



Kimmo Rossi

European Commission
DG CONNECT

Unit G.3 - Data Value Chain



I will present topics of interest from the current state of:

- Horizon 2020
- Connecting Europe Facility
- Structural funds

Disclaimer: the programmes described in this presentation are at the level of proposal/draft and subject to change following a multi-party adoption process. Only the versions formally adopted by the legislative bodies can be considered final and binding.



- A single programme for Research and Innovation
- Innovation is intertwined with research, part of every level and phase, not only the "final" stage
- More coordination across programme and between programmes
- Fast, flexible, open funding schemes for SMEs
- Challenge-based, less prescriptive



- Open: no themes, any good idea goes
- Continuous calls – only for SMEs
- Three separate phases:
 - 1. Concept and feasibility assessment: short proposal, quick evaluation, small lump sum grant**
 - 2. Validation & piloting & scaling up: innovation grant 12-18 months.**
 - 3. Commercialisation: pre-commercial procurement (PCP), access to risk finance, label of recognition, indirect support**



Some likely topics:

- Big Data – Research
- Big Data – Innovation
- Cracking the language barrier
- Multimodal & natural interaction



(Big/Smart/Open) Data represents significant potential for job creation and wealth (revenue) generation

European Data landscape is fragmented

- data producers: public sector, private sector, sensors, devices, transactions...
- data industry: a few large companies, high number of SMEs
- language barriers, legal & institutional obstacles
- data users: practically all industry sectors

We need to create European Data ecosystem, where

- data flows without unnecessary obstacles (open data, interoperability)
- data creates **value** (by analysing, linking, visualizing...)
- data and the added value is **accessible** to all actors
- we have appropriate **capacities** for value creation (skills, infrastructures, markets, technology)



Improve EU capacities, develop generic technology and address **entire data value chains and markets - cross-sector, cross-border, cross-language**

1. Innovation actions: promote open data value chains and data markets, involve SMEs, technology transfer and data exchange/reuse in cross-sector/cross-border settings, network of European skills centres for big data analytics technologies and business development

2. Research actions: novel data structures, algorithms, architectures, optimization and language understanding technologies etc. for analysis and visualisation tasks on extremely large and diverse data streams; competitions/prize schemes around specific big data challenges (e.g. prediction, deep analysis) arising from **key industrial domains** (e.g. geo-spatial, energy, finance, health, skills/employment, agriculture, climate/weather, product recommendations...)

Implies: strong attention to **application areas:** building the (sub)communities and links to sector-specific Societal Challenges



Research

1. Collaborative **projects** addressing analysis, prediction, visualization on extremely large, diverse (multilingual, structured/unstructured) data
2. Collaborative **projects** to set up benchmarking and evaluation settings for big data analysis and prediction
3. Prize schemes to stimulate excellence in deep analysis and prediction on Big Data

Innovation

1. Open Data reuse **incubator** for SMEs: spin off mini projects building experiments and proofs-of-concept for business models and value-adding chains based on reuse of Open Data.
2. Collaborative **projects** on cross-sectoral, cross-border, cross-lingual analytics solutions/services, technology transfer, market validation, clear business cases.
3. Horizontal actions:
 - PPP preparation, cross-program coordination on Big Data
 - Network of skills centres, curricula, training and education
 - Networking, clustering, legal issues etc.



Problem: European Digital Single Market is fragmented by language barriers. Current (e.g. Google) machine translation solutions fall short in quality and coverage (languages, text types, topics) and are not customizable. Lack of **cross-lingual technology** equally hampers progress in multi- and cross-lingual analytics

Solution: explore new avenues, methods, approaches to achieve *significant improvement in translation quality* in fully automatic MT. Special emphasis on: all (difficult, small) EU languages as **target language**. Self-learning/self-improving systems, making best use of available data and language resources. Special focus on the EU languages "facing digital extinction".

Implications: close collaboration and clustering with other actions supporting **language resources infrastructure** (META, Connecting Europe Facility, national programs, structural funds)



Approach:

explore new avenues, methods, approaches to achieve *significant improvement in translation quality* in MT.

