



Kimmo Rossi

European Commission

CNECT.G.3

Warsaw, 27 September 2012

last series of calls under the current programme:

- Challenge 4 – *Digital Content & Languages* of the ICT work programme 2012-2013, released in July
 - Call-10: open Jul 12; close Jan 13; budget: 27 M
 - SME call: same schedule; STREP budget: 15 M (for data + language)
- **Call-10**, R&D, objective 4.1:
 - a) cross-media **content analytics**
 - b) high-quality **machine translation**
 - c) speech-enabled **multimodal interaction**
- **SME-Call**, innovation, objective 4.3:
 - b) **industrialisation** of promising technologies, first-use validation, business models & exploitation paths

Call 10 - 4.1 Content analytics & language technologies

- **deadline for proposals 15 Jan 2013 at 17.00 Brussels time**
- **thematic continuity** with respect to Call 7:
- **"bridge"** to Horizon 2020 (2014+): higher ambition, cross-disciplinarity

- **as with previous calls, a few common features:**
 - written and/or spoken language, as required
 - multi-lingual (i.e. multiple in/out languages), where relevant cross-lingual (“translation”)
 - handle everyday language, social media & user-generated content
 - cope with massive volumes & diverse sources
 - cater for contextualisation & personalisation
 - technologies are generic & adaptive (language, domain, task)
 - but testing within specific application environments
- **new emphasis on**
 - *go beyond pure text or speech:*
 - 1) **multi-media** content & **multi-modal** interaction, and therefore
 - 2) **multi-disciplinary approaches & partnerships**

3 research lines (“outcomes”):

- a. (cross-media) content analytics
 - b. (high-quality) machine translation
 - c. (natural) spoken & multimodal interaction
- no cross-over, choose **one** of the above!
 - **for each line** there is room for:
 - **1-3 STREP projects** (75% funding) investigating new approaches & research avenues beyond the current SoA
 - up to **1 support action** (100%) laying the foundations of coordinated efforts under H2020 (eg project clusters)
- each line provides opportunities for
- **ambitious** forward-looking efforts, as well as
 - active co-operation with **progressive** vendors & users

projects: a) analytics

- some features:
 - exploring, interpreting, 'understanding'... rich, unstructured content: interplay of text, speech, audio & video
 - exploiting diverse, multilingual sources
 - delivering significantly better-than-SoA accuracy across languages & media mixes (EN text being the benchmark)
 - coupling accuracy with efficiency & broad coverage
 - based upon linguistics, semantics, statistics & whatever else it takes...!

features: b) translation

- significant improvement over today's MT *quality*
- not just re-combination of existing methods & techniques
- more than another MOSES-based platform with no substantial innovation
- no more post-edited MT, CAT & MT

features: c) interaction

- clear *grounding* in existing or emerging ICT platforms
- a prominent *dialogue* element, multiple modalities
- emphasis is NOT on speech recognition, transcription, signal processing

roadmapping actions (*CSA projects*)

- *sectorial* (see a-b-c)
- centred on needs & application scenarios, core technologies & research tasks, trials... defined & agreed with *industry*
- based upon common tools, data sets & standards; agreed reference architecture(s)
- common integration/evaluation methods & facilities...
- ... yielding strong alliances & credible plans

budget breakdown:

- projects (STREP, 75%): no less than 21 M
- roadmapping actions (CSA, 100%): no more than 6 M
- no predefined allocation between the 3 project lines

typical scope/size:

- projects: up to 3 years, up to 3 M funding
- support actions: up to 2 years, up to 2 M funding
- strategic involvement of vendors & users a definite asset
- 'multiple' languages: it depends, 3 is a minimum
- size of the consortium: as big (as small) as it takes

SME Call - 4.3 SME initiative on analytics

- **3 project lines:**
 - a. Open Data Reuse Incubator
 - b. *Easing Transfer and Take-up of LT*
 - c. Software and Applications based upon Reuse of Open Data
- **key parameters** (different from Y2011 call):
 - up to 2 years, up to 1.5 M funding
 - 2+ SMEs, > 30% of the funding
 - 1-step evaluation, 20 pages maximum

4.3.b Easing transfer and take-up of LT

... bringing ["intelligence" related] language **technologies** closer to **commercial maturity** through an **industrialisation process** including:

- (i) *engineering* of promising but commercially untested technologies
- (ii) *integration* within existing or upcoming products and services;
- (iii) *first-use experimentation* and validation;
- (iv) *in-depth assessment* along technical, user related and economic dimensions;
- (v) *identification* of possible exploitation paths, viable business models, suitable sources of funding.

