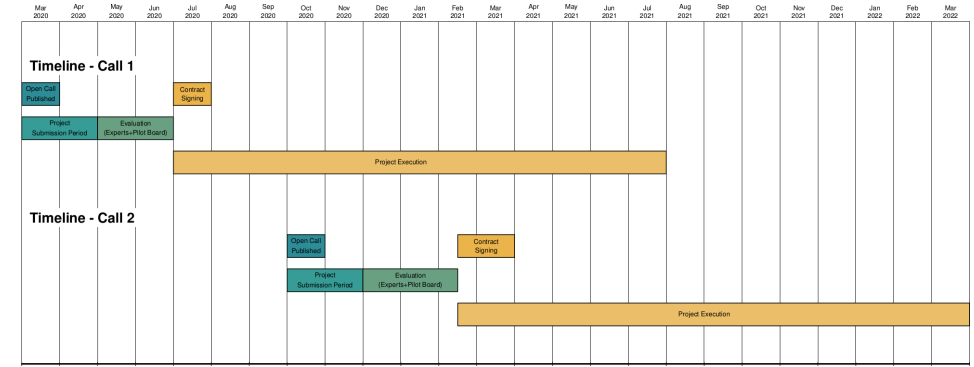
- **Type A:** 79 proposals
- **Type B:** 31 proposals

Total amount requested: 16.9M€

10 projects selected on 29 June 2020

Open Call #2 – Statistics

106 project proposals submitted



European Language Grid: An Overview

Georg Rehm¹, Maria Berger¹, Ela Elsholz², Stefanie Hegel³, Florian Kintzel⁴, Katrin Marheinecke⁵, Stelios Piperidis⁶, Miltos Deligiannis⁷, Dimitrios Galanis⁸, Katerina Gkirtzou⁹, Penny Labropoulou¹⁰, Kalina Bontcheva¹¹, David Jones¹², Ian Roberts¹³, Jan Hajič¹⁴, Jana Hamřlová¹⁵, Lukáš Kačena¹⁶, Khalid Choukri¹⁷, Victoria Arranz¹⁸, Andrejs Vasiljevs¹⁹, Orians Anvari²⁰, Andis Lagzdins²¹, Jūlija Melnikā²², Gerhard Backfried²³, Erinc Dikić²⁴, Miroslav Janasik²⁵, Katja Prinz²⁶, Christoph Frinz²⁷, Severin Stampel²⁸, Dorothea Thomas-Aniola²⁹, José Manuel Gómez Pérez³⁰, Andres Garcia Silva³¹, Cristian Berriol³², Ulrich Germann³³, Steve Renals³⁴, Ondrej Klejch³⁵

¹ DFKI GmbH, Germany • ² ILSP/Athena RC, Greece • ³ University of Sheffield, UK • ⁴ Charles University, Czech Republic • ⁵ Evaluations and Language Resources Distribution Agency (ELDA), France • ⁶ Tilde, Latvia • ⁷ SAIL LABS Technology GmbH, Austria • ⁸ Expert System Iberia SL, Spain • ⁹ University 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 a world-class, with technologies that outperform the global players. It 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, LT community to deposit and upload their technologies and data sets into their 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 (LTC) for out

Keywords: LR Infrastructures and Architectures, LR National Interactions

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 (Rehm and Uszkoreit, 2012; Komai, 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 Hegel, 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

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 Marheinecke², Stefanie Hegel³, Stelios Piperidis⁴, Kalina Bontcheva⁵, Jan Hajič⁶, Khalid Choukri⁷, Andrejs Vasiljevs⁸, Gerhard Backfried⁹, Christoph Prinz¹⁰, José Manuel Gómez Pérez¹¹, Luc Meertens¹², Paul Lukowicz¹³, Josef van Genabith¹⁴, Andrea Linsch¹⁵, Philipp Shussek¹⁶, Morten Irgens¹⁷, Patrick Gattelier¹⁸, Joachim Köhler¹⁹, Laure Le Bars²⁰, Dimitria Anastasiou²¹, Albina Aukoraitė²², Nürja Bel²³, António Branco²⁴, Gerhard Bodin²⁵, Walter Daelemans²⁶, Koenraad De Smedt²⁷, Radovan Garabik²⁸, Maria Gavrilidou²⁹, Dagmar Gromann³⁰, Svetla Kovac³¹, Simon Krek³², Cveta Krstev³³, Kristin Lindén³⁴, Bernardo Maguin³⁵, Jan Odijk³⁶, Maciej Ogrodniczak³⁷, Erikur Rignavaldson³⁸, Mike Rosner³⁹, Bettina Sandford Pedersen⁴⁰, Jagna Skadipa⁴¹, Marko Tadić⁴², Dan Tufuș⁴³, Tamás Váradi⁴⁴, Kadri Vider⁴⁵, Andy Way⁴⁶, François Yvon⁴⁷

