# What's next for Multilingual Europe? Introducing the European Language Grid

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

09-09-2019, ELSE-IF 2019 – Karlsruhe, Germany http://www.european-language-grid.eu

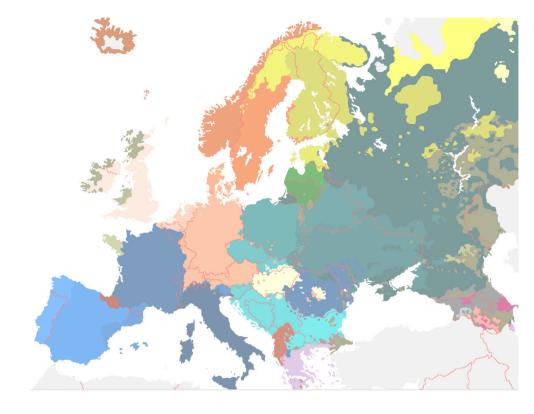


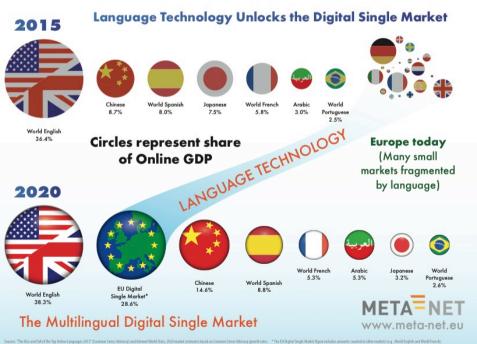
What's next for Multilingual Europe? Introducing the European Language Grid 09-09-2019, ELSE-IF 2019 – Karlsruhe, Germany

- 1. History Motivation Context
- 2. European Language Grid (2019-2021)

# **Multilingualism in Europe**

- Multilingualism is at the heart of the European idea
- 24 EU languages they all have the same status
- Dozens of regional and minority languages as well as languages of immigrants and trade partners
- Many economic, social and technical challenges:
  - The Digital Single Market needs to be multilingual
  - Cross-border, cross-lingual, cross-cultural communication
  - Fragmentation of the LT market and landscape (including availability of LTs and LRs)
  - There's no LT platform for Europe! (yet)







# 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**
- 900+ members in 67 countries



Published in 2013

# **META-NET (T4ME)**

**CESAR** 

# **META-NORD**

# **METANET4U**





# META FORUM 2010

## T<sub>4</sub>ME



**H**ELG

### T4ME

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

### META-NET

Established in 2010, META-NET is a network of Excellence consisting of 60 research centres from 34 countries, building the technological foundations of a multilingual European information society

### META-FORUM 2010

2010

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



### META-NORD

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

2011

2012

### METANET4U

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



#### CESAR EU-funded project (ICT Policy Support Programme) functioning as a part of META-NET to standardise language resources and tools (2011 – 2013)

### META-FORUM 2011

Conference "Solutions for Multilingual Europe" (Budapest – June 27/28, 2011)







#### META-FORUM 2012

Conference "A Strategy for Multilingual Europe" (Brussels – June 20/21, 2012)

#### META-NET White Papers

Release of 32 volumes on 31 languages, revealing that there is a severe threat of digital extinction for at least 21 European languages 2013 (December 2012)

#### "State of the Art of Machine Translation -Current Challenges and Future Opportunities 2014 Workshop on "State of the Art of Machine

Translation", commissioned by the EU Parliament (December 2013)

#### META-SHARE

Initiated in 2013, META-SHARE has functioned as an open and secure network of repositories for sharing and exchanging language data, tools and services

#### Strategic Research Agenda for Multilingual Europe 2020

Launch of the Strategic Research Agenda for Multilingual Europe 2020 (January 2013)

#### META-FORUM 2013

Conference "Connecting Europe for New Horizons" (Berlin - September 19/20, 2013)





### CRACKER

EU-funded project CRACKER (Horizon2020) pushing towards an improvement of MT research in terms of efficiency and effectiveness (2015 - 2017)

#### **Cracking the Language Barrier Federation**

Founded in 2015, the federation has been assembling European research and innovation projects as well as all related community organisations working on multilingual technologies

#### META-FORUM 2015

Conference "Technologies for the Multilingual Digital Single Market" (Riga – April 27, 2015)

#### Strategic Agenda for the Multilingual Digital Single Market (Version 0.5)

Launch of the Strategic Agenda for the Multilingual Digital Single Market titled "Technologies for Overcoming Language Barriers towards a truly integrated European Online Market" (April 2015)

### **Riga Summit on the Multilingual**

**Digital Single Market** Summit "Shape the future of the multilingual

digital single market" (April 27–29,2015)

#### Strategic Research and Innovation Agenda (Version 0.9)

Launch of the Strategic Research and Innovation Agenda titled "Language as a Data Type and Key Challenge for Big Data" (July 2016)

#### META-FORUM 2016

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





## META FORUM 2016





#### "Language Equality in the Digital Age"

Workshop on "Language Equality in the Digital Age", commissioned by the EU Parliament's Science and Technology Options Assessment Committee (STOA) (January 2017)

