



EUROPEAN LANGUAGE GRID

European Language Grid: An Overview

Georg Rehm (DFKI) – ELG Coordinator
georg.rehm@dfki.de

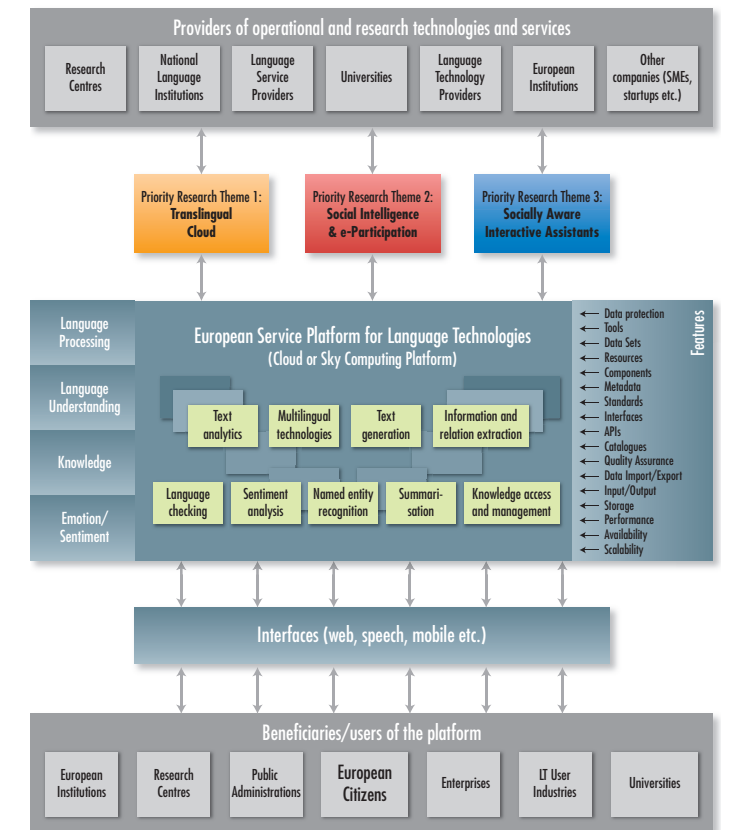
08-10-2019, META-FORUM 2019 – Brussels, Belgium
<http://www.european-language-grid.eu>

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.
- **META³NET** vision: **European Service Platform for Language Technologies**.
- First mention in the **Strategic Research Agenda for Multilingual Europe 2020** (published in early 2013).



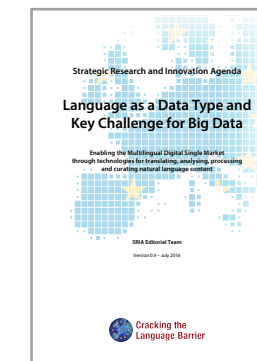
META³ FORUM 2013



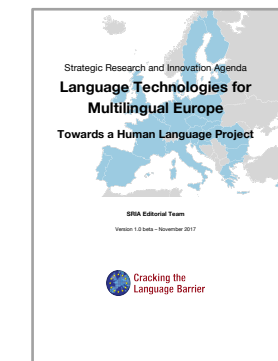
The concept was later refined in the three LT SRIAs (2015, 2016, 2017).



META³ FORUM 2015



META³ FORUM 2016



META³ FORUM 2017

Why “European Language Grid”?

Call ICT-29-2018 – The action shall:

i. develop the architecture and components for a public, open and interoperable grid connecting resources and tools, sharing and combining resources to support effective development and deployment of [LT] (software and services) across Europe. It shall provide **easy access to basic natural language processing tools and services for European languages**. The action shall cater for both consolidation of existing and a seamless inclusion of new resources and tools available for free or/and for a fee [...].

ii. coordinate the work of the European Language Grid and all actions supported under this topic and address the interoperability issues. It shall identify barriers for deploying multilingual services and **establishing language infrastructure at European scale** [...]. The action shall [...] facilitate **coordination between various European, national and regional activities** through a **structured dialogue** [...].

iii. pilot the European Language Grid in specific sectors of high commercial and/or societal impact, through **small scale demonstrators** geared towards an innovative integration of [LT] in specific operating processes/operations. [...] (Financial Support for Third Parties)

iv. establish competence centres [...] in Member and Associated States. It shall build on the previous EC-funded actions ([...] FP7, H2020, CEF).

