

*HORIZON 2020 – Work Programme 2016-2017
Information and Communication Technologies*

ICT-29 Multilingual Next Generation Internet

ICT Proposer's Day Budapest - 9-10 November

Aleksandra Wesolowska

DG CONNECT Unit G3 "Learning, Multilingualism & Accessibility"

Next Generation Internet

Artificial Intelligence

Interactive Technologies

Internet of Things

Multilingual

Global Social Sphere

Inclusive

NGI - Open Internet Initiative

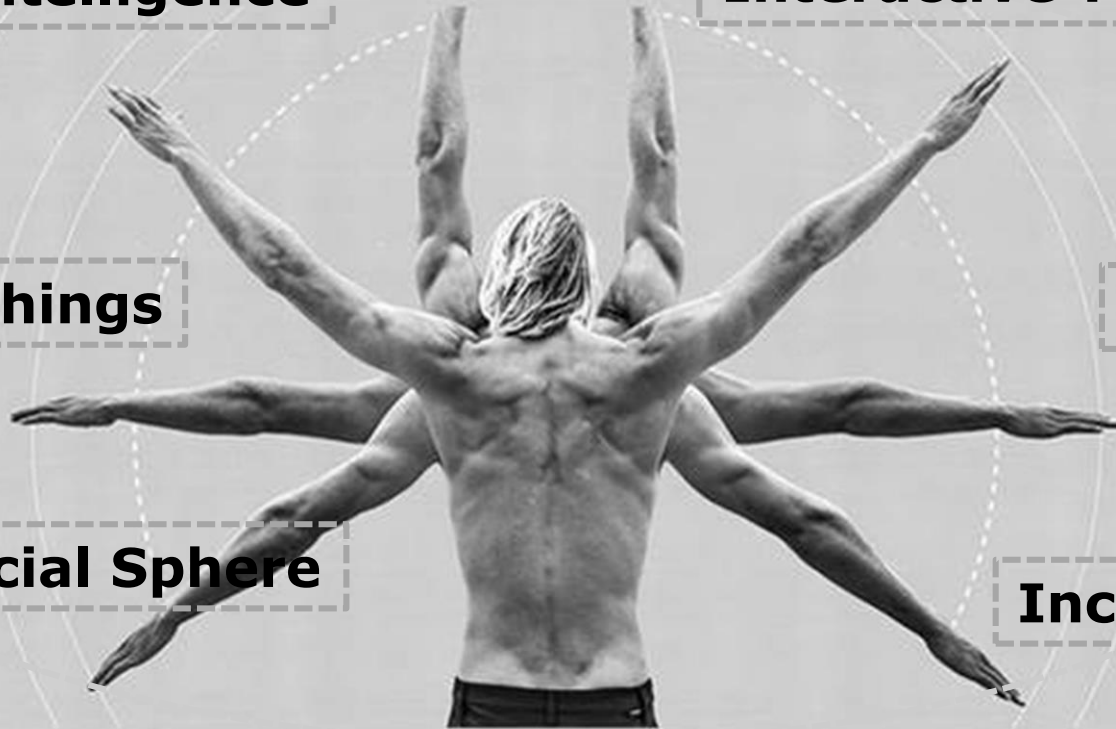
5G

Cloud

HPC

Big
Data

Cyber
Security



Cultural and Linguistic Diversity

24 official EU languages and over 60 regional or minority languages spoken by approximately **400 million** people

Language technologies are key

- to seize the opportunities and benefits of the Digital Single Market
- to cross-border provision and access to online content and services
- inclusiveness - access to knowledge and information



Language Technologies: Key Enabling Technology

- **Embedded in mobile communications, social media, intelligent assistants**
- **Enable innovative IT products and services for e-commerce, e-government, education, healthcare, media, tourism...**
- **Bridge barriers to cross-border online services (commercial & public)**
- **Achieve high quality technological solutions for all European languages**
- **Allow all European citizens and businesses to fully participate in and benefit from the Digital Single Market**

ICT-29-2018: A multilingual Next Generation Internet

Budget: 25 MEUR

- a) Innovation Action: A **European Language Grid**

Budget: 7 MEUR → one action

funding rate: 70% (100% for for non-profit legal entities)

- b) Research and Innovation Actions: **Domain-specific/challenge-oriented Human Language Technology**

Budget: 18 MEUR → several actions

funding rate: 100%

Call opened: 31 October 2017
Call closes: 17 April 2018

a) IA: A European Language Grid (1)

i. *architecture and components for a **public, open and interoperable grid** connecting resources and tools to support effective development and deployment of language technologies (software and services) across Europe*

- ecosystem for accessing tools, services and data resources
- enable effective pooling and usage of existing but often "dispersed" tools, services and data resources for EU languages
- lowering barrier of entry/facilitating growth for SMEs active in LT
- **not only technology integration issues** but also **organisational aspects and obstacles**, e.g. maintenance responsibility; connection modalities; services type and scope; billing models if you combine tools/services from several providers; confidentiality requirements for language resources data sets etc.
- end users, i.e. developers of multilingual solutions and services based on LT **must be tightly involved in the design, testing and appreciation of the grid.**

a) IA: A European Language Grid (2)

ii. **Coordinate** the work of European Language grid...

