



Contribution ID: 11

Type: **not specified**

Big Data from Crowdsourcing in the Humanities: Potentials and Challenge

Tuesday 24 September 2013 14:30 (30 minutes)

Mobile devices, enabling easy interaction and providing advanced sensors and audio and video recording for virtually everybody, have a great potential to revolutionize science –in particular, they will raise crowdsourcing to a new level, providing “big” data streams for the humanities.

App-based crowd-sourcing has a potentially transforming effect in several phases in the research process:

- * Data creation, elicitation and collection: Many more subjects can be easily actively involved, overcoming shortcomings such as unbalanced or lacking participation in questionnaires

- * Data storing, pre-processing, and management: Advanced back-end-infrastructures shall be built to receive and distribute data, pre-process it (assess, select, curate), and organize it for later archiving and analysis, dealing with data magnitudes bigger than what we have currently

- * Data “processing” (compile, enrichment, annotation): Overcoming the bottleneck of annotation by massive contribution from the crowd or citizen-specialists

- * Dissemination of results, evaluation: Data-sets can be made easily accessible, accompanying publications or as published results in their own right, with visualization app technology

Several technological and societal challenges have to be addressed to make the most of the potential of the big data from crowd-sourcing:

- * How to engage the users: Making them aware of the app, and making them engage in using it

- * How to receive, manage, pre-process, and store the data sustainably: Setting up an appropriate central back-end infrastructure; adding metadata; automatically pre-processing; storing large amounts of dynamic data; distributing stimuli and resulting datasets

- * Provenance information and quality assessment: It is crucial to have data quality assessment and information is who was the user, when and where the data were generated, on the basis of which stimulus or in which context, etc.

- * How to curate the data: A strategy for data curation has to be included in the plans for an app infrastructure and workflow right from the start

- * Privacy, intellectual property, authorship, access restrictions: Protecting the privacy of contributors or curators and give the appropriate credits for contributions

- * Life-cycle: policy-based handling and de-commitment: Dealing with data sets in a systematic manner requires policy-based automated treatment possibly including for de-commitment.

Generally, successful employment of crowd-sourcing and in particular apps for mobile devices depends on much more than a well-designed and programmed app. This implies longer development time and costs that have to be taken into consideration. In particular, the back-bone infrastructure needs careful planning and installation in order to be prepared to deal adequately with the incoming data streams and sets. This seems to be a task for data centres which have already some experience with the handling of complex and large sets of digital data.

Primary author: Dr DRUDE, Sebastian (MPI for Psycholinguistics)

Presenter: Dr DRUDE, Sebastian (MPI for Psycholinguistics)