#### "Language equality in the digital age: Towards a Human Language Project"

Launch of the study on "Language equality in the digital age: Towards a Human Language Project", commissioned by the EU Parliament (March 2017)

#### META-FORUM 2017

2016

Conference "Towards a Human Language Project" (Brussels – November 13/14, 2017)

#### Strategic Research and Innovation Agenda (V1.0)

Launch of the Strategic Research and Innovation Agenda titled "Language Technologies for Multilingual Europe – Towards a Human 2017 2015 Language Project" (December 2017)



**ELG Final Proposal Submission** 

Final submission on Feb. 20. 2018



EU-funded (ICT PSP)



















# 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

http://www.meta-net.eu/whitepapers

**#ELG** 09-09-2019, ELSE-IF 2019 – Karlsruhe, Germany

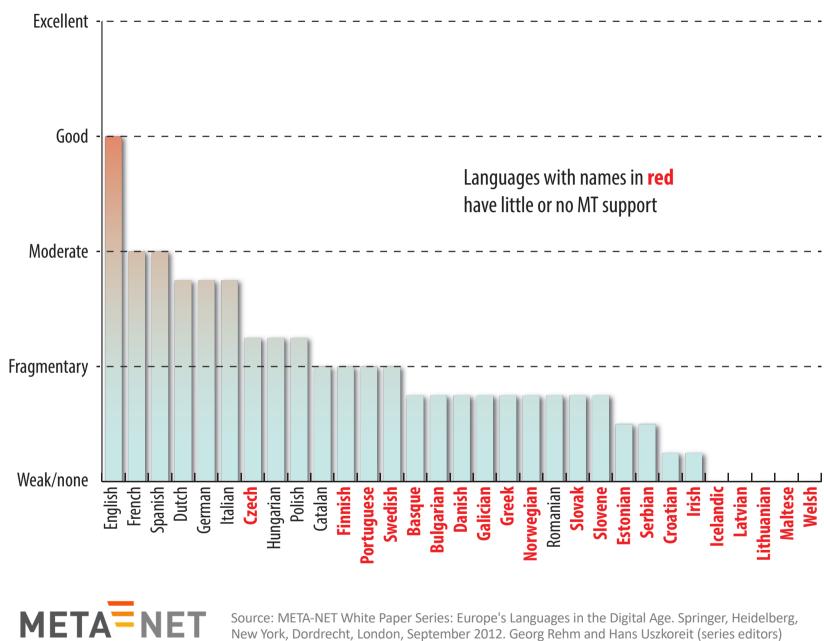
Polish\* Portuguese\* Romanian\* Serbian Slovak\* Slovene\* Spanish\* Swedish\* Welsh

\* Official EU language

		Excellent support	Good support	Moderate support	Fragmentary support
The Normedian involves in the Caractan involves in the Service Language in Service	Machine Translation		English	French, Spanish	Catalan, Dutch, German Hungarian, Italian, Polish Romanian
THE DIGITAL AGE – SVENSK DERVENDAR HEAVEN GUTAN CALL - HEAVENDAR HEAVENDAR GUTAN CALL - HEAVENDAR GUTAL AGE – SVENSK JEZIK V DIGITAL GUTAL AGE – CENCOKI JEZIK V DIGITAL GUTAL AGE – LEVOVENSK TUROVA MAUTUA FLERA VI AGE – LEUNGVA NUELVENNME AVZ DIGITAL AGE – LA LINGVA MAUTUA FLERA VI AGE – LA LINGVA TALLANA NELVE E DIGITAL AGE – DIE DENUSCHE SPROG I DI AL AGE – DIE DENUSCHE SPROG I DI AL AGE – CEŠTINA V DIGITALINIM VĚKU DIGITAL AGE – LEVOMEN KIELI DIGITALAU E DIGITAL AGE – LEVOMEN KIELI DIGITALAU E DIGITAL AGE – LET NEDERLANDS IN HEI DIGITAL AGE – LET NEDERLANDS IN HEI DIGITAL AGE – LEVOMEN KIELI DIGITALAU E DIGITAL AGE – LEVOMEN KIELI DIGITALAU E DIGITAL AGE – LEVOMEN KIELI DIGITALAU CHAGENDA FOR MULTUINGUAL EURO CHAGENDA FOR MULTUNGUAL EURO CHAGENDA FOR M	Text Analytics		English	Dutch, French, German, Italian, Spanish	Basque, Bulgarian, Catala Czech, Danish, Finnish, Galician, Greek, Hungaria Norwegian, Polish, Portuguese, Romanian, Slovak, Slovene, Swedish
BEN BITALA TIDSALDERN BITALA TIDSALDERN BITALA TIDSALDEREN FOIGITALE E INDSALDEREN DIGITALE FUDSALDEREN DIGITALE TIDSALDER E IM DIGITALE TIDSALDER SELLA AIKAKAUDELLA L FOIGITALE TIDSALDER FOIGITALE TIDSALDER A L ' ERE DU NUMÉRIQUE SELLA AIKAKAUDELLA L FOIGITALE TIDSALDER A L ' ERE DU NUMÉRIQUE SELLA AIKAKAUDELLA L FOIGITALE TIDSALDER EAN PE 2020	Speech		English	Czech, Dutch, Finnish, French, German, Italian, Portuguese, Spanish	Basque, Bulgarian, Catala Danish, Estonian, Galician Greek, Hungarian, Irish, Norwegian, Polish, Serbia Slovak, Slovene, Swedish
	Resources		English	Czech, Dutch, French, German, Hungarian, Italian, Polish, Spanish, Swedish	Basque, Bulgarian, Catala Croatian, Danish, Estonia Finnish, Galician, Greek Norwegian, Portuguese Romanian, Serbian, Slova Slovene
	ų	ELG	09-09-2019,	ELSE-IF 2019 – Karlsruhe, (	Germany