- relevant results of small-scale projects and RIA actions become part of the grid (interoperability)
- structured dialogue with relevant stakeholders and activities
- ... leading to an improved approach underlying the grid (legal, organisational, technological aspects)

iv. **establish competence centres / nodes in Member States**

- to facilitate coordination with national and regional activities
- building on previous EC-funded actions within the FP7, H2020 and CEF and existing networks/federations Member States (CLARIN, EFNIL, CEF Automated Translation National Anchor Points...)

a) IA: A European Language Grid (3)

iii. pilot the European Language Grid ... through small scale demonstrators (mini-projects for third parties)

- cascading grants instrument (Annex K of the general Annexes)
 - best-practice examples: H2020 ODINE and DATA PITCH projects
- see in particular "Guide for Applicants" documents
<https://opendataincubator.eu/apply/>, <https://datapitch.eu/apply/>
- aspects to consider: support activities for mini-projects; KPIs for measuring success of the mini-project-scheme for LT; relationship between mini-projects and the grid (results, feedback...)

a) IA: A European Language Grid

- not another website/portal/observatory with a directory/links to organizations providing tools, services, language resources
- not another data or tools repository
- relationship with and input to to ICT-26 Artificial Intelligence platform

b) RIA: Domain-specific/challenge-oriented Human Language Technology

- researchers and industrial users of language technologies working together on R&I challenges:
 - in a **specific sector** of high commercial and/or societal impact, or
 - a **technological challenge** common/relevant to several sectors
- challenges defined by industrial users
- aim of sectorial projects: *create a sustainable ecosystem of multilingual applications and services tailored for the specific needs of the addressed sector*
- *detailed analysis of the expected advances in LT-related research*

Expected impacts

- Provide European research and language technology industry with a **better access to and usage of quality language resources and tools**;
- Increase in the **quality and coverage of multilingual solutions used by industrial players** in sectors relevant to the emergence of the Digital Single Market;
- **Increase in the uptake of language technologies** in Europe in various sectors;
- **Cost savings** for private and public sector users of language technology solutions.

Language technologies - topic evolution

- support for language technologies through FP7, CIP ICT-PSP, and H2020 Call 1
- "*infrastructural*" projects in FP7: META-NET (META-SHARE) (<http://www.meta-share.org/>) and MLI project <http://mli-project.eu/> Towards a Multilingual Data & Services infrastructure
- **H2020 Call 1 ICT-17: Cracking the language barrier (15 MEUR)** had a narrow focus on machine translation
- some LT-related projects emerged from H2020 Big Data calls
- Overview of recent and ongoing EU-funded projects related to LT: <http://www.cracking-the-language-barrier.eu>

examples of H2020 LT projects

Call 1 ICT-17 projects on machine translation

- [MMT](#): adaptive domain-sensitive MT infrastructure
- [QT21](#): SoA neural MT technology
- [TraMOOC](#): Translation for Massive Open Online Courses
- [HimL](#): automated translation of public health information



Examples of H2020 Big Data projects with LT dimension

- [FREME](#): semantic enrichment of digital content
- [Kconnect](#): ML search for medical information
- [SUMMA](#): media analytics Big Data platform - automated analysis of media streams across many languages



actors and initiatives

- European associations & organisations: **META-NET**, **LT-Innovate**, CLARIN, EFNIL, ELDA, ...
- Initiatives and actors relevant for "language grid":
 - non-European commercial players: Google Cloud Natural Language API, Microsoft Cognitive Services...,
 - open-source NLP toolkits, e.g. StanfordCore NLP, GATE...,
 - NLP tools and services provided by national labs/research centres/companies for specific languages, e.g. Wrocław University of Technology, Charles University in Prague, Tilde tools and services for Baltic languages...

additional/background documents

- Language Infrastructures for a Leading Europe – **LT-Innovate position paper** from 2016
- [European Platform for the Multilingual for Digital Single Market: Conceptual Proposal](#)
- **Strategic Agenda for Multilingual Digital Single Market** – version 1.0 to be presented at upcoming META-FORUM
- [Report from the roundtable on language technologies](#) on 13 December 2016 in Luxembourg
- ODINE [Final report on lessons learnt and best practice](#)
- Other documents: META-NET White Papers, Conversational interaction Technologies Roadmap (CITIA)...
- **CEF Automated Translation** – deployment of mature language technologies for public online services