EUROPEAN LANGUAGE GRID

European Language Grid

Raivis Skadiņš (TILDE)
raivis.skadins@tilde.lv

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

05.03.2020.
DIGITAL LATVIAN LANGUAGE RESOURCES AND TOOLS IN COMMON RESEARCH INFRASTRUCTURE
http://www.european-language-grid.eu

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
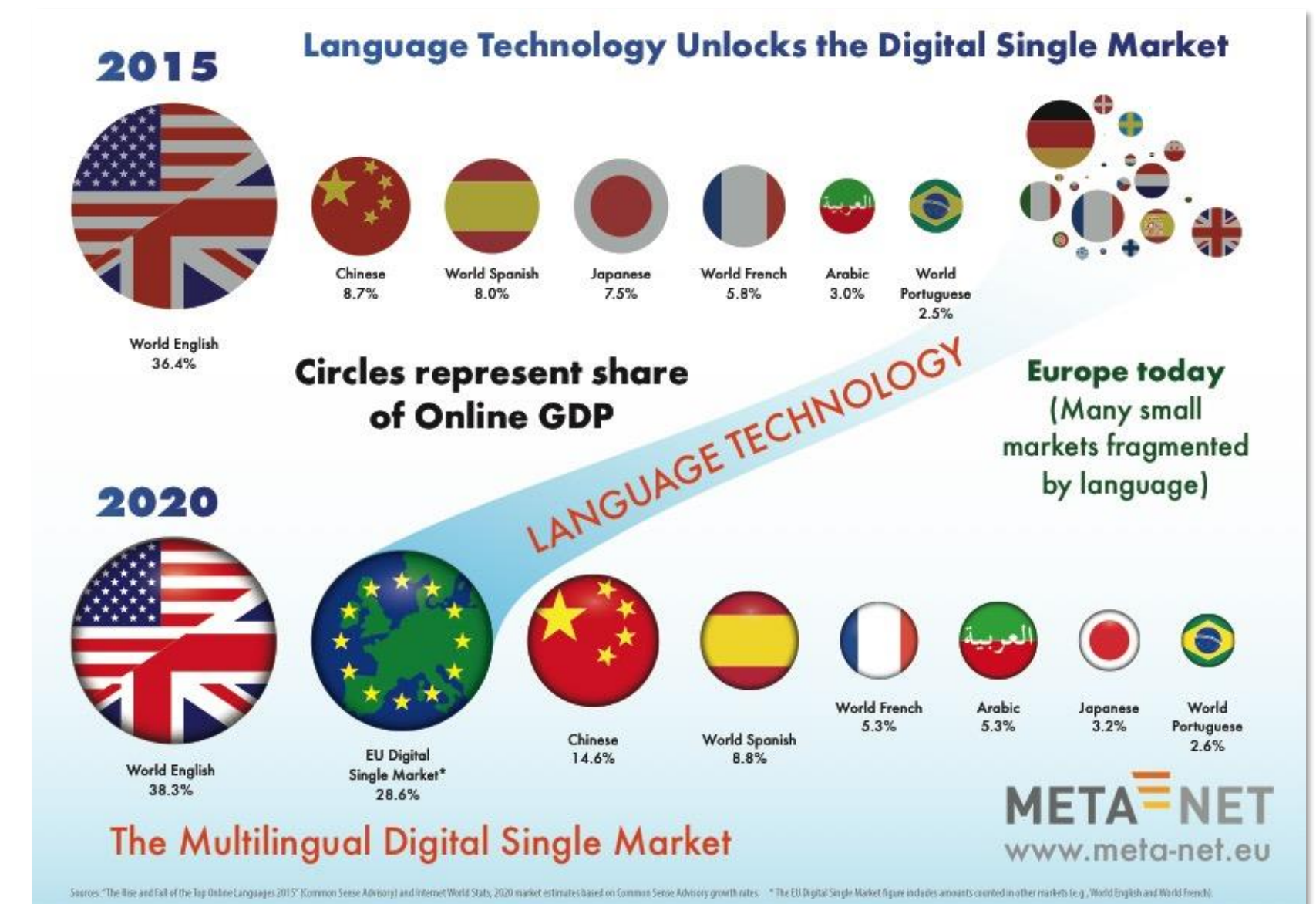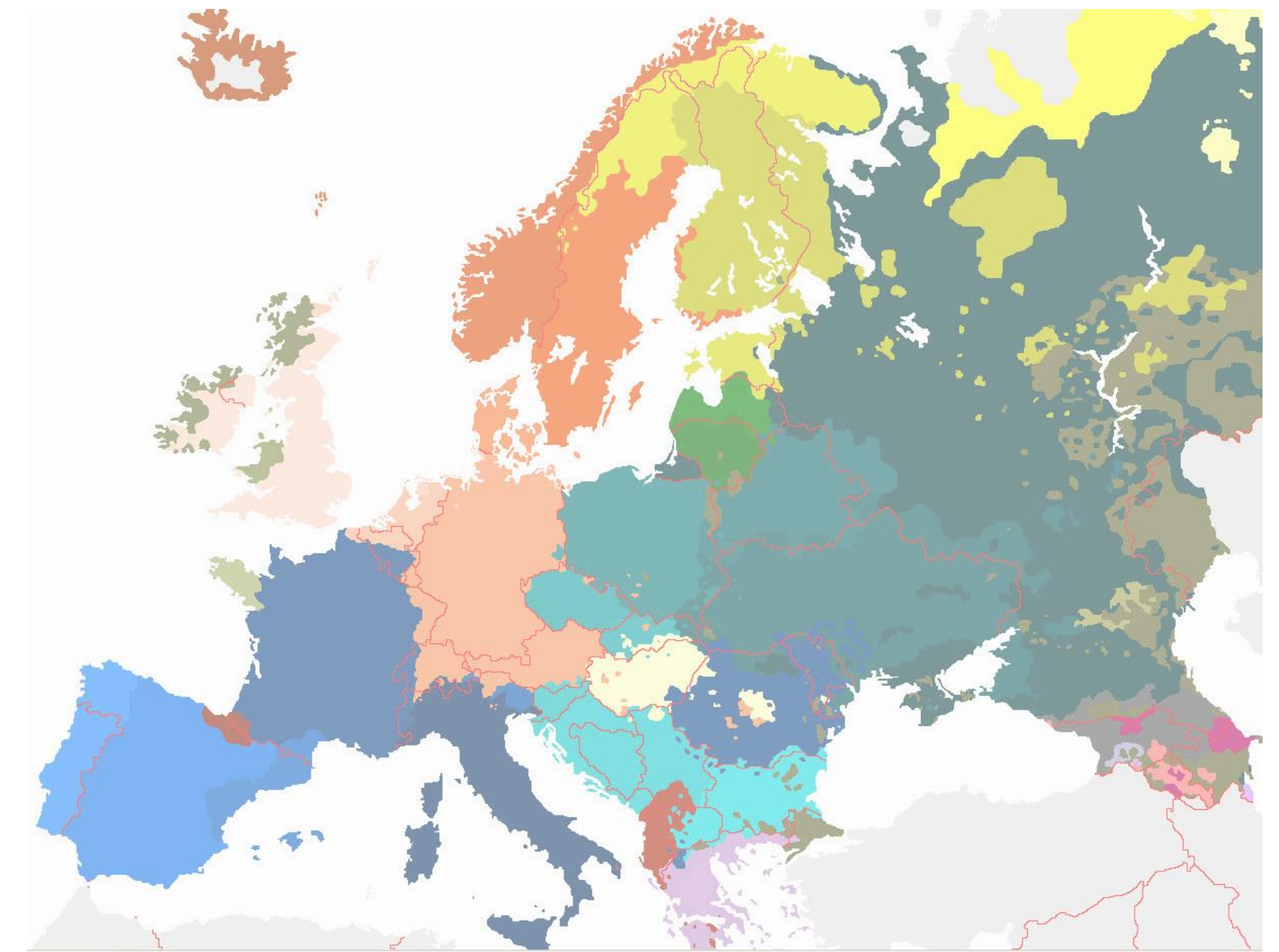
META-NET
(T4ME)