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

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

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

Abstract

Language is the cornerstone of Europe and firmly anchored in the European treaties including its impacting business, cross-lingual and cross-cultural communication are still powerful means to break down these barriers. While the last decade has seen various and technologies tailored to Europe's specific needs, there is still an immense level of fragmentation in the European Information and Communication Technology landscape, describing funding programmes, activities, actions and efforts, including the current state of play in industry and the LT market. We present a high-level overview of the European Language Technology landscape in 2020, including the EU level in the last ten years and develop strategic guidance with regard to four key areas: international projects, infrastructure issues, policy issues, infrastructures, multilingualism and cross-cultural communication.

Introduction

Research in Language Technology (LT) has not only enabled a highly visible and growing research community but also a thriving LT industry. Commercial LTs are indispensable in our day-to-day lives, from mobile communication to business and government, and, thus, a key to the success of the European Union. The LT industry is crucial in Europe's multilingual and trade partners. The European LT landscape is in a state of rapid change, with new players entering the market, and existing players expanding their reach. This paper provides an overview of the European Language Technology landscape in 2020, including the EU level in the last ten years and develop strategic guidance with regard to four key areas: international projects, infrastructure issues, policy issues, infrastructures, multilingualism and cross-cultural communication.

Since 2010, the topic has been receiving increasing attention, recently also increasingly in the commercial LT space in Europe. This paper provides an overview of the European Language Technology landscape in 2020, including the EU level in the last ten years and develop strategic guidance with regard to four key areas: international projects, infrastructure issues, policy issues, infrastructures, multilingualism and cross-cultural communication.

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

Georg Rehm¹, Dimitrios Galanis², Penny Labropoulou³, Stelios Piperidis⁴, Martin Well⁵, Ricardo Usbeck⁶, Joachim Köhler⁷, Miltos Deligiannis⁸, Katerina Gkirtzou⁹, Johannes Fischer¹⁰, Christian Chiaros¹¹, Nils Feldhus¹², Julián Moreno-Schneider¹³, Florian Kintzel¹⁴, Elena Montiel¹⁵, Victor Rodriguez Donce¹⁶, John P. McCrac¹⁷, David Laqua¹⁸, Irina Patricia Theile¹⁹, Christian Dittmar²⁰, Kalina Bontcheva²¹, Ian Roberts²², Andrejs Vasiljevs²³, Andis Lagzdins²⁴

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

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

Abstract

AI, we concentrate on two core aspects: (1) cross-platform search and uniform service workflows. We devise five different levels (or increasing levels) of interoperability, from the lowest level of interoperability (implementing a wider federation of AILT platforms) to the highest level of interoperability (implementing a wider federation of AILT platforms). We illustrate the different levels of interoperability.

International Projects, Tools, Systems, Applications, Web Services

from their respective unique offerings. Only by discussing and agreeing upon standards as well as technical and operational concepts for AILT 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 A4EU 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 pre-requisites 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 A4EU (Section 2.1), ELG (Section 2.2), QURATOR (Section 2.3), Lynx (Section 2.4) and SPEAKER (Section 2.5).

2.1. A4EU

In January 2019, the A4EU 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) and the design of a European AI on-demand platform that provides 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

- Three publications on ELG or ELG-related topics at LREC 2020.
- Please cite these papers if you want to refer to ELG in publications.
- IWLTTP 2020 workshop organised at LREC 2020 (30 submissions, 17 papers).
- Coming up in Q1/2021: CFP for special issue of LRE journal on LT platform topics.

European Language Equality (ELE)



**EUROPEAN
LANGUAGE
EQUALITY**



**EUROPEAN
LANGUAGE
GRID**

Consortium: 53 partners from all over Europe.

Coordinator: ADAPT Centre (Dublin City University).