Self-learning/self-improving autonomous, fully automatic systems, making best use of available data and language resources.

Emphasis on: all (difficult, small) EU languages as **target language**.
Special focus on the EU languages "facing digital extinction".

Implications:

close collaboration and clustering among all actions supporting a **language infrastructure** (META, CLARIN-ERIC, Connecting Europe Facility, national programs, structural funds...)



- One deep and broad research project
 - kick off a multidisciplinary research action
 - focus on points where current systems fail (adaptation, quality, need of large corpora...)
 - break the glass ceiling of quality improvement
- A few advanced pilots – the "intertwined innovation" element
 - test, validate, evaluate quality improvement in realistic use situations, e.g. online services
 - address "poorly served" languages
 - connect, contribute & make use of platforms and infrastructures for language resources, open data...
- One Coordination Action: promote a common language resources infrastructure for MT, benchmarking, best practices evaluation, interoperability, metadata harmonisation...



Rationale: while systems and devices are becoming more and more powerful, the interface to humans is lagging behind

Objective: achieve transparency and invisibility of technology – effortless, effective human-machine **dialogue**, easy use of complex and powerful systems, easy **access to information**

Links to: creative industries, communication technologies, language (especially: speech) processing, cognitive & behavioural analysis

Uses: search, information retrieval, elderly, people with special needs, designers/artists...



Rationale: while systems and devices are becoming more and more powerful, the interface to humans is lagging behind

Objective: achieve transparency and invisibility of technology – effortless, effective human-machine **dialogue**, easy use of complex and powerful systems, easy **access to information**

Links to: creative industries, communication technologies, language (especially: speech) processing, cognitive & behavioural analysis

Uses: search, information retrieval, elderly, people with special needs, designers/artists...



Connecting Europe Facility (CEF)



What is CEF?

- A funding programme for **infrastructures** and **deployment** of digital services
- procurement and deployment of **mature** technology to build a "core platform"
- grants for "generic services" building on and linking to core platform

What CEF is not?

- research or innovation (Horizon 2020 is for that)



Rationale

- Automated Translation (AT) is a "building block"
- AT will serve the other Digital Services Infrastructures in CEF
- AT = whatever it takes to make DSIs actually multilingual

Features

- Adaptable machine translation and relevant Language Resources are central
- Other likely key areas: CAT, CMS, terminology, semantic interoperability, interfaces to various systems and data types
- Human element is essential: service provision, quality control, validation, post-editing, on-demand response...

Instruments: mostly procurement (calls for tender)

National dimension: CEF implies and encourages partnerships with member states and regions, e.g. use of structural funds for language resources and, technology and translation on a "local" basis



Language industry

- providers of language technology, especially MT
- language service providers

Language competence centres: provision of language resources, tools, validation and evaluation

Member states/regions: coordinating local/national/regional programs and initiatives with CEF to reach critical mass



In the current programme for **Structural funds**, 19 BEUR was available for ICT projects, but only 15 BEUR was used.

We need better strategic planning, communication and coordination for the use of structural funds

Where are the LT/LR programs funded from structural funds???

Structural funds are the right instrument to support individual (national, regional) languages!

I am aware of very few examples so far:

- "Lithuanian language in the information society": 10 MEUR funding for machine translation, language resources, multilingual digital services, mostly from EU structural funds.
- An MT/LR program is being started in Latvia.

How can we replicate this in other Member States?



The Smart Specialization Platform (S3 platform):

Toolbox and information resource for Member States (and other players) on the opportunities to support Digital Agenda technologies from structural funds – together, in a coordinated fashion

<http://s3platform.jrc.ec.europa.eu>

The future of language-specific LT/LR actions will largely depend on how these funding opportunities can be seized.



Thank you!

Kimmo.Rossi@ec.europa.eu

<http://cordis.europa.eu/info-management>

<http://cordis.europa.eu/fp7/ict/language-technologies/>