Oct 27, 2017

A multilingual Next Generation Internet

ID: ICT-29-2018

Type of action:

◦ RIA Research and Innovation action , IA Innovation action

Deadline Model : single-stage

Opening: **31 October 2017**

Deadline: **17 April 2018 17:00:00**
Brussels time

Closed

a) Innovation Action: A European Language Grid

The action shall:

i. develop the architecture and components for a public, open and interoperable grid connecting resources and tools, sharing and combining resources to support effective development and deployment of language technologies (software and services) across Europe. It shall provide easy access to basic natural language processing tools and services for European languages. The action shall cater for both consolidation of existing and a seamless inclusion of new resources and tools available for free or/and for a fee, enabling providers to control access rights reflecting their policies. The end-users of the grid shall be closely involved in the process.

ii. coordinate the work of the European Language grid and all actions supported under this topic and address the interoperability issues. It shall identify barriers for deploying multilingual services and establishing language infrastructure at European scale, including any skills gap. The action shall address legal and organisational obstacles, facilitate coordination between various European, national and regional activities through a structured dialogue and the establishment and exchange of best practices.

iii. pilot the European Language Grid in specific sectors of high commercial and/or societal impact, through small scale demonstrators geared towards an innovative integration of language technologies in specific operating processes/operations. The action shall provide facilities for collaboration, technical and linguistic guidance, access to open-source tools and open language resources (available through the grid), access to venture capital, and promotion and dissemination events. The results of all small scale demonstrators should be made available through the European Language grid under appropriate licensing conditions. The action shall select these small scale demonstrators through the use of financial support to third parties. Up to 30% of the EU funding of the action should be allocated to the financial support of these third parties, typically of the size of EUR 100 000 to 200 000 per third party^[1] and a duration of about 9 to 12 months. Financial support to third parties should in line with the conditions set out in Part K of the General Annexes.

iv. establish competence centres / nodes in Member and Associated States. It shall build on the previous EC-funded actions within the FP7, H2020 and CEF^[2].

The Commission considers that proposals requesting a contribution from the EU of about 7 million would allow this area to be addressed appropriately. Nonetheless, this does not preclude submission and selection of proposals requesting other amounts.



EUROPEAN LANGUAGE GRID

Objectives (Selection)

1. Establish the ELG as the primary platform for LT in Europe.
2. ELG as a platform for commercial and non-commercial, industry-related LTs, both functional and non-functional.
3. Establish the ELG as the primary European market place for the highly fragmented European LT landscape (demand ↔ supply).
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 (pilot projects).
6. Help establish the Multilingual Digital Single Market.



EUROPEAN LANGUAGE GRID

Consortium



Kick-off meeting, 22/23 January 2019

Good relationship with all relevant initiatives

- META-NET and META
- Cracking the Language Barrier
- ELRC, CEF AT, e-translation
- LT Innovate
- CLARIN
- Big Data Value Association
- AI4EU, HumanE AI
- EOSC *and several others*



EUROPEAN LANGUAGE GRID

Concept and Methodology

- **Grid Platform**
- **Grid Content**
- **Grid Community**



EUROPEAN LANGUAGE GRID

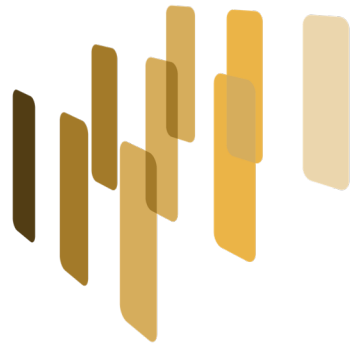
Concept and Methodology

- **Grid Platform**

- 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 directory of
 - Functional services, data sets, tools, technologies, models etc.
 - LT companies, research organisations, research projects, service and application types, languages etc.