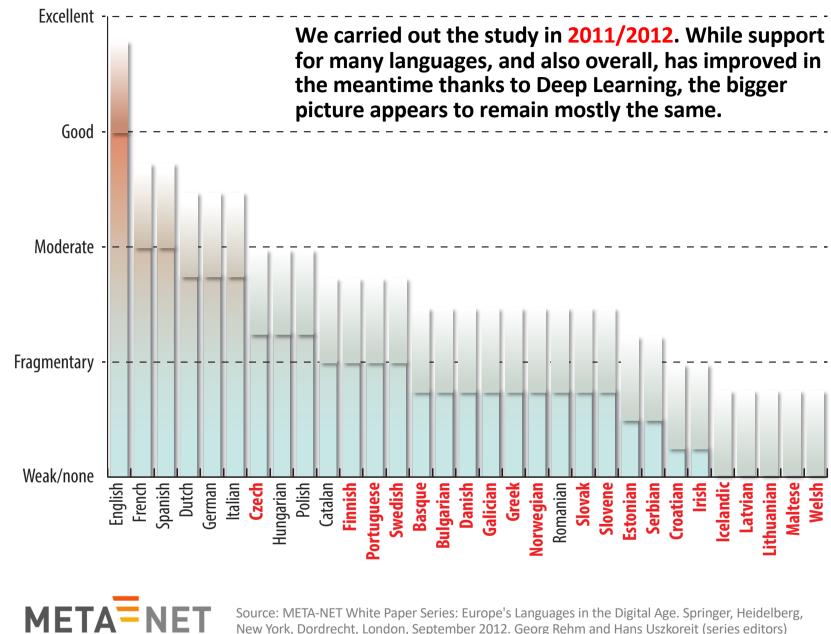
port	Weak or no support through LT
rman, <sup>2</sup> olish,	Basque, Bulgarian, Croatian, Czech, Danish, Estonian, Finnish, Galician, Greek, Icelandic, Irish, Latvian, Lithuanian, Maltese, Norwegian, Portuguese, Serbian, Slovak, Slovene, Swedish, Welsh
atalan, nish, garian, sh, nian, vedish	Croatian, Estonian, Icelandic, Irish, Latvian, Lithuanian, Maltese, Serbian, Welsh
atalan, alician, Irish, erbian, vedish	Croatian, Icelandic, Latvian, Lithuanian, Maltese, Romanian, Welsh
atalan, tonian, ireek, uese, Slovak,	Icelandic, Irish, Latvian, Lithuanian, Maltese, Welsh





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)



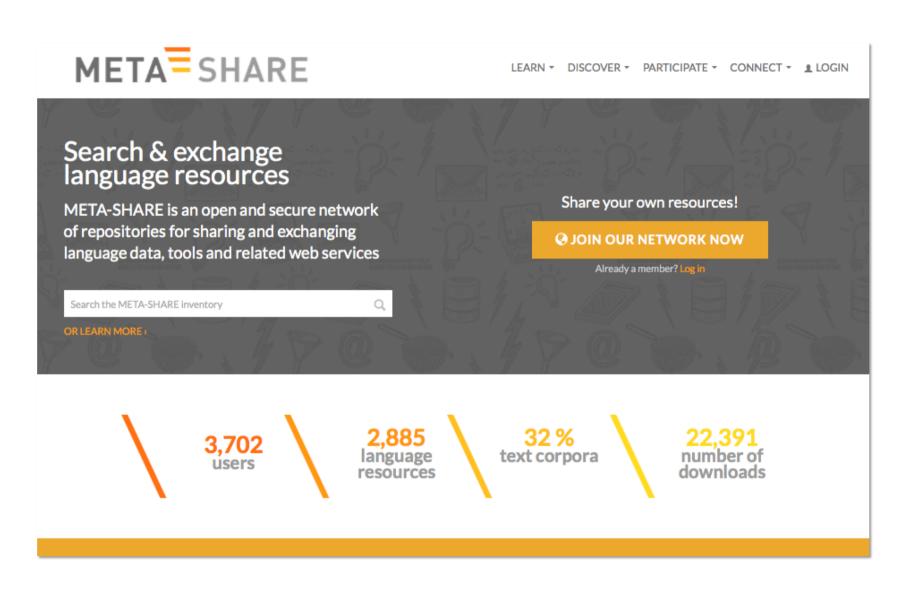


New York, Dordrecht, London, September 2012. Georg Rehm and Hans Uszkoreit (series editors)

9

# **META-SHARE**

- META-SHARE addresses technology and infrastructure needs with regard to the
  - Visibility
  - Documentation
  - Identification
  - Availability
  - Preservation
  - Interoperability of LRs and LTs
- 35 META-SHARE members and organisations in 25 countries
- http://www.meta-share.org



# **Motivation and Context**

- The European LT landscape and market are very fragmented.
- 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 mentioned in the Strategic Research **Agenda for Multilingual Europe 2020** (published in early 2013).



