FEW Nexus – a Data Collection and Integration Challenge

Dieter Pfoser, George Mason University, dpfoser@gmu.edu

1. Overview
The FEW nexus concept recognizes that the provision and consumption of food, energy and water are inextricably interlinked. Hence the three concepts have to be studied jointly and the workshop goal is “understanding, appreciating and visualizing the interconnections and interdependencies in the FEW system of systems at local, regional and global levels.” This statement identifies the following data science challenges.

2. Data integration
Although a considerable amount of data has been collected with respect to all three aspects of the FEW systems, “differences in data collection protocols, and data representation standards” create a data integration challenge. Consequently, data analysis is fragmented and cannot generate conclusive results.

The key concepts that could facilitate interoperability and a shared understanding of the data are ontologies to denote the meaning and structure of concepts and Linked Data, which represents a method to publish structured data so that it can be interlinked and become more useful to a greater user group. Linked Data\(^1\) builds upon standard Web technologies such as HTTP, RDF and URIs to share data (rather than Web pages). This enables data from different sources to be connected and queried using standard WWW technology, i.e., no central repository is needed to integrate datasets. Linked Data allows one to share data models (Ontologies) as well as the data itself.

3. Data collection, Crowdsourcing
Providing FEW Nexus-level data analysis requires “a better understanding of the problem, the connections, and the impacts by generating, using, and openly sharing improved sources of data.” I.e., in many cases, we do not have the data we actually need to “monitor a variety of Earth resources and trends.”

With the Internet becoming a medium for data publishing and information exchange, we have seen an explosion in the amount of online content available on the Web. When channeled properly, such user-generated content can be used to complement authoritative data sources provided by governments and NGOs. Consider here for example Openstreetmap, a community mapping effort in which contributors typically add their local knowledge to a global dataset. A similar approach could be envisioned for collecting FEW related data. On the ground observations provided by locals could complement authoritative datasets in real-time. The feasibility of such an approach has been demonstrated in crisis response scenarios using mobile phones and dedicated collection platforms such as Ushahidi\(^2\). It has been shown that user-generated content does not require sophisticated platforms and networks, but simple text messages suffice. Besides dedicated collection efforts, implicit content reflects derived information as the original, user-generated content was produced with a different purpose in mind. Implicit content is also often embedded in social media contributions (e.g., blogs, micro-blogs, social multimedia). Using various data mining approaches, implicit content can be a valuable “sensor data” resource providing detailed observations in real time.

\(^1\)http://linkeddata.org
\(^2\)http://www.ushahidi.com