- **Grid Content**

- **Grid Community**



EUROPEAN LANGUAGE GRID

Concept and Methodology

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



EUROPEAN LANGUAGE GRID

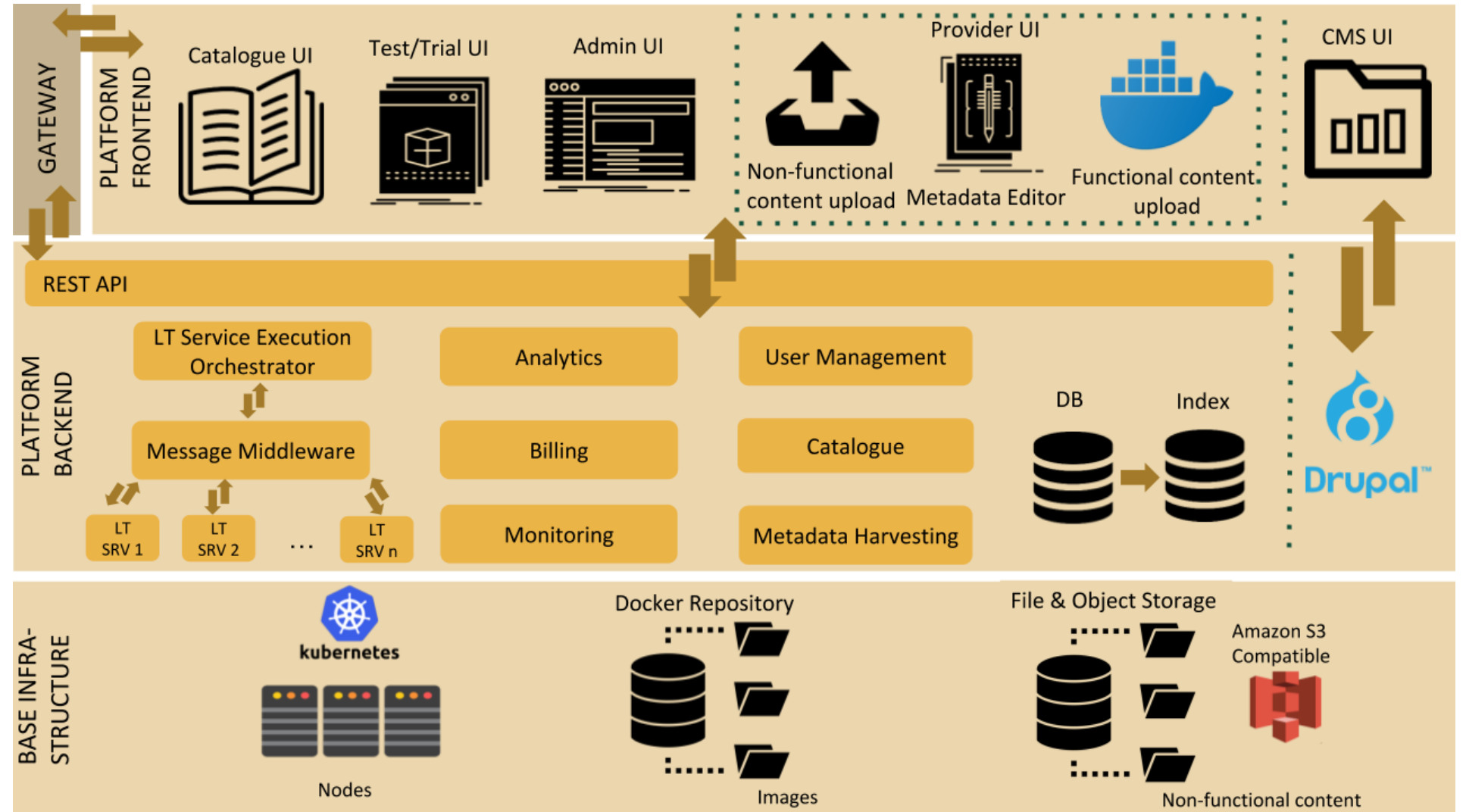
Concept and Methodology

- **Grid Platform**
- **Grid Content**
- **Grid Community**
 - Stakeholders: LT provider and buyer companies, research centres, universities, administrations, NGOs etc.
 - Plus: 15-20 pilot projects, the six ICT-29b) RIA projects, META-NET, LT Innovate and other networks and 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 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

Architecture (Current Working Version)



What's contained in the ELG Platform? Four different Types of Content.