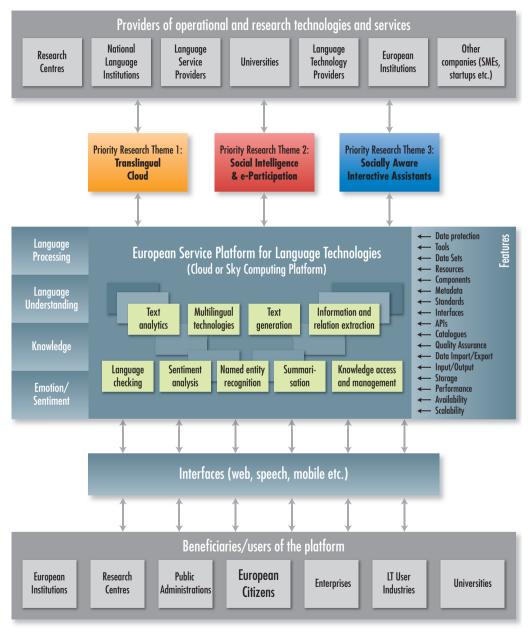
The concept was later refined in the three LT SRIAs (2015, 2016, 2017).



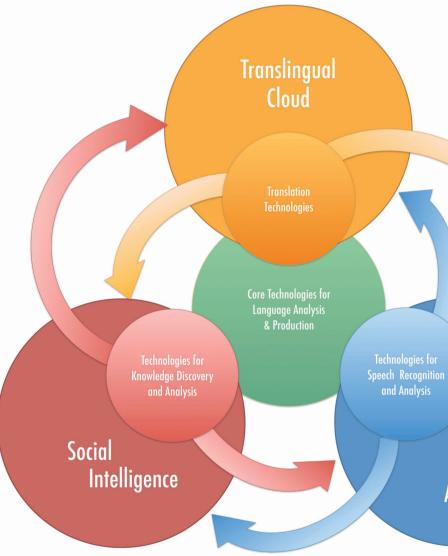


### META FORUM 2017

# **European LT Service Platform**



### 3+2 Priority Research Themes



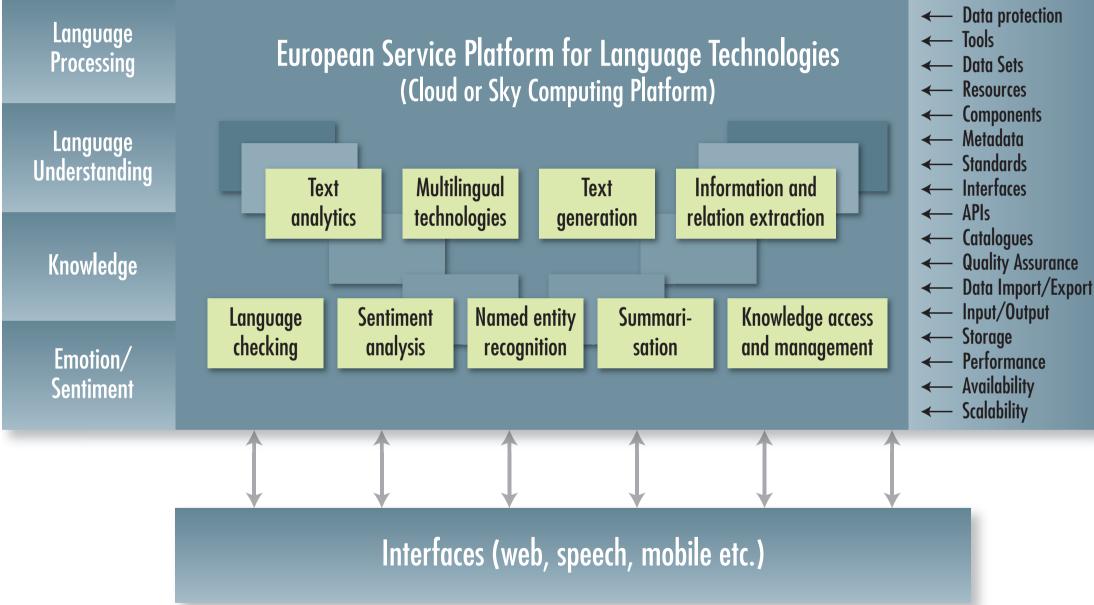
Socially Aware Interactive Assistants

### From: META-NET SRA (2013)



12

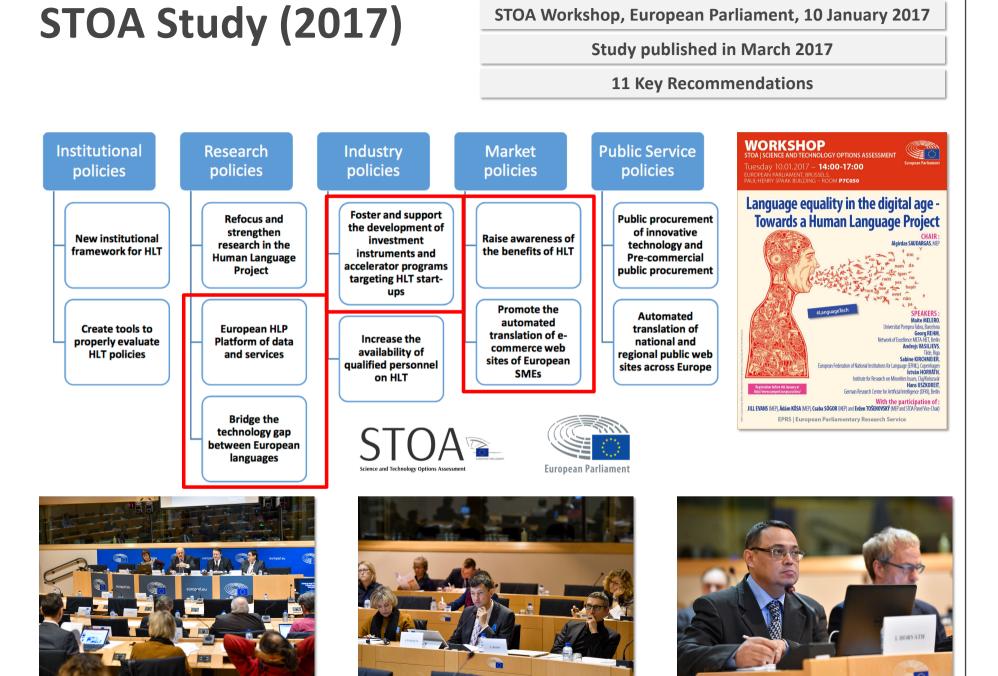
# **European LT Service Platform**



# Features

### From: MFTA-NFT SRA (2013)







# Language equality in the digital age

Towards a Human Language Project

STUDY

Science and Technology Options Assessment

EPRS | European Parliamentary Research Service Scientific Foresight Unit (STOA) PE 581.621

# "Language Equality" EP Resolution (2018)

EP Resolution "Language equality in the digital age" P8 TA(2018)0332 – partially based on the STOA study

Voting (11 Sept. 2018): **592 yes** – 45 no

**Recommendations addressed by European Language Grid:** 

- 29. Create a European LT platform for sharing of services
- 41. Enable and empower European SMEs to use LTs
- 26. ICT integrators should be given economic incentives for LT
- 27. Europe has to secure its leadership in language-centric AI
- 32. Set up LT financing platform; emphasise R&D in Deep NLU
- 40. Develop investment instruments and accelerator programs

European Parliament 2014-2019

P8 TA-PROV(2018)0332

Language equality in the digital age European Parliament resolution of 11 September 2018 on language equality in the digital age (2018/2028(INI))

The European Parliament.

- having regard to Articles 2 and 3(3) of the Treaty on the Functioning of the European Union (TFEU),