Pre-proposal service:

- free assessment of your proposal idea
- non-committing, does not prejudice evaluation
- send to CNECT-G3@ec.europa.eu
- compact: 1-2 pages
- closes 4 weeks before call deadline, but if you want *useful* feedback: send no later than 1 Dec.

new programmes 2014-2020

- research (H2020)
 - innovation (H2020)
 - infrastructure (CEF)

The new programmes (2014-2020)

Horizon 2020 (H2020) - research & innovation

Connecting Europe Facility (CEF) - pan-European service platforms/service infrastructure

Both have the status of legislative proposal, adopted by the European Commission and addressed to the "legislators" (the Council and EP)

H2020 (2014-2020):

- **budget (EC proposal): 80 B**, ~20% for ICT
- **research & innovation**
 - excellent science (ERC, FET, research infra)
 - *industrial leadership* (most mainstream ICT themes)
 - societal challenges (e.g. health, ageing, education...)
- **access to risk finance**
 - consists of Debt Instrument and Equity Instrument
- **innovation in SMEs: dedicated SME instrument**
 - funding & financing early stage high risk research and innovation, stimulating break-through innovation, commercialisation of research results

H2020 ICT part

EC proposal sets out 6 activity lines, expected to cover the full range of needs:

- new generation of components & systems
- next generation computing
- future internet
- **content technologies and information management**
- advanced interfaces and robotics
- micro- and nanoelectronics, photonics

H2020 ICT part – Content technologies and information management

Proposed objectives and content:

- provide new tools to create, exploit and preserve content **in any language**
- provide new tools to model, analyse and visualise vast amounts of **data**
- new technologies for **language**, learning, interaction, preservation, content access and analytics
- IIM systems based on **advanced data mining, ML, statistical analysis** and visual computing

H2020 What happens now & next

- **on-going discussions with Council & EP**
 - outcome depends on concomitant discussion of *MFF* (= money for all EU policies)
 - so far, support for H2020 & ICT
- **scope of individual ICT themes is not decided**
 - stakeholder consultations will be launched after summer 2012
- **early orientations by Q1 2013**
 - first series of calls towards Q4 2013

CEF (2014-2020):

- **TEN model:** pan-European, public-interest infrastructures
- 3 pillars: Transport, Energy & ICT (new, 9 B proposed)
- for ICT: (i) Broadband & (ii) ***Digital Service Platforms***
 - eGovernment services, eHealth, eCulture...
 - **Multilingual Access to online Services** (new)
 - Open Data Portal (data.eu)
- discussions on-going with Member States (Council working parties) and European Parliament, update in the autumn
- discussions have been positive and constructive; main challenge is effective communication, concrete illustration of needs, benefits, features, implementation...

CEF Multilingual:

"... enable providers of online services to offer their content & services in as many EU languages as possible, in a most cost effective way."

addressing:

- **public services** esp. eGovernment, eHealth, eCulture, Open Data portals
- **commercial services** eg eCommerce
- where appropriate, **user-generated content** (social media)

- less than 40% of Europeans (outside UK, IE) understand English, less than 10% *accept to use* online services in English
- multilingual services are necessary to:
 - unify the EU online market – from 20+ “sub-markets” to one single market of 500+ million consumers
 - enable the provision of public services across national borders
 - unleash the power of global online communities

Thank you!

cordis.europa.eu/fp7/ict/language-technologies/

CNECT-G3@ec.europa.eu

kimmo.rossi@ec.europa.eu