Presenter / Contributor name: Spiros Mouzakitis

Type of the presentation proposed: in-use contribution

Title of the presentation: ENGAGE e-infrastructure platform

Summary of the presentation (100 words):

ENGAGE (http:www.engagedata.eu) is a collaborative space for researchers, data journalists, enterprises and citizens with a view to extracting, highlighting and enhancing the re-use value of PSI datasets through crowd-sourcing. This presentation will present the platform and its re-usable components, current experience gained through its operation, as well as the vision of the ENGAGE service.

Extended abstract of the presentation (up to 4 pages):

ENGAGE (http://www.engagedata.eu) is space for researchers, data journalists, citizens and other potential end users, who rely on open public data for professional or personal use. ENGAGE is a centralized and collaborative e-Infrastructure providing a single point of access to distributed and diverse PSI datasets, as well as relevant tools in order to cover the needs of researchers and citizens. PSI can be anything varying from election results, population statistics, unemployment figures, earnings and transportation to fire incidents, criminal records and illegal immigration.

The vision of the ENGAGE e-Infrastructure is to extract, highlight and enhance the re-use value of PSI datasets. This is achieved by moving from low-structured, non-machine process able, isolated and difficult to find data (e.g. PDF) to high-structured, easy to process and linked datasets (e.g. CSV, RDF).

And for this to be achieved, ENGAGE empowers its user community to find, curate, cleanse, enrich and extend PSI datasets and their metadata in a collaborative, crowd-sourcing approach.

The envisaged impact of ENGAGE is to contribute towards removing the constraints of access and usability as well as the barriers between disciplines and borders, to enable more effective scientific collaboration around PSI. Through social collaboration, Public Sector data will move towards more standardized, effective and semantically enriched structures that will allow PSI re-use to reach its full potential.

The main services of the ENGAGE platform to be presented are:

- Faceted, elastic search through datasets, which may have been harvested by various open data repositories, using keywords or filters on the meta-data
- Internationalization and Localization features supporting various languages (including Farsi, Chinese traditional, Indonesian, Hebrew)

- Automatic translation of Search results, Metadata Values and actual values of parsed datasets (resource files) that can be manually and collaborative refined through Microsoft Translator hub.
- Implementation of dataset extension workflow and graphical representation of the evolution of extended datasets. The current extension types include the following: "Conversion to other formats, Data Enrichment, Metadata enrichment, Snapshots of real-time data, Mashups"
- Implementation of User Groups for the collaborative management and maintenance of datasets
- A dedicated section for users to request datasets from the community members under "Requests Section" instantiating a Yahoo Answers-like functionality.
- Improved community-based Data Rating system with view of the distribution of votes, quality metrics and individual ratings
- Linking Datasets with derived items (including Applications, Web Services, Reports, Papers, Journal)
- K-Means clustering module for open and big data analytics
- ENGAGE Plug-in for Open Refine
- Integration of the core visualization capabilities for tabular datasets including grid view, chart view, and map view.
- Integration with external visualization service, providing users with the ability to create arbitrary visualizations using the best-in-class visualization engines, such as IBM's ManyEyes and embedding live visualizations as related items along ENGAGE datasets.
- Online pre-viewers for various types of documents (including live CSV viewer)
- RESTFUL ENGAGE API for all entities of the ENGAGE ecosystem and SPARQL Interface (Linked data)
- Sign on using Social Media accounts (Twitter, Facebook, Google+, Linkedin)

The presentation will conclude with lessons learned so far, as well as an outlook for the Open and Big Data:

- The value of community in harmonizing data (researchers, data journalists, enterprises, citizens IT experts)
- Curation and trust through crowdsourcing for improving the RE-USE value of datasets.
- Digitization of public documents through crowdsourcing: a win-win situation