- having regard to Articles 21(1) and 22 of the Charter of Fundamental Rights of the European Union.
- having regard to the 2003 UNESCO Convention for the Safeguarding of the Intangible Cultural Heritage.
- having regard to Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information<sup>1</sup>,
- having regard to Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information<sup>2</sup>,
- having regard to Decision (EU) 2015/2240 of the European Parliament and of the Council of 25 November 2015 establishing a programme on interoperability solutions and common frameworks for European public administrations, businesses and citizens (ISA2 programme) as a means for modernising the public sector<sup>3</sup>,
- having regard to the Council resolution of 21 November 2008 on a European strategy for multilingualism (2008/C 320/01)<sup>4</sup>.
- having regard to the Council decision of 3 December 2013 establishing the specific programme implementing Horizon 2020 - the Framework Programme for Research and

OJ L 345, 31.12.2003, p. 90. OJ L 175, 27.6.2013, p. 1.

OJ L 318, 4.12.2015, p. 1.

OJ C 320, 16.12.2008, p. 1.



#### TEXTS ADOPTED Provisional edition

# **Objectives**

- 1. Establish the ELG as the primary platform for LT in Europe.
- 2. ELG as a platform for commercial and non-commercial LTs, both functional and non-functional.

A similar platform does not exist yet – not even globally. The ELG platform will be the very first of its kind.

- 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. Unleash enormous potential for innovation (pilots etc.)
- 5. Establish the European Language Grid as the primary European market place for LT to connect demand and supply.
- 6. Help establish the Multilingual Digital Single Market.

