



**Kimmo Rossi**

**Obj. 4.1 Content analytics and  
language technologies**

***CNECT.G.3***

*Warsaw, 27 September 2012*

## **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: 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



## Pre-proposal service:

- free assessment of your proposal idea
- non-committing, does not prejudice evaluation
- send to [CNECT-G3@ec.europa.eu](mailto: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.

***Thank you!***

*[cordis.europa.eu/fp7/ict/language-technologies/](https://cordis.europa.eu/fp7/ict/language-technologies/)*

*[CNECT-G3@ec.europa.eu](mailto:CNECT-G3@ec.europa.eu)*

*[kimmo.rossi@ec.europa.eu](mailto:kimmo.rossi@ec.europa.eu)*