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)
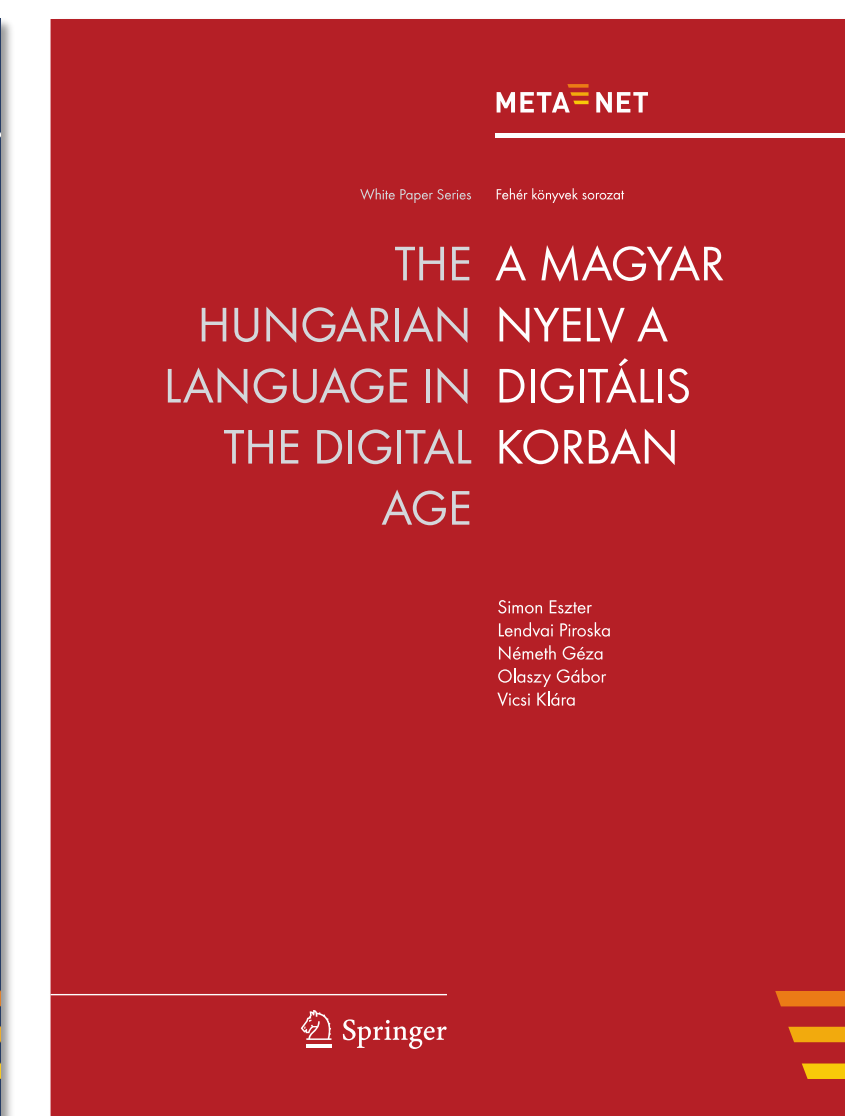
- General Secretary:
  Georg Rehm (DFKI)

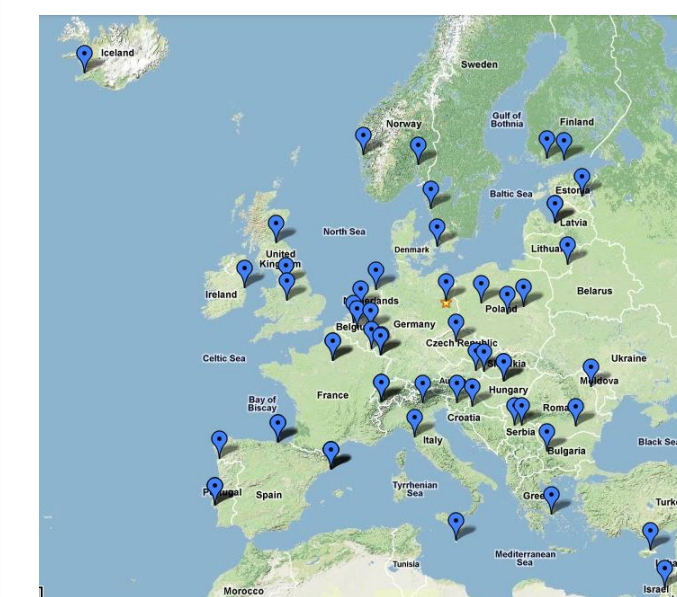- Multilingual Europe
  Technology Alliance

META=NET

STRATEGIC
RESEARCH
AGENDA FOR
MULTILINGUAL
EUROPE 2020

edited by the
META Technology Council

META=NET

White Paper Series   Fehér könyvek sorozat

THE   A MAGYAR
HUNGARIAN   NYELV A
LANGUAGE IN DIGITÁLIS
THE DIGITAL KORBAN
AGE

Simon Eszter
Lendvai Piroska
Németh Géza
Olaszy Gábor
Vicsi Klára

Springer

Published in 2013   31 volumes, published in 2012

ELG   European Language Grid

META=FORUM 2015

META=FORUM 2017

MULTILINGUAL DIGITAL SINGLE MARKET — Riga Summit 2015

META=FORUM 2010

META=SHARE

T4ME

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)

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

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

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

Strategic Research and Innovation Agenda (V1.0)
Launch of the Strategic Research and Innovation Agenda titled "Language Technologies for Multilingual Europe – Towards a Human Language Project" (December 2017)

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

META=FORUM 2012

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

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

META=NET

2010   2011   2013   2015   2017

META NORD

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

2012

METANET4U
EU-funded project (ICT Policy Support Programme) 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)

CESAR — CEntral and South-east europeAn Resources

"State of the Art of Machine Translation – Current Challenges and Future Opportunities"
Workshop on "State of the Art of Machine Translation", commissioned by the EU Parliament (December 2013)

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

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

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

2014

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)

2016

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

2018

EUROPEAN LANGUAGE GRID

2019-2021

META=FORUM 2011

META=FORUM 2013

META=FORUM 2016

COMPRISE   ELITR   Prêt-à-LLOD   EMBEDDIA   gourmet — Global Under Resourced MEdia Translation