# Consortium











# **All relevant Initiatives covered**

- META-NET and META
- Cracking the Language Barrier
- ELRC, CEF AT Tools & Services
- CLARIN

DA

- LT Innovate
- Big Data Value Association
- AI4EU, HumanE AI
- EOSC and several others



# **EUROPEAN** Concept and Methodology **LANGUAGE GRID** • Grid Platform

- Grid Content
- Grid Community

# **EUROPEAN** Concept and Methodology LANGUAGE **Grid Platform** GRID

- Catalogue of functional services, data sets, tools, technologies, models etc.
- Catalogue of LT companies, research organisations, service and application types, languages etc.
- Interactive modern web interface with different views for users, LT providers, potential buyers, and administrators.
- Information about conference and training events, pilot projects, open calls and other pieces of content.
- Built with robust, scalable, reliable, widely used technologies that are constantly developed further.
- Ability to scale with the growing demand and supply of resources.
- **Grid Content**
- Grid Community

# **EUROPEAN** Concept and Methodology LANGUAGE **Grid Platform** GRID

# **Grid Content**

- **Content:** services, resources, data sets, tools.
- **Functional Content:** running services, can be uploaded and integrated into other systems. Stakeholders will be able to upload and make available their services.
- LT-as-a-Service: create an image of an LT service that can be run on a different system through a virtual machine.
- Easy and efficient for LT providers to create and upload images of LT services, easy and efficient for ELG to receive and make images available.
- **Non-functional Content:** catalogue entries of companies; file storage; description of a web interface (REST API).

# Grid Community

# **EUROPEAN** Concept and Methodology LANGUAGE **Grid Platform** GRID

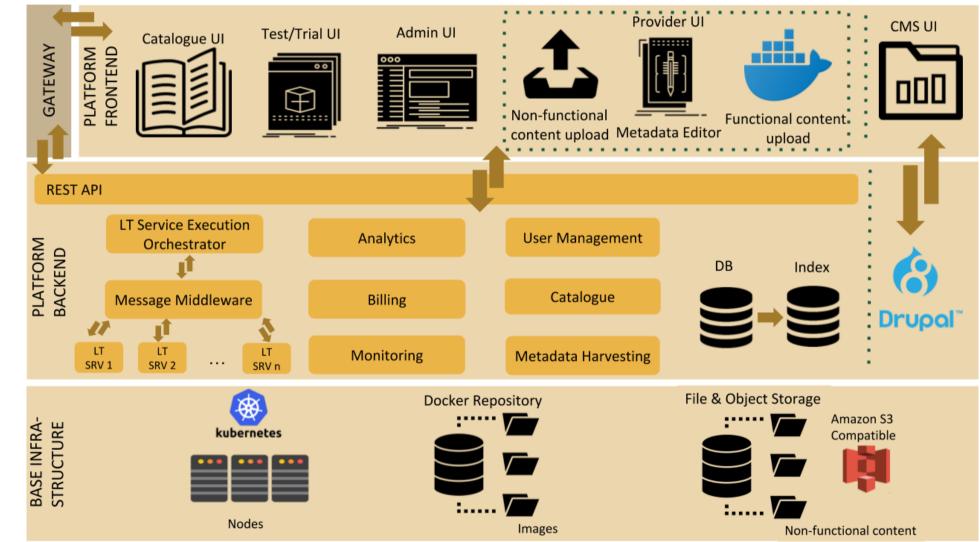
**Grid Content** 

# Grid Community

- Stakeholders: LT provider and buyer companies, research centres, universities, administrations, NGOs etc.
- Plus: 15-20 pilot projects, the six new ICT-29b) RIA projects, META-NET, IT Innovate and other networks and initiatives
- Organisation of conferences, training events, presentations, publications, social media, blog posts etc.
- 32 NCCs as a strong international network to identify content for the ELG, interest companies in using the ELG for making available their services.
- European LT Council as an international, pan-European body, in which LTrelated matters can be discussed and coordinated.

# Architecture

(Current Working Version)



# EUROPEAN<br/>LANGUAGEGRIDPlatform Catalogue – Four different layers

	Туре	Description	ELG Consortium bri
1)	ELG Language Technology Tools, Services – functional Grid content	Containerised services (from, e.g., tokenisation or POS tagging to complex workflows); can be uploaded and deployed through the ELG; can be integrated in other systems through the ELG	GATE, GATECloud, DKT, UDPipe UEDIN's MT services, Weblicht, sentiment (polarity) detection, a detection tools, etc.
2)	ELG Language Technology Tools, Services – functional Grid content, remotely invoked	Remote APIs (REST) can be registered, described, searched and integrated with the help of the ELG platform	Research and commercial servic Cloud with 65+ NLP services, UE and UEDIN's MT services, SAIL's services, etc.
3)	ELG LRs/LTs – non-functional Grid content	Upload, describe, search, download data sets, source code, models etc.	Data sets from META-SHARE, EL resources, models, tools etc.
4)	ELG Information and metadata	ELG catalogue entries, e.g., of an LT provider company (no language processing functionality or code available in ELG, no data)	CRACKER, ELRA, ELRC (NAPs etc CLARIN VLO, LingHub, LT World