Type	Description	ELG Consortium brings in
1) Functional grid content: LT Tools and Services	Containerised services; can be uploaded and deployed as well as integrated through ELG	GATE, GATECloud, DKT, UDPipe, Tilde's and UEDIN's MT services, SAIL's ASR, KWS, sentiment (polarity) detection, age & gender detection tools, etc.
2) Functional grid content (remotely invoked): LT Tools and Services	Remote APIs can be registered, described, searched and integrated with the help of ELG	Research and commercial services (e.g. GATE Cloud with 65+ NLP services, UDPipe), TILDE's and UEDIN's MT services, SAIL's and EXPSYS's services, etc.
3) Non-functional grid content: LRs, data sets etc.	Upload, describe, search, download data sets, source code, models etc.	Data sets from META-SHARE, ELRC-SHARE, ELRA, resources, models, tools etc.
4) Yellow Pages of European LT	ELG catalogue entries, e.g., of an LT provider company	CRACKER, ELRA, ELRC (NAPs etc.), META-SHARE, LingHub, LT World etc.



EUROPEAN LANGUAGE GRID

Services and Tools

	Language Group			Totals
	A	B	C	
Basic linguistic processing	146	30	63	239
Entity recognition and linking	102	7	11	120
Classification (sentiment, opinion, topic, language etc.)	40	4	18	62
Machine Translation	105	4	14	123
Other text processing (including parsing, summarisation etc.)	35	7	13	55
Speech tools (ASR, gender detection, synthesis etc.)	26	5	14	45
Totals	454	57	133	644

Services and tools to be ingested by the ELG partners

A – Official EU languages: Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovenian, Spanish, and Swedish

B – EU candidate languages and Free Trade Partners: Albanian, Basque, Catalan, Galician, Icelandic, Norwegian, Scottish Gaelic, Welsh, Serbian, Turkish, Ukrainian

C – Languages spoken by EU immigrants; languages of political and trade partners: Afrikaans, Arabic, Berber, Cebuano, Chinese, Hebrew, Hindi/Urdu, Indonesian, Japanese, Korean, Kurdish, Latin, Malay, Pashto, Persian (Farsi), Russian, Tamil, Vietnamese



EUROPEAN LANGUAGE GRID

	Open Access	Language Group A			Language Group B			Language Group C			Totals
		Corpora	Lexicons	Models	Corpora	Lexicons	Models	Corpora	Lexicons	Models	
META-SHARE	yes	617	447	16	55	54	0	84	51	1	1325
	no	582	550	1	44	65	0	198	94	0	1534
ELRC-SHARE	yes	317	114	0	3	1	0	0	0	0	435
	no	74	16	0	2	1	0	0	0	0	93
ELDA	no	563	1012	0	35	18	0	250	54	0	1932
ELG	mixed	74	108	43	0	0	12	4	1	21	263
Totals		2227	2247	60	139	139	12	536	200	22	5582

Datasets and resources to be provided by the ELG partners

A – Official EU languages: Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovenian, Spanish, and Swedish

B – EU candidate languages and Free Trade Partners: Albanian, Basque, Catalan, Galician, Icelandic, Norwegian, Scottish Gaelic, Welsh, Serbian, Turkish, Ukrainian

C – Languages spoken by EU immigrants; languages of political and trade partners: Afrikaans, Arabic, Berber, Cebuano, Chinese, Hebrew, Hindi/Urdu, Indonesian, Japanese, Korean, Kurdish, Latin, Malay, Pashto, Persian (Farsi), Russian, Tamil, Vietnamese



EUROPEAN LANGUAGE GRID

Pilot Projects

- ELG will provide 1.9M€ for pilot projects using Financial Support for Third Parties, selected through two open calls.
- Usage of LTs in specific applications, processes and operations.
- Pilot projects should demonstrate the usefulness of ELG.
- Results of the projects will be made available through ELG.
- Lightweight, open, transparent and expert-evaluation based selection process.
- Single project up to 200,000€, recommended minimal amount of 100,000€; duration expected in the 9-12 months range.