CRACKER — Cracking the Language Barrier

Cracking the Language Barrier

ELG consortium — DFKI · IEA LSP · The University Of Sheffield · ELDA · TILDE · SAILLABS TECHNOLOGY · EXPERT SYSTEM · THE UNIVERSITY of EDINBURGH
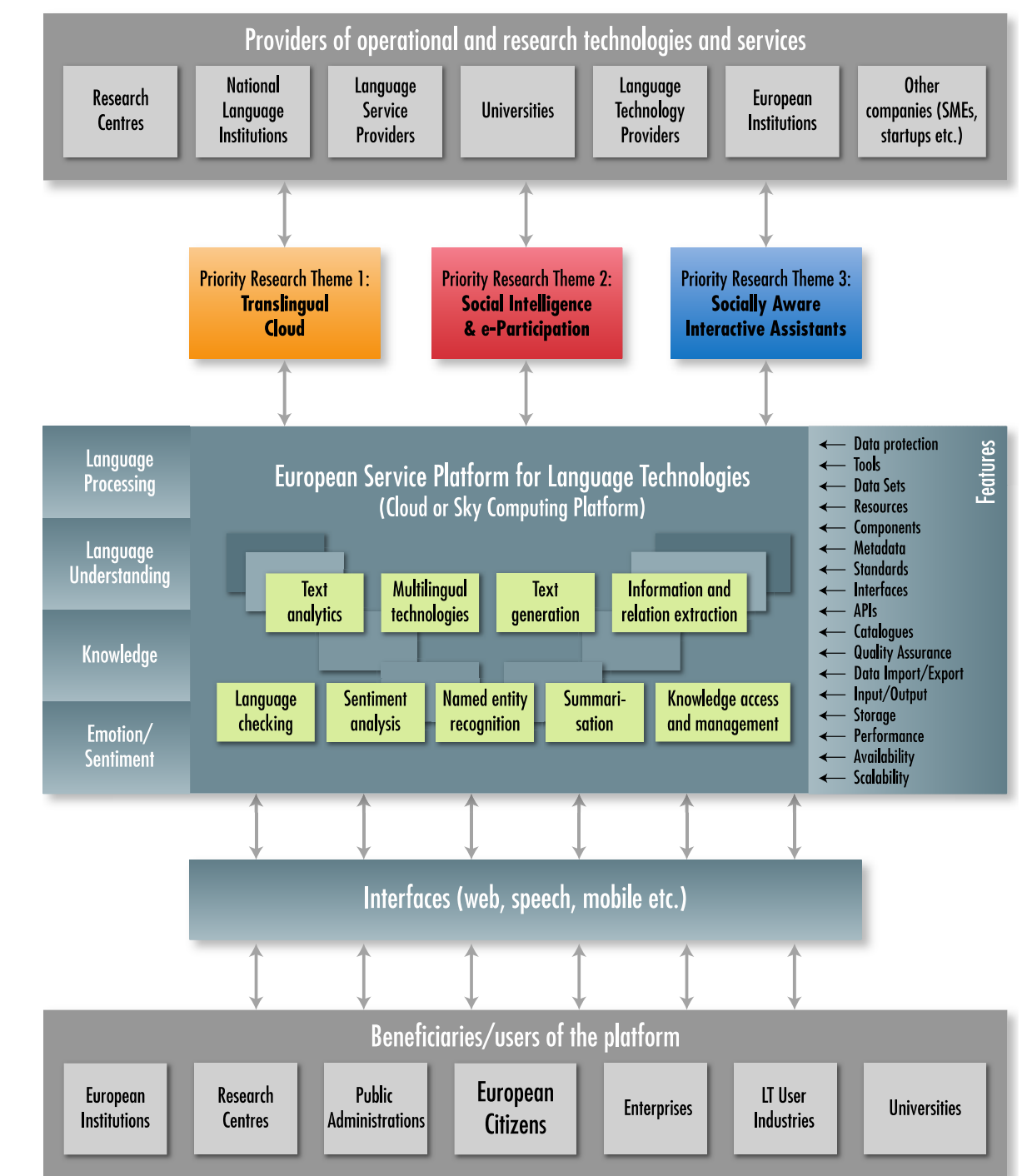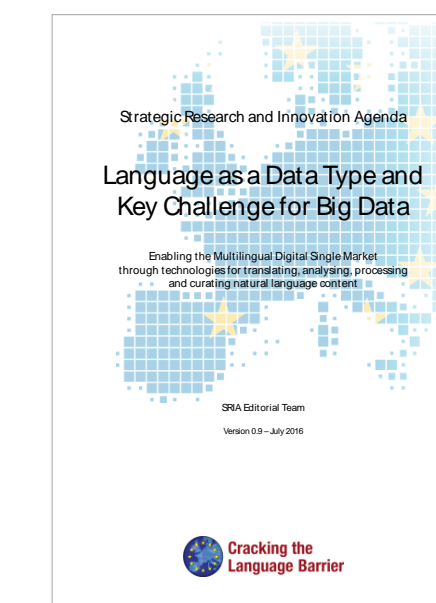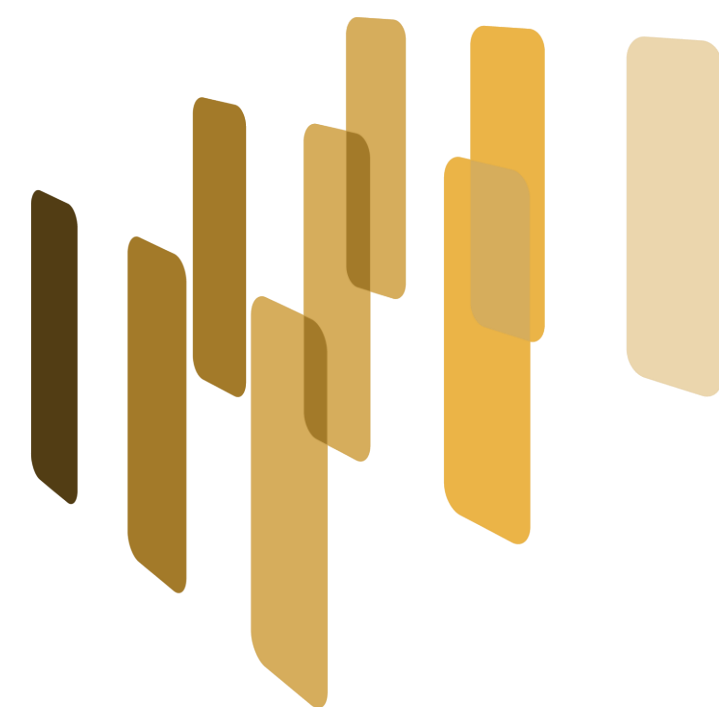
ELG   European Language Grid

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).
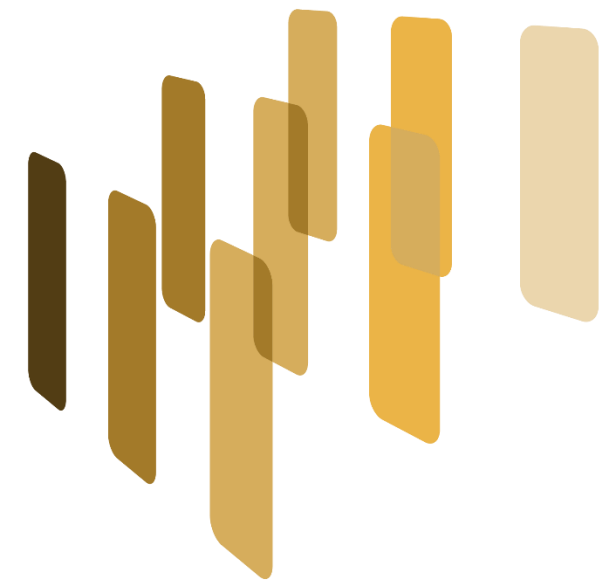
META FORUM 2013

