



**Roberto Cencioni**

**European Commission**  
**DG Information Society & Media**

*Digital Content & Cognitive Systems*  
*Language Technologies*

*Brussels, 19-21 June 2012*



## EC has supported Language Technologies since 1980s

- a fresh start since 2008
  - renewed political commitment
  - explosion of online content, esp. social networks
  - sizeable industry of language services
  - promising technical advances
- approx **150 M funding** in the last 5 years
- nearly **60 projects** underway
- next *call for proposals* in July



## main features of our projects

- multi-party, multi-nation partnerships
  - member states, associated countries, 3<sup>rd</sup> countries
- academia + industry + "launching" users
- focus on enabling technologies
- multiple languages
- impact
  - target groups, application scenarios, demonstrators
- significant involvement of commercial players



## main themes of our projects

*"... effective solutions that support business and inter-personal communication and enable people to make sense of online content and services in Europe's many languages."*

- global content processing, eg automated translation
- mining of unstructured information, eg text analytics
- natural interaction, eg speech understanding & dialogue

in and between multiple languages



# Upcoming calls for proposals

## *ICT Work programme 2012-2013*

- Challenge 4 “Technologies for Digital Content & Languages”
- 2 objectives, 2 calls both launched in July:

**Call 10 - 4.1:**                      open July 2012  
   close Jan 2013  
   budget: 27 M  
   - dedicated to language

**SME Call - 4.3:**                      same schedule  
   budget: 20 M  
   - both content (data) & language

- approx. 12 LT submissions retained for funding



## Call 10 - 4.1 Content analytics & language technologies

- **new title** to reflect a broader, multi-disciplinary approach
- **continuity** in terms of project lines:
  - *content analytics*  
aka information extraction, mining of unstructured content, categorisation & summarisation, social media mining & sentiment analysis, business intelligence...
  - *machine translation*
  - *spoken & multimodal interaction*
- **discontinuity** in terms of ambition & timeframe  
- paving the way for work under H2020 (2014+)



- **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 & UGC
- 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**
- *progress across languages & media*



### 3 research lines (“outcomes”):

- a. (cross-media) content analytics
  - b. (high-quality) machine translation
  - c. (natural) spoken & multimodal interaction
- no cross-over, stay within the line you’ve chosen
  - **for each line** there is room for:
    - **1-3 projects** (75% funding) investigating new approaches & research avenues beyond the current SoA
    - no more than **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...!
- what we don't need:
  - just "*yet another project*" with no significant S&T innovation



## projects: b) translation

- what we don't need:
  - projects delivering no significant improvement over today's MT quality
  - a mere re-combination of existing methods & techniques
  - another MOSES-based project with no substantial innovation
  - things others can do: post-edited MT, CAT & MT, crowd sourcing...

## projects: c) interaction

- what we don't need:
  - research with no clear *grounding* in existing or emerging ICT platforms
  - speech recognition, transcription, translation, acoustics... without
  - ... a prominent *dialogue* component
  - ... the interplay of speech & other *modalities*



## **roadmapping actions** (*blueprint for future R&I*)

- *sectorial* (see a-b-c), not for the LT community as a whole
- centred on needs, scenarios, tasks... defined & agreed with industry
- based upon common tools, data sets & standards; agreed reference architecture(s)
- joint integration & evaluation facilities ...
- ... yielding credible plans, efforts & timescales

## **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



## as with previous calls:

- single entry-point (Mr Kimmo Rossi)
- help-desk for inquiries & outline proposals
- info-days ...
  - *further details in due time*
- candidate experts/evaluators are welcome



## **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.



***Thank you!***

**FP7 ICT programme:**

[http://cordis.europa.eu/fp7/ict/programme/home\\_en.html](http://cordis.europa.eu/fp7/ict/programme/home_en.html)

**Horizon 2020:**

[http://ec.europa.eu/research/horizon2020/index\\_en.cfm](http://ec.europa.eu/research/horizon2020/index_en.cfm)