### ings in

e, Tilde's and , SAIL's ASR, KWS, age & gender

ices (e.g. GATE DPipe), TILDE's s and EXPSYS's

### ELRC-SHARE, ELRA,

### c.), META-SHARE, d etc.

# **Pilot Projects**

- ELG will provide close to 30% of the overall budget for small scale demonstrator projects (pilots) using Financial Support for Third Parties, selected among SME applicants through two open calls.
- Usage of LTs in specific applications, processes and operations; pilots should also demonstrate the usefulness of the ELG as being built within the project.
- Results will be made available through the ELG.
- Lightweight, open, transparent and expert-evaluation based selection process, following established successful practices of previous EC funded projects (e.g., ODINE).
- Single project up to 200,000€, with recommended minimal amount of 100,000€; duration of the projects is expected in the 9-12 months range.

# **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, resources, tools, technologies, projects, research centres, experts, communities, companies, etc.; organise an event such as, e.g., a training workshop; promote ELG in their region; help with outreach and dissemination activities.
- NCCs as the backbone of the European LT Council.

Impact

- Impact on Business and Industry
  - Tackle fragmentation
  - Enable European SMEs to use LTs to expand their business online across many languages
  - Open up markets and foster growth •
  - Connect demand and supply •
  - Cost reductions
  - Reduced time-to-market
- **Impact on Innovation** 
  - Provide HQ resources for fast and efficient experimenting with new methods and technologies
  - Enable businesses to grow and benefit from scaling up.
- Impact on the Digital Single Market
  - Unlock the DSM by making it multilingual and socially inclusive.
  - MT of e-commerce shops will enable SMEs to offer their services in more languages, take advantage of the DSM
- **Impact on Society and Multilingual Europe** 
  - Prevent digital language extinction •
  - Enable cross-lingual and cross-cultural communication







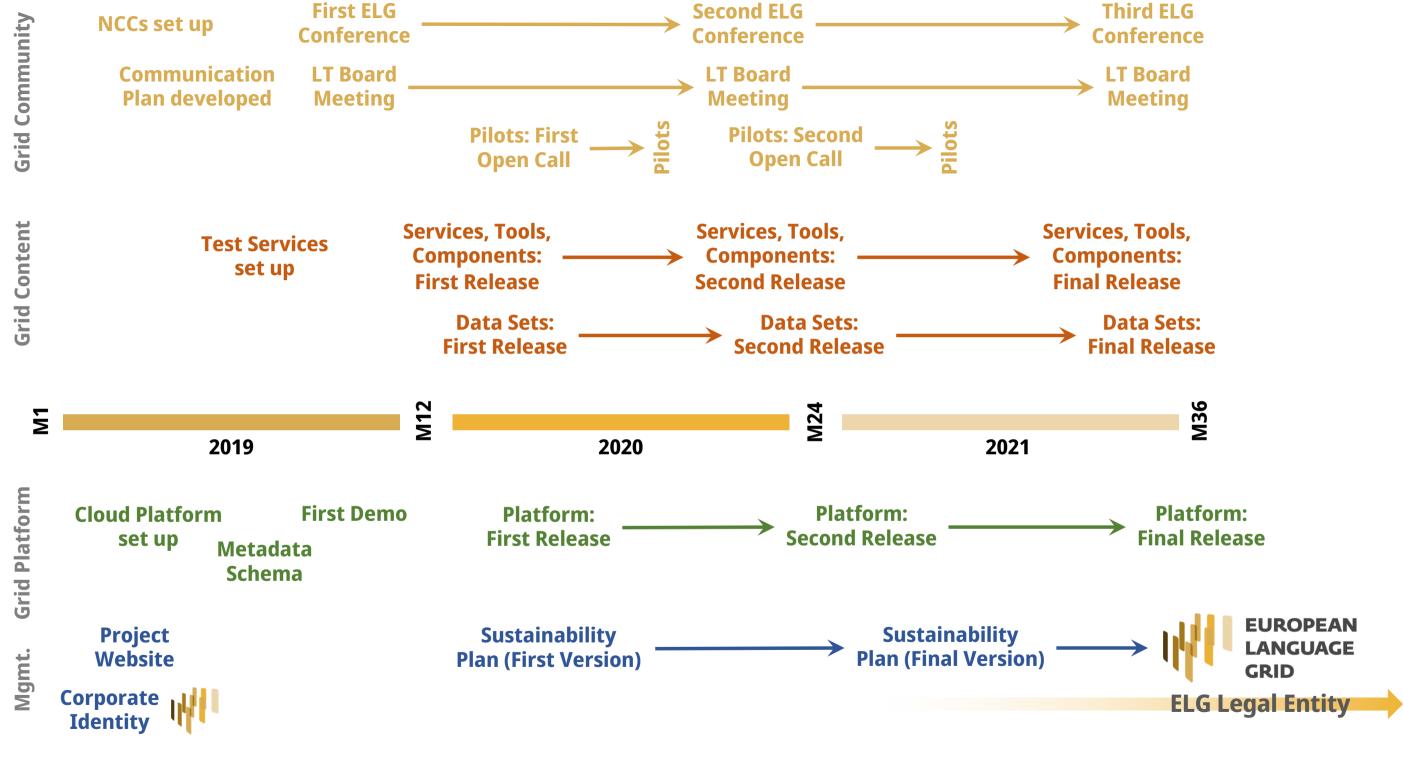
Global Under Resourced MEdia Translation