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

ELG Project
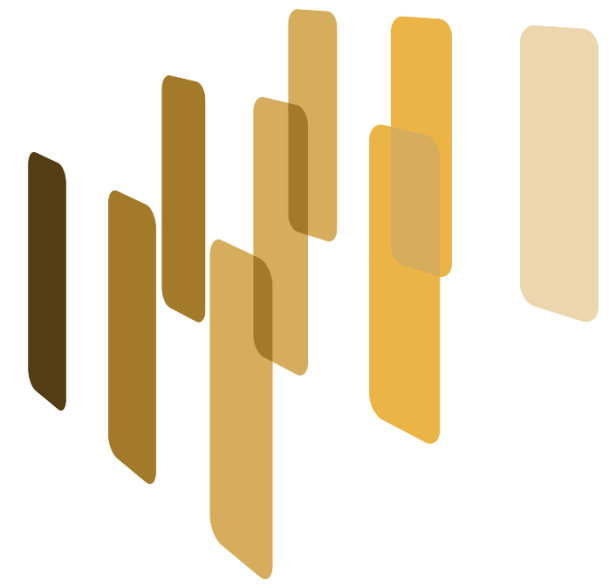
EUROPEAN LANGUAGE GRID

Objectives (Selection)

1. Establish the ELG as the primary LT 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

Kick-off meeting, 22/23 January 2019

Good relationship with all relevant initiatives

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

EUROPEAN LANGUAGE GRID

Concept and Methodology

- **Grid Platform**

- **Grid Content**

- **Grid Community**

EUROPEAN LANGUAGE GRID

Concept and Methodology

Highlights in Year 1

- Three-layered platform architecture ready: frontend; backend; base infrastructure
- Technical decisions for all components
- Base infrastructure up and running
- Operated on a Kubernetes cluster in the data centre of a Berlin-based cloud provider (SysEleven)
- All components running as Docker containers

- **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**

Highlights in Year 1

- Two iterations of API development

- Several functional services integrated in the MVP: IE (USFD, EXPSYS, CUNI); MT (TILDE); ASR (SAIL, TILDE); TTS (TILDE)

- Ensuring maximum coverage of EU languages through prioritisation

- Commercial-grade service integration and robustness through industrial and open-source NLP/LT leaders responsible for integration

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.

Highlights in Year 1

- Calls for Open Pilots fully set up (Pilot Board, call, survey etc.)

- NCC network set up

- META-FORUM 2019

- Collaborations: ICT-29b projects, AI4EU, several other projects and initiatives

- Sustainability plan in progress

- LTC concept/vision and population

- LREC 2020 workshop and tutorial

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

Community: NCCs and LTC
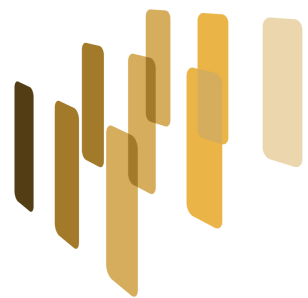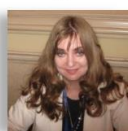
32 National Competence Centres (NCCs)

- Strong international network of national networks to broaden ELG's reach, identify content for the ELG and interest companies in using the ELG.

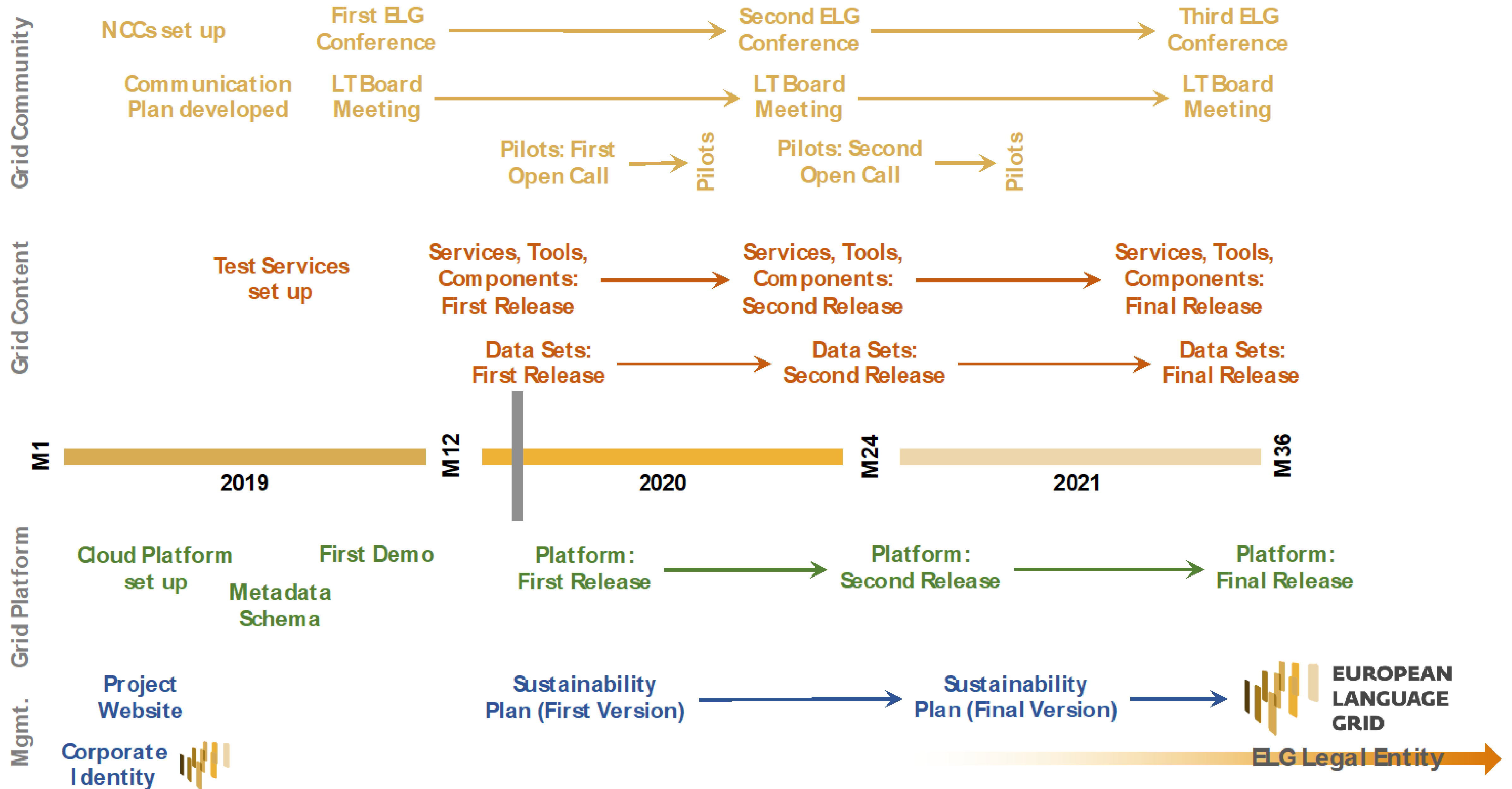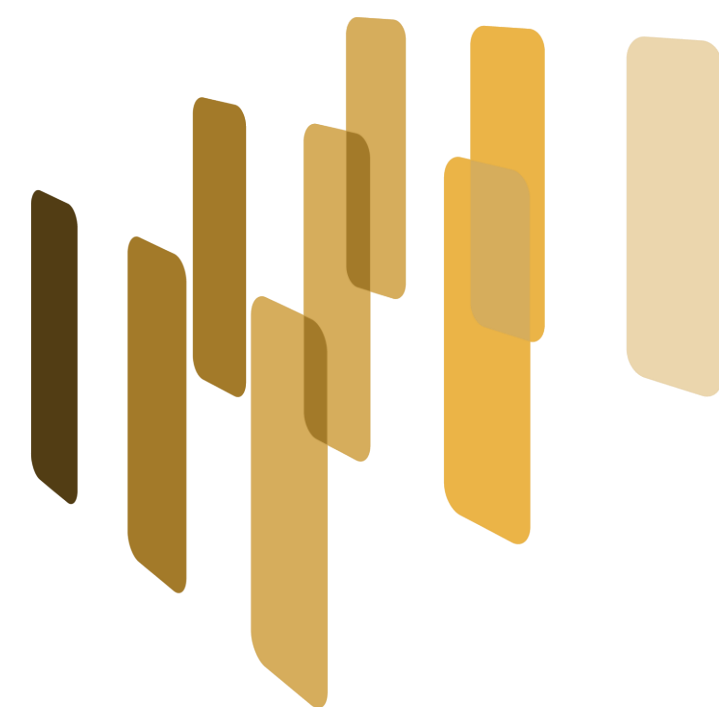- Main goal: support the mission of the ELG project.

