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 • 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)

META-NET (T4ME) • Multilingual Europe Technology Alliance

META-NORD

METANET4U • Published in 2013

31 volumes, published in 2012

Multilingual Europe Technology Alliance
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 0.9)
Launch of the Strategic Agenda for the Multilingual Digital Single Market titled "Language as a Data Type and Key Challenge for Big Data" (April 2016)

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

"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)

META-FORUM 2015
Conference "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)

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

2011
 META-FORUM 2011
Conference "Connecting Europe for New Horizons" (Berlin – September 18/20, 2011)

2012
 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

2013
 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

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)

2015
 META-FORUM 2016
Conference "Beyond Multilingual Europe" (Lisbon – July 4/5, 2016)

2016
 ELG Final Proposal Submission
Final submission on Feb. 20, 2018

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

2018
 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)

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: 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)

META-NET
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-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
EU-funded project (ICT Policy Support Programme) to enhance the European Linguistic Infrastructure (2011 – 2013)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

META-NET
EU-funded project (ICT Policy Support Programme) functioning as a part of META-NET to standardise language resources and tools (2011 – 2013)
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\textsuperscript{NET}** vision: European Service Platform for Language Technologies.

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

ELG Project
Objectives (Selection)

1. Establish the ELG as the primary LT platform and marketplace 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.
Consortium

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
Europeana 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

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

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**
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.
ELG Platform – Overview of Release 1
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

### ELG R1: Services

<table>
<thead>
<tr>
<th>Service Type</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASR</td>
<td>24</td>
</tr>
<tr>
<td>IE &amp; Text Analysis</td>
<td>618</td>
</tr>
<tr>
<td>MT</td>
<td>11</td>
</tr>
<tr>
<td>ASR</td>
<td>18</td>
</tr>
<tr>
<td>TTS</td>
<td>2</td>
</tr>
</tbody>
</table>

### ELG R3: Services

<table>
<thead>
<tr>
<th>Service Type</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASR</td>
<td>721</td>
</tr>
</tbody>
</table>

### Table:

<table>
<thead>
<tr>
<th>Service</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>26</td>
</tr>
<tr>
<td>ASR</td>
<td>12</td>
</tr>
<tr>
<td>IE &amp; Text Analysis</td>
<td>377</td>
</tr>
</tbody>
</table>

### Table:

<table>
<thead>
<tr>
<th>Service</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT</td>
<td>618</td>
</tr>
<tr>
<td>ASR</td>
<td>11</td>
</tr>
<tr>
<td>TTS</td>
<td>2</td>
</tr>
</tbody>
</table>

### Table:

<table>
<thead>
<tr>
<th>Service</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>5</td>
</tr>
<tr>
<td>ASR</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table:

<table>
<thead>
<tr>
<th>Service</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT</td>
<td>1</td>
</tr>
<tr>
<td>ASR</td>
<td>1</td>
</tr>
<tr>
<td>TTS</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table:

<table>
<thead>
<tr>
<th>Service</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>1</td>
</tr>
<tr>
<td>ASR</td>
<td>1</td>
</tr>
<tr>
<td>TTS</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table:

<table>
<thead>
<tr>
<th>Service</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT</td>
<td>1</td>
</tr>
<tr>
<td>ASR</td>
<td>1</td>
</tr>
<tr>
<td>TTS</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table:

<table>
<thead>
<tr>
<th>Service</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>1</td>
</tr>
<tr>
<td>ASR</td>
<td>1</td>
</tr>
<tr>
<td>TTS</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table:

<table>
<thead>
<tr>
<th>Service</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT</td>
<td>1</td>
</tr>
<tr>
<td>ASR</td>
<td>1</td>
</tr>
<tr>
<td>TTS</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table:

<table>
<thead>
<tr>
<th>Service</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>1</td>
</tr>
<tr>
<td>ASR</td>
<td>1</td>
</tr>
<tr>
<td>TTS</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table:

<table>
<thead>
<tr>
<th>Service</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT</td>
<td>1</td>
</tr>
<tr>
<td>ASR</td>
<td>1</td>
</tr>
<tr>
<td>TTS</td>
<td>1</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Resource Type</th>
<th>#LRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td></td>
</tr>
<tr>
<td>Corpus</td>
<td>49</td>
</tr>
<tr>
<td>Lexical conceptual database</td>
<td>6</td>
</tr>
<tr>
<td>Audio</td>
<td></td>
</tr>
<tr>
<td>Corpus</td>
<td>48</td>
</tr>
<tr>
<td>Video</td>
<td></td>
</tr>
<tr>
<td>Corpus (combined with text)</td>
<td>1</td>
</tr>
<tr>
<td>Corpus (combined with audio)</td>
<td>2</td>
</tr>
</tbody>
</table>

### ELRC-SHARE

<table>
<thead>
<tr>
<th>Linguality Type</th>
<th>#LRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corpus</td>
<td></td>
</tr>
<tr>
<td>Monolingual</td>
<td>2</td>
</tr>
<tr>
<td>Bilingual</td>
<td>168</td>
</tr>
<tr>
<td>Multilingual</td>
<td>10</td>
</tr>
<tr>
<td>Lexical conceptual database</td>
<td></td>
</tr>
<tr>
<td>Bilingual</td>
<td>6</td>
</tr>
<tr>
<td>Multilingual</td>
<td>1</td>
</tr>
</tbody>
</table>

### META-SHARE

<table>
<thead>
<tr>
<th>Resource Type</th>
<th>#LRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>DFKI</td>
<td></td>
</tr>
<tr>
<td>Corpus</td>
<td>7</td>
</tr>
<tr>
<td>ILSP</td>
<td></td>
</tr>
<tr>
<td>Corpus</td>
<td>182</td>
</tr>
<tr>
<td>Linguistic description</td>
<td>5</td>
</tr>
<tr>
<td>Lexical conceptual database</td>
<td>18</td>
</tr>
</tbody>
</table>
Open Calls and Pilot Projects
Open Calls for Pilot Projects: Summary

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

Timeline - Call 1
- Open Call Published
- Evaluation (Experts+Pilot Board)
- Project Submission Period
- Contract Signing
- Project Execution

Timeline - Call 2
- Open Call Published
- Evaluation (Experts+Pilot Board)
- Project Submission Period
- Contract Signing
- Project Execution
Summary and Next Steps
• 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

• 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
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