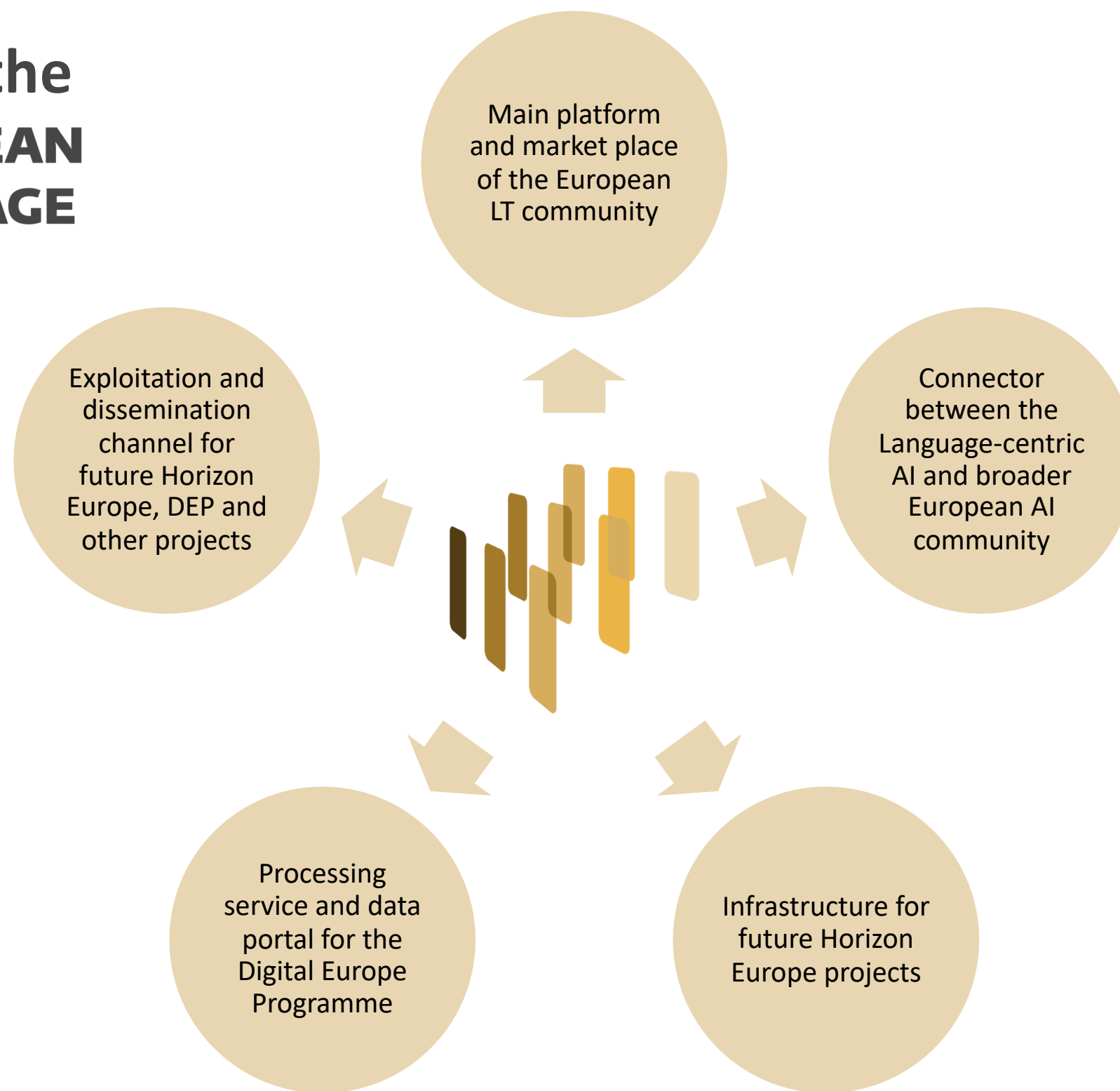


EUROPEAN LANGUAGE GRID

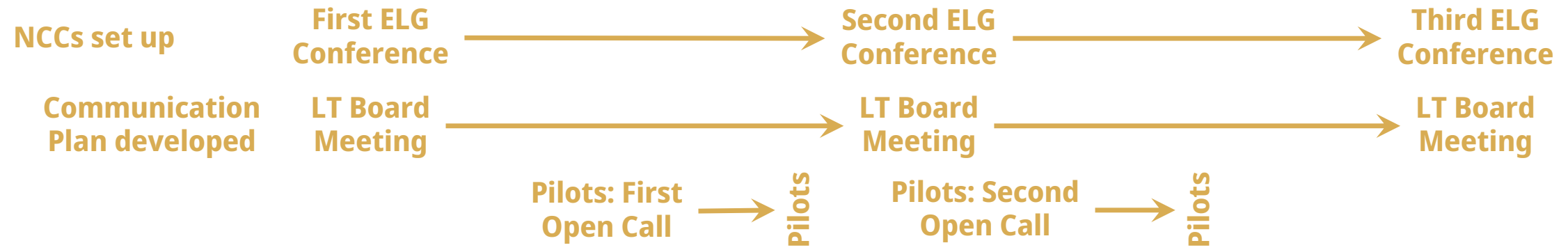
National Competence Centres (NCCs)

- Network of 32 National Competence Centres (NCC) for outreach, communication and promotion activities.
- All academic organisations – to guarantee independence.
- Act as local bridges to the ELG consortium and platform.
- Selected based on participation in existing initiatives, taking into account scientific standing and connections to industry.
- Responsibilities: collect national/regional information about services, data sets, tools, technologies, projects, research centres, companies, etc.; organise a training workshop; promote ELG in their region; help with outreach and dissemination activities.
- NCCs are the backbone of the European LT Council.

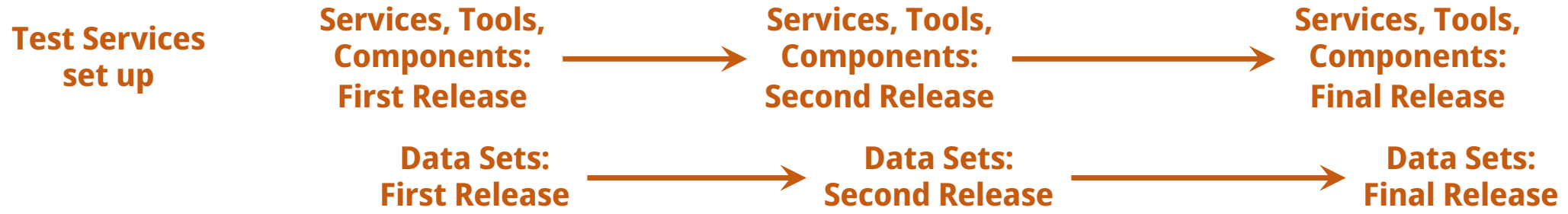
Future roles of the **EUROPEAN LANGUAGE GRID**



Grid Community



Grid Content



Grid Platform



Mgmt.





EUROPEAN LANGUAGE GRID

Summary

- Establish ELG as the primary platform for (industry-relevant) LT in Europe, bringing together and uniting a network of European experts.
- Market place for the European LT business space, ELG will strengthen Europe's position in this field.
- There is currently no other scalable cloud platform that serves as a broker for such a broad variety of multilingual services, resources and data sets.
- ELG is a project and initiative *from* the European LT community *for* the European LT community.



European Language Grid An Overview

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, ELG Coordinator)
georg.rehm@dfki.de

08-10-2019, META-FORUM 2019 – Brussels, Belgium
<http://www.european-language-grid.eu>