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

GERHARD BUDIN
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 TADIC
Department of Linguistics, University of Zagreb

Cyprus

FRYNI KAKOYIANNI DOA
University of Cyprus

Czech Republic

JAN HAJIČ
Institute of Formal and Applied Linguistics, Charles University Prague

Denmark

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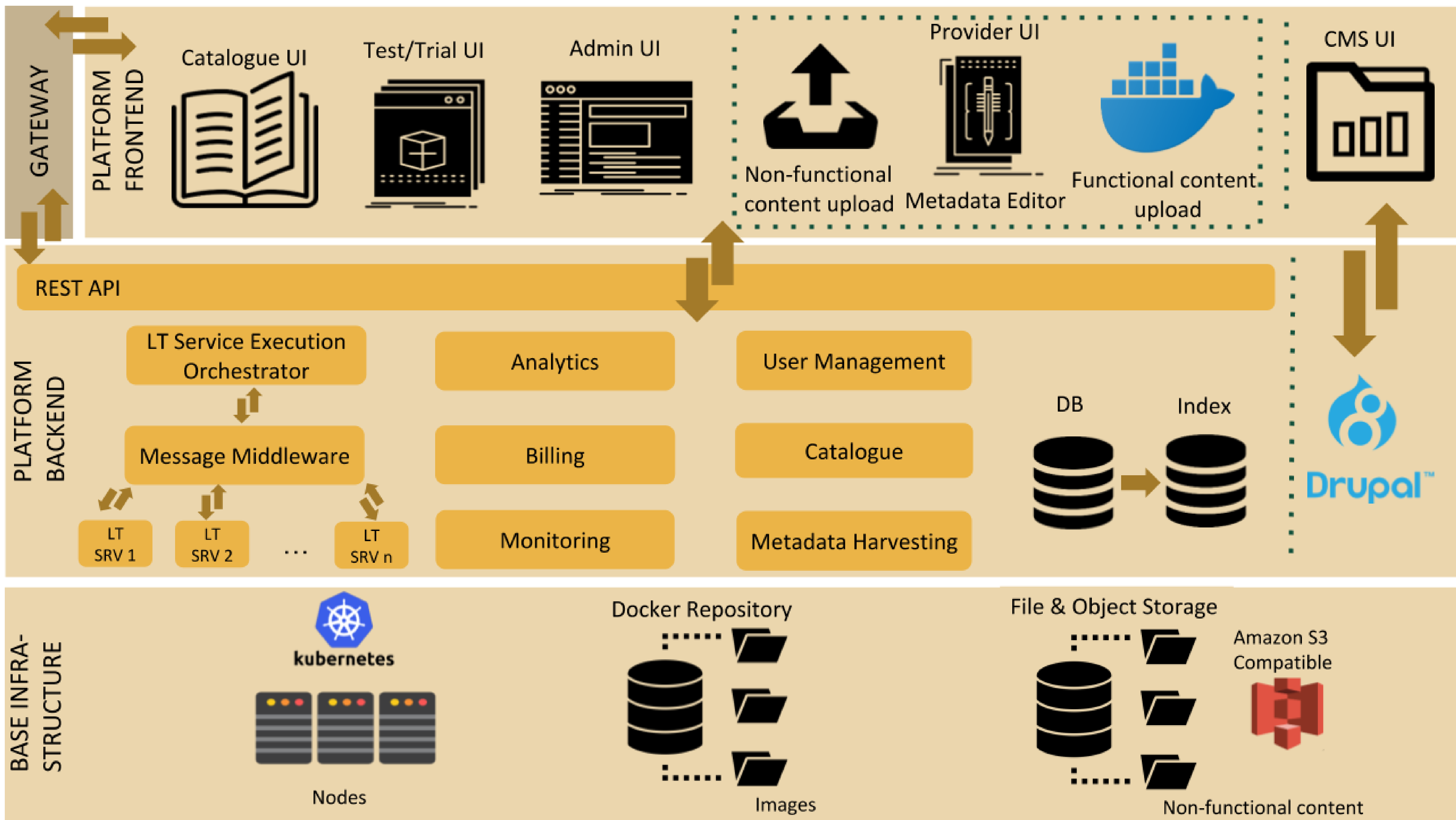Centre for Language Technology, Department of Nordic Studies and Linguistics, University of Copenhagen

Grid Community

NCCs set up   First ELG Conference → Second ELG Conference → Third ELG Conference

Communication Plan developed   LT Board Meeting → LT Board Meeting → LT Board Meeting

Pilots: First Open Call → Pilots   Pilots: Second Open Call → Pilots

Grid Content

Test Services set up

Services, Tools, Components: First Release → Services, Tools, Components: Second Release → Services, Tools, Components: Final Release

Data Sets: First Release → Data Sets: Second Release → Data Sets: Final Release

M1 — 2019 — M 12 — 2020 — M24 — 2021 — M 36

Grid Platform

Cloud Platform set up   First Demo   Metadata Schema

Platform: First Release → Platform: Second Release → Platform: Final Release

Mgmt.

Project Website

Sustainability Plan (First Version) → Sustainability Plan (Final Version) → EUROPEAN LANGUAGE GRID

Corporate Identity

ELG Legal Entity →

European Language Grid

ELG Platform – Overview of Release 1

GATEWAY

PLATFORM FRONTEND

Catalogue UI

Test/Trial UI

Admin UI

Provider UI

Non-functional content upload

Metadata Editor

Functional content upload

CMS UI

REST API

PLATFORM BACKEND

LT Service Execution Orchestrator

Analytics

User Management

Message Middleware

Billing

Catalogue

LT SRV 1

LT SRV 2

...