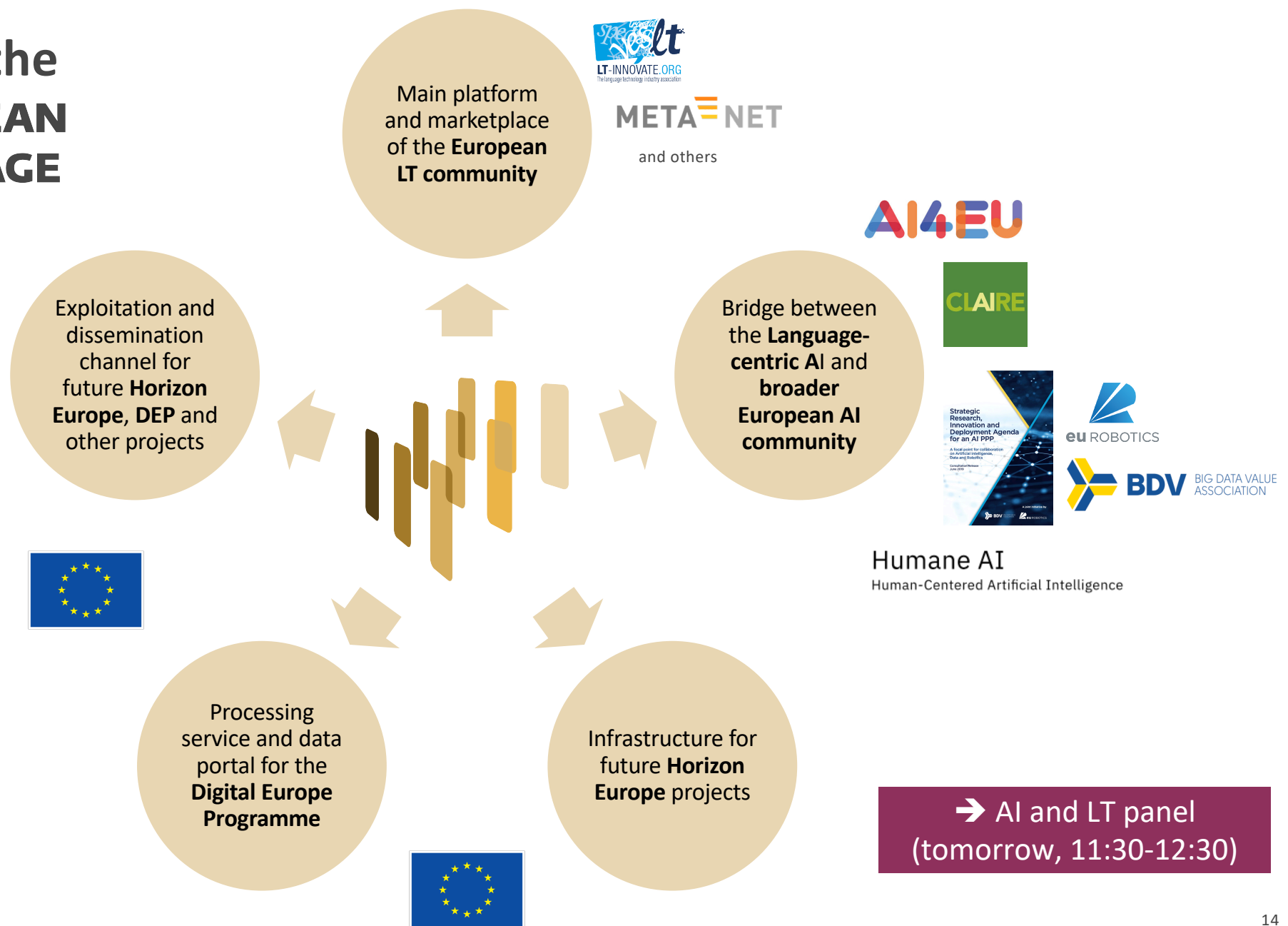
Objective: *Development of a strategic research, innovation and deployment agenda to achieve digital language equality in Europe by 2030.*

Runtime: 18 months – ELE and ELG will both end in June 2022.

Start on 1 January 2021.

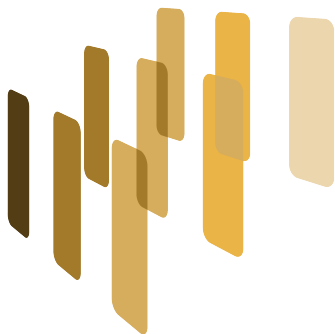
→ ELE session (tomorrow, 09:00-09:30)

Future roles of the EUROPEAN LANGUAGE GRID



Summary and Next Steps

- Establish ELG as the primary platform and marketplace for Language Technology in Europe.
- An initiative *from* the European LT community *for* the European LT community.
- European LT landscape is **highly fragmented**: ELG aims to provide just the right **umbrella platform**.
- Global market size by 2025 is enormous: we want the European LT community to be a **key player**.
- We want to **increase the visibility and reach** of all members of the European LT landscape.
- ELG is a long-term initiative: we will establish a **legal entity** for sustainability.
- Contribute to the long-term goal of **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**: develop and populate **ELG Release 2** with functional services and resources (early 2021); develop **ELG Release 3** (early 2022); establish the **ELG legal entity** (late 2021).



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).

Georg Rehm (DFKI), Kalina Bontcheva (University of Sheffield), Stelios Piperidis (ILSP/R.C. Athena)
georg.rehm@dfki.de, k.bontcheva@sheffield.ac.uk, spip@athenarc.gr

01/02/03-12-2020 META-FORUM 2020 – Piloting the European Language Grid (virtual conference)
<http://www.european-language-grid.eu>