**Home** 

Ċ

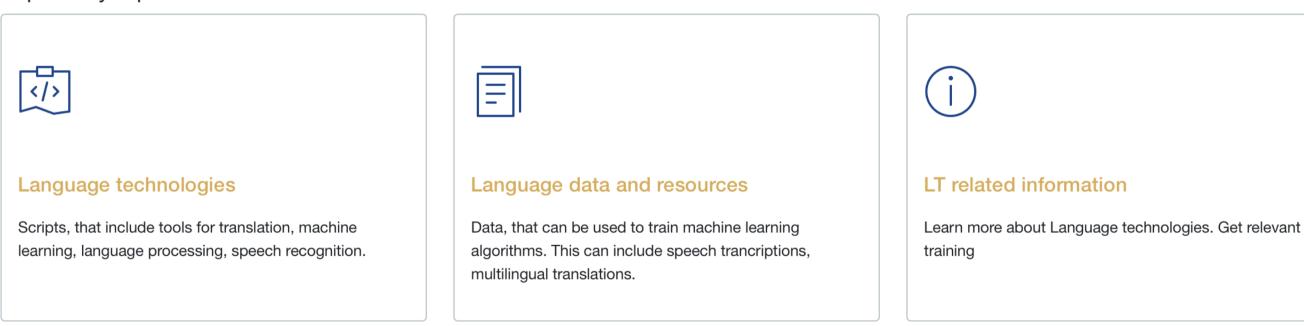
# European Language Grid

All European language technologies, data and information in one place.

=



### Explore by topic



### What are language technologies?

Language technology, often called human language technology (HLT), studies methods of how computer programs or electronic devices can analyze, produce, modify or respond to human texts and appear it consists of natural language processing (NILD) and computational linguistics (CL) on the one hand, and appear to change at the other. It also includes many application oriented

#### Information About ELG

0

ᠿ

 $\Box$ 

				Ċ	
EUROPEAN LANGUAGE GRID				Home Technologies	Resources
Home > Search					
Q Enter search criteria			Search		
All	Technologies	Resources	Informationa		
Categories Licence Resource type	10	0 results ANNIE English Nan	ned Entity Recogniser		Туре
		Named entity recognitio		entity types, such as Person, Location, sions	
		Cogito Discover Pa	rt of Speech Annotator		Туре

Part of Speech Annotator based on Cogito Discover. Cogito Discover is an Expert System scalable production-ready multilingual platform for text analytics. It can be integrated in any application and workflow, deployable as on-premises solution, as SaaS on a cloud platform, or on hybrid environments.

### **Cogito Discover Semantic Annotator**

Semantic Annotator based on Cogito Discover. Cogito Discover is an Expert System scalable production-ready multilingual platform for text analytics. It can be integrated in any application and workflow, deployable as on-premises solution, as SaaS on a cloud platform, or on hybrid environments.

### Information About ELG

0

Û

ŋ

				Ċ	
EUROPEAN LANGUAGE GRID			Home	Technologies	Resources
Home > Search					
Q Enter search criteria		Search			
Back to search results					
ANNIE English Na	med Ent	titv Recoanise	er		
Metadata Documentation	Try out	Code samples			
GATE NER English					
Named entity recognition pipeline that identifies	basic entity types, suc	h as Person, Location, Organizatior	n, Money amou	unts, Time and Date	expressions
Creators					
Name: Ian Roberts					
Metadata creation date: 2019-07-29					

### Information About ELG

• • • +

k Ho





					Ċ	
EUROPEAN LANGUAGE GRID				Home	Technologies	Resources
Home > Search						
Q Enter search criteria		Search				
Back to search results						
ANNIE English Na	med Ent	tity Reco	ognise	r		
Metadata Documentation	Try out	Code	samples			
GATE NER English						
Named entity recognition pipeline that identifies b	asic entity types, suc	h as Person, Locatio	n, Organization, I	Money amou	nts, Time and Date	expressions
Creators						
Name: Ian Roberts						
Metadata creation date: 2019-07-29						

# Information About ELG

∶k He



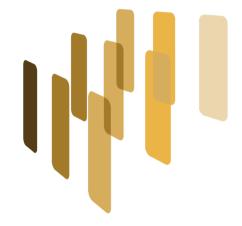


Summary

- The ELG will be the primary platform for LT in Europe, uniting a network of European experts in the 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.

# • *From* the European LT community *for* the community

- New resources and services will be added continuously.
- As a market place for the European LT business space, ELG will strengthen Europe's position in this field.
- SMEs and research centres can apply for pilot projects to validate and evaluate the European Language Grid.



What's next for Multilingual Europe? Introducing the 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, ELG Coordinator) georg.rehm@dfki.de

09-09-2019, ELSE-IF 2019 – Karlsruhe, Germany http://www.european-language-grid.eu