LT SRV n

Monitoring

Metadata Harvesting

DB

Index

Drupal™

BASE INFRA-STRUCTURE

kubernetes

Nodes

Docker Repository

Images

File & Object Storage

Amazon S3 Compatible

Non-functional content

ELG   European Language Grid

17

ELG R1: Services

- **ELG R1** (April 2020): 181 services
  - IE and Text Analytics: 154
  - MT: 18
  - ASR: 7
  - TTS: 2
- **ELG R3** (Dec. 2021): 721 services

	A	B	C	D	Total
ASR	12	3	9		24
IE & Text Analysis	377	58	157	26	618
Other	7	1	1	2	11
MT					
A	57		6	1	64
C	4				4
Total	**457**	**62**	**173**	**29**	**721**
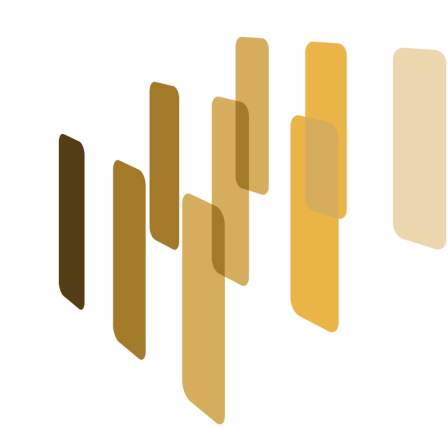
MT

CUNI
- Czech-English
- English-Czech
- English-French
- French-English

DFKI
- English-French
- English-German
- English-Spanish

ILSP
- English-Greek
- Greek-English

Tilde
- English-Bulgarian
- English-Latvian
- English-Polish
- Latvian-English
- Polish-English

UEDIN
- Czech-English
- English-Czech
- English-German
- German-English

ASR

SAIL LABS
- English
- French
- German
- Greek
- Spanish

Tilde
- Latvian

UEDIN
- Czech

TTS

Tilde
- Latvian
- Lithuanian

| Etiquetas de fila	Czech	English	French	German	Greek	Latvian	Spanish	Total general
CUNI	**6**	**6**	**5**	**5**	**5**	**5**	**5**	**37**
Dependency Parsing	1	1	1	1	1	1	1	7
Lemmatisation	1	1	1	1	1	1	1	7
Morphological analyser	1	1	1	1	1	1	1	7
Named Entity Recognition	1	1						2
Part of Speech tagging	1	1	1	1	1	1	1	7
Tokenization	1	1	1	1	1	1	1	7
DFKI	**1**	**8**	**2**	**6**	**1**	**1**	**2**	**21**
Categorization		1						1
Language identification	1	1	1	1	1	1	1	7
Morphological analyser		1	1	1			1	4
Named Entity Recognition		2		2				4
Sentence splitting		1	1					2
Summarization		1						1
Tokenization		1		1				2
Expert System	**1**	**7**	**6**	**5**	**1**	**1**	**6**	**27**
Categorization		1					1	2
Language identification	1	1	1	1	1	1	1	7
Lemmatisation		1	1	1			1	4
Named Entity Recognition		1	1	1			1	4
Part of Speech tagging		1	1	1			1	4
Sentiment Analysis		1	1					2
Summarization		1	1				1	4
ILSP		**1**		**4**				**5**
Information Extraction				2				2
Named Entity Recognition		1		1				2
Sentiment Analysis				1				1
SAIL LABS	**2**	**3**	**3**	**3**	**2**		**3**	**16**
Language identification	1	1	1	1	1		1	6
Named Entity Recognition	1	1	1	1	1		1	6
Sentiment Analysis		1	1	1			1	4
USFD	**1**	**27**	**5**	**9**	**1**	**1**	**4**	**48**
Categorization		4						4
Language identification		1	1	1		1		4
Morphological analyser		1						1
Named Entity Recognition		7	2	4				13
NER Disambiguation		4	1	1			1	7
Number annotation		1						1
Opinion Mining		2						2
Part of Speech tagging	1	3	1	1	1	1	1	9
Sentence splitting		1		1				2
Summarization		1					1	2
Tokenization		2		1				3
Total general	**11**	**52**	**21**	**28**	**14**	**8**	**20**	**154**

IE and Text Analytics

ELG R1: Data Sets and LRs

- **ELG R1** (April 2020): 502 LRs
  - ELRA-SHARE: 103 LRs
  - ELRC-SHARE: 187 LRs
  - META-SHARE: 212 LRs
- Focus on freely available language resources
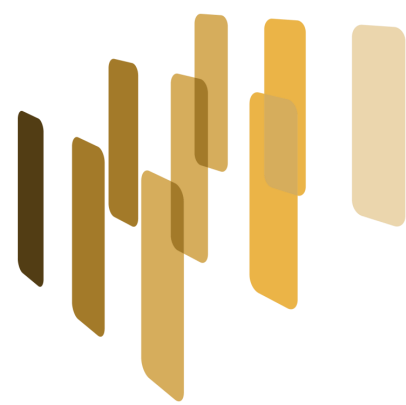
ELRA-SHARE	Resource Type	#LRs
Text	Corpus	49
	Lexical conceptual database	6
Audio	Corpus	48
Video	Corpus (combined with text)	1
	Corpus (combined with audio)	2

ELRC-SHARE	Linguality Type	#LRs
Corpus	Monolingual	2
	Bilingual	168
	Multilingual	10
Lexical conceptual database	Bilingual	6
	Multilingual	1

META-SHARE	Resource Type	#LRs
DFKI	Corpus	7
ILSP	Corpus	182
	Linguistic description	5
	Lexical conceptual database	18

European Language Grid
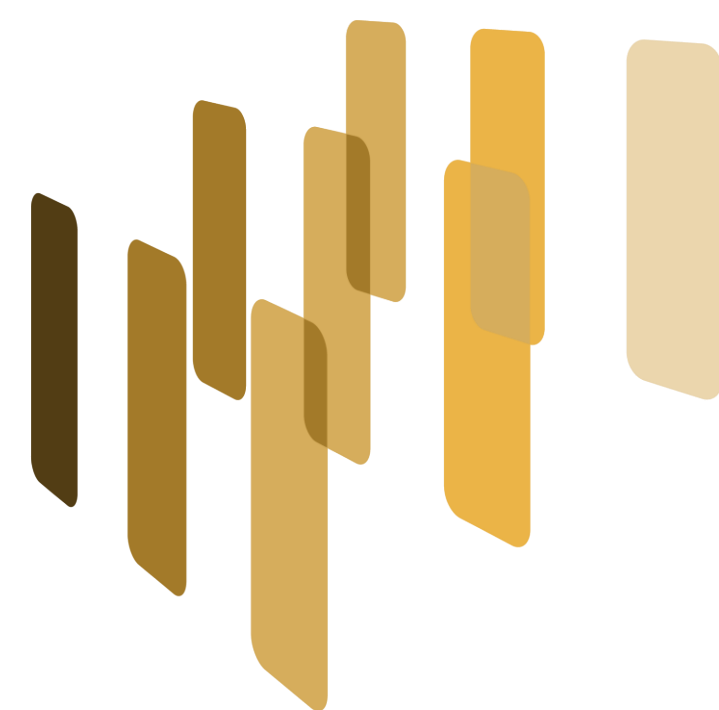
European Language Grid

Open Calls and Pilot Projects

Open Calls for Pilot Projects: Summary

EUROPEAN LANGUAGE GRID

Two open calls for demonstrator or integration projects

- 1st Call: **March – April 2020**, 2nd Call: **September – October 2020**

Pilot projects shall

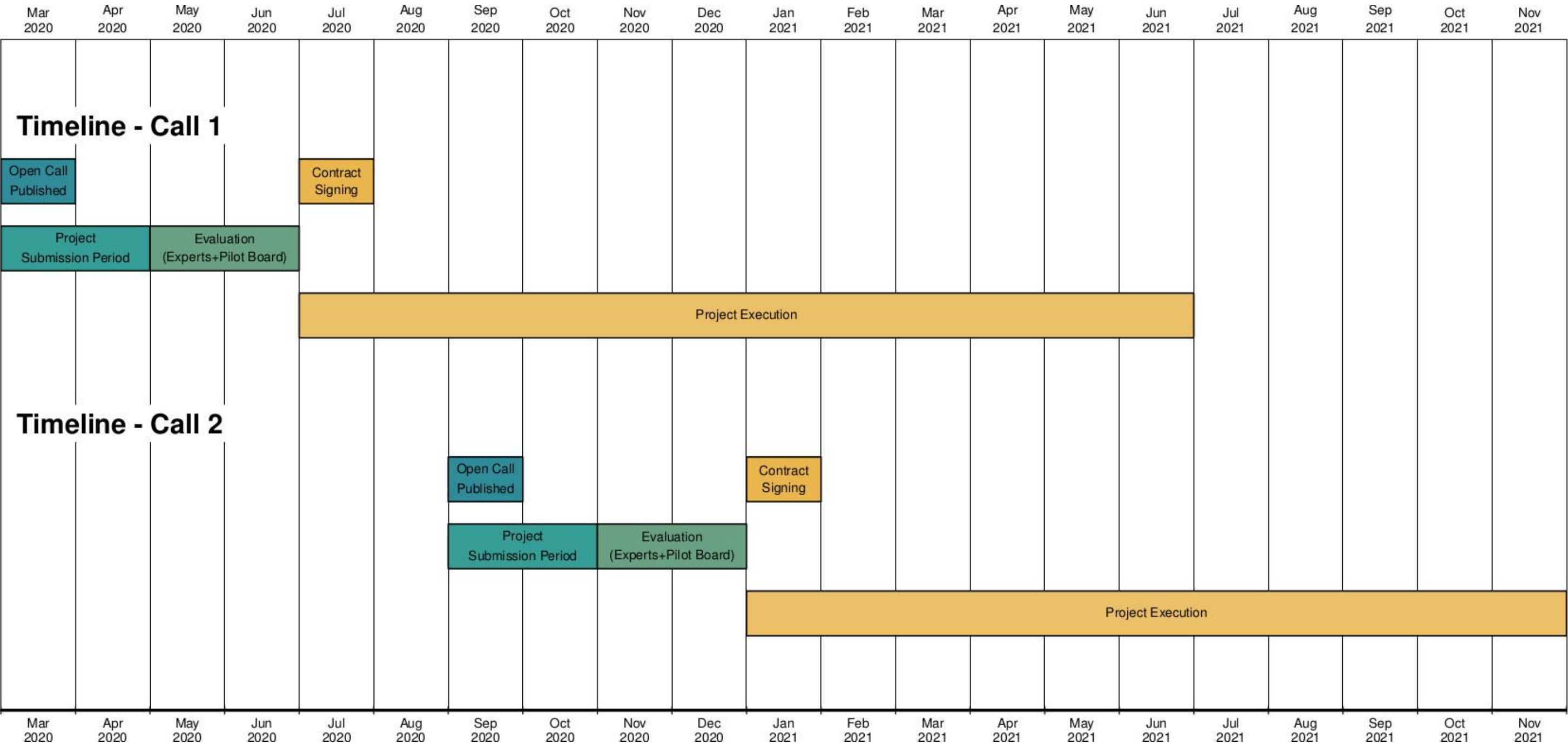- broaden ELG's portfolio of language technologies and data sets

- demonstrate usefulness of ELG
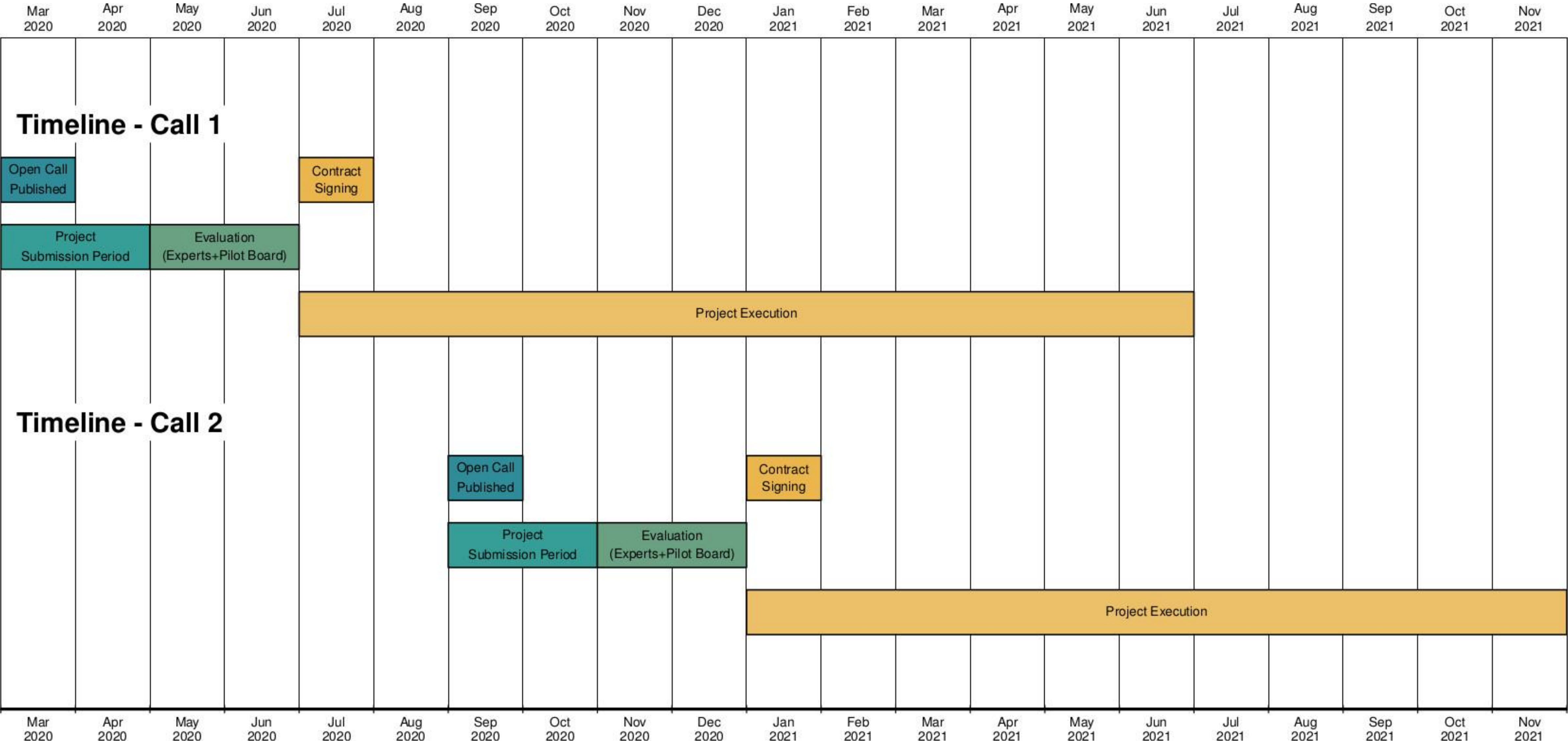
Up to €200,000 per project
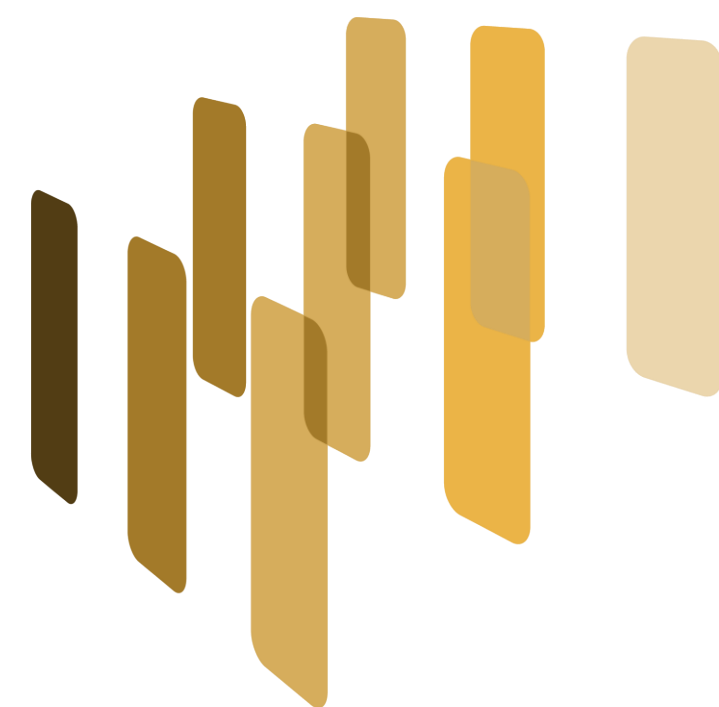
- Approx. €2,000,000 in total

Duration: **9-12 months**

Eligibility: **SMEs**, **research orgs.**

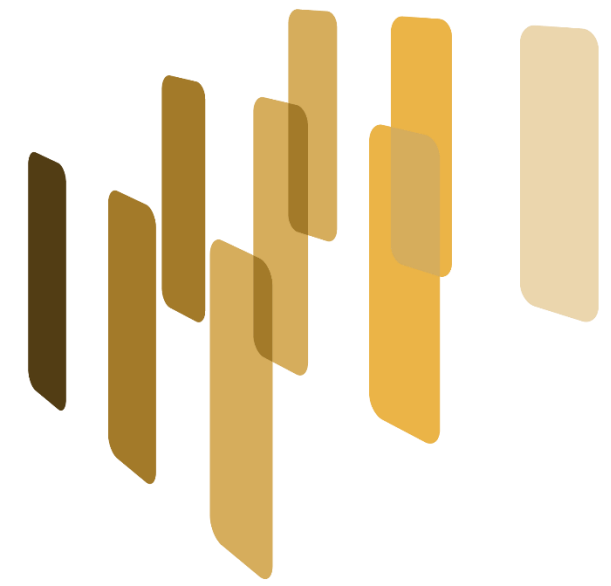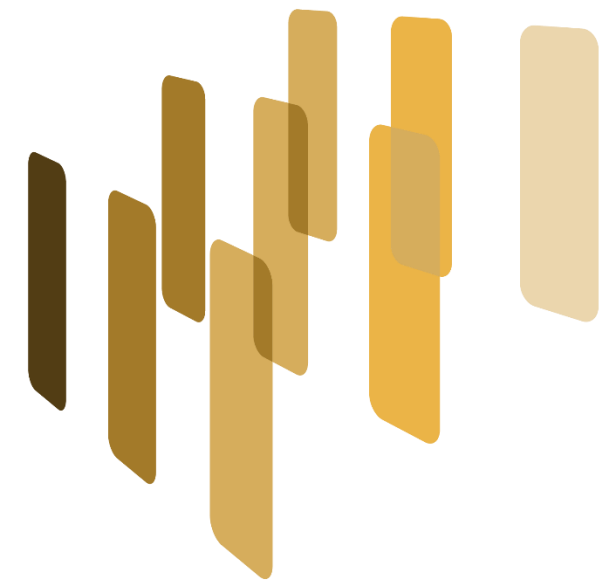# Open Calls for Pilot Projects: Timeline

European Language Grid

Summary and Next Steps

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.

- ELG caters for both LT industry and LT research.

- 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.

Next Steps

EUROPEAN LANGUAGE GRID

- 29 Feb. 2020: **Launch ELG Release 1 alpha**

- 01 March 2020: **Publication of first open call for pilots**

- 30 April 2020: **Launch ELG Release 1**

- 16 May 2020: **ELG Tutorial 2020: How to ingest your Language Technology service or data set into the European Language Grid** (at LREC 2020)

- 16 May 2020: **IWLTP 2020 – 1st International Workshop on Language Technology Platforms** (at LREC 2020)

- 23 June 2020: **ELG – An Introduction and Overview. 1st Regional ELG Workshop: Switzerland, Austria, Germany**

- **META-FORUM 2020** – Second Annual ELG Conference

META-FORUM 2019
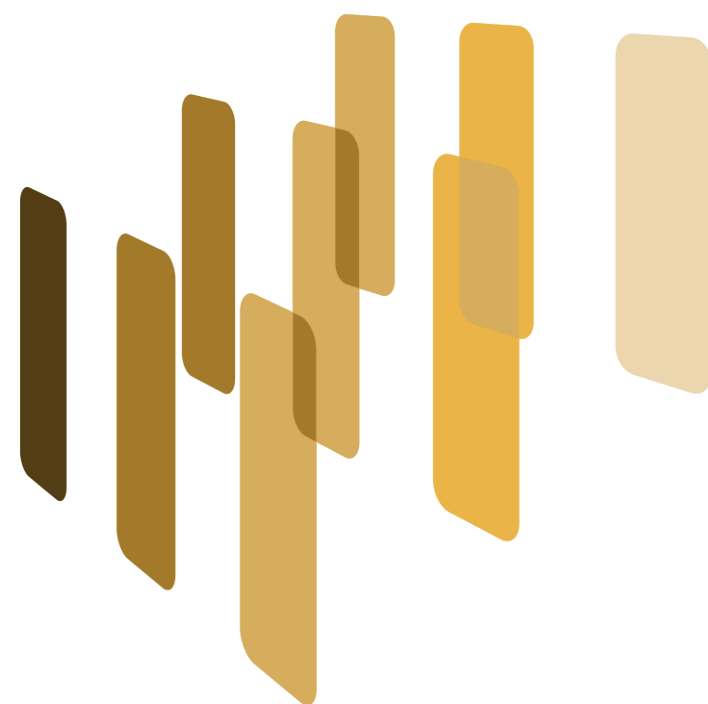
European Language Grid

Thank you!

Raivis Skadiņš (TILDE)
raivis.skadins@tilde.lv

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

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

Interested in participating?

Please get in touch:

contact@european-language-grid.eu