A relaxed approach to handling semantics in data management.

Paolo Diviacco, Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS) (Italy), pdiviacco@ogs.trieste.it
Alessandro Busato, OGS (Italy), alessandro.busato@ogs.trieste.it
Peter Fox, Rensselaer Polytechnic Institute - RPI (USA), pfox@cs.rpi.edu

A very important aspect of geophysical data management is how to handle heterogeneity of the data. This can stem from different vintages, from differences in acquisition practices, resolution, data type, format and media that result in many issues in accommodating such heterogeneity within a single framework. In addition, the same dataset can be used in different contexts, with different research objectives. Data discovery can then be complex because data "tagging" is generally based upon data usage. If we "cage" data into preset possible cases, every time we have a situation or a need that does not match the pre-cooked cases, it will be rather difficult to hit what we are looking for.

Metadata is of course an important aspect of data discovery, but can "portray" only some of the many end user needs. It is necessary then to add to the traditional practices, other means to get to useful data.

Data, Information and Knowledge are the different layers that structure the environment in which each research takes place. Data management is traditionally relegated to the lower layer, under the assumption that the higher layers do not influence it. This, we think, is a gross mistake.

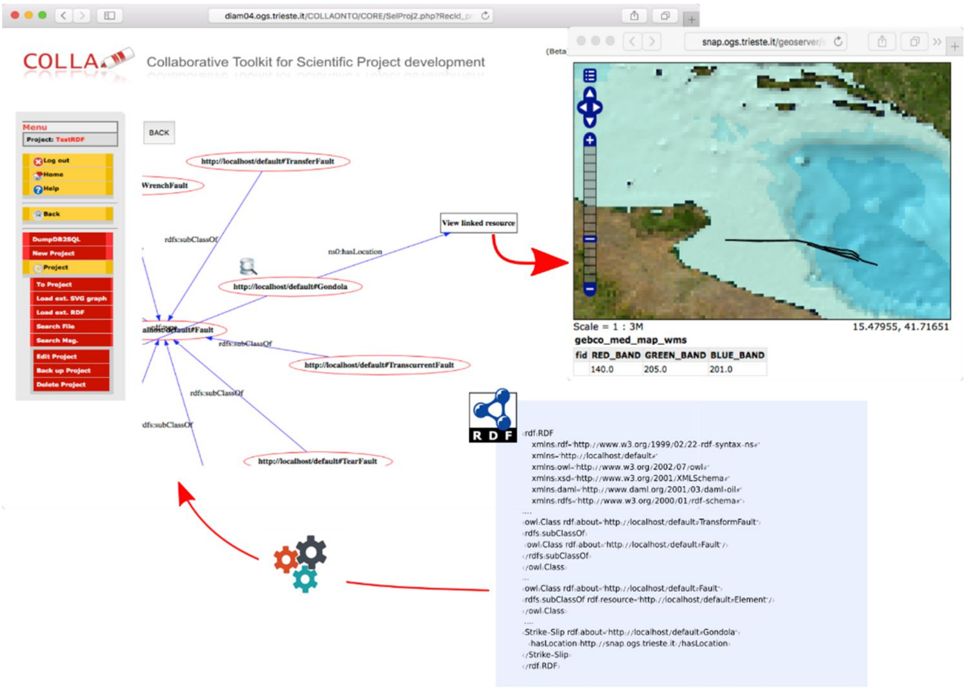
Information introduces meaning into the picture of basic measurement, and knowledge introduces context and cultural issues. A large literature in modern epistemology and sociology of science (see for example: Hanson, 1958; Duhem and Wiener, 1954; Putnam, 1975) informs us that measurements and observation are largely conditioned by theories; experiments conform to previous knowledge, and perception itself is conditioned by cognitive models. The aforementioned conditions must all be considered when providing data management and in particular, data search tools. Consequently, tools like controlled vocabularies and ontologies are filling the gap described above, introducing a semantic level where meaning is encoded. At the same time these very tools can introduce side effects. In fact, since semantic tools refer to a model "portraying" reality they need an explicit formalization of it. Considering that web ontologies are based on a denotative approach, a change in the structure of the model may inevitably lead to a change in the set of results.

Can formalization of a domain vocabulary be achieved? Some authors in the line of view of Polanyi (1966) are rather pessimistic on this.

We propose a more "relaxed" approach that integrates, metadata, visualization, representation, and semantic tools that together can help the end user to establish a path to get to what he or she is looking for.

This is based on the introduction of graphic representations of the domain of knowledge in which data and information are to be used. These can be thought as maps that, following Suchman (1987), show the possible ways to travel from one place (here concepts) to another without conditioning users in their choice. Besides, even if referring to the same data, multiple and concurrent representations (maps) can be designed for the same or different domains of knowledge. Tools have been developed to produce such maps and to organize data and information in them. One implementation we found very promising is based on the use of RDF files, that easilly can introduce in the maps the use of linked data.

Prototypes of the system have already been developed and are currently under study, while first tests show encouraging results.



fid	RED_BAND	GREEN_BAND	BLUE_BAND
	140.0	205.0	201.0

```
<rdf:RDF
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns="http://localhost/default#"
   xmlns:owl="http://www.w3.org/2002/07/owl#"
   xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
   xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
....
<owl:Class rdf:about="http://localhost/default#TransformFault">
<rdfs:subClassOf>
 <owl:Class rdf:about="http://localhost/default#Fault"/>
</rdfs:subClassOf>
</owl:Class>
...
<owl:Class rdf:about="http://localhost/default#Fault">
<rdfs:subClassOf rdf:resource="http://localhost/default#Element"/>
</owl:Class>
 ....
<Strike-Slip rdf:about="http://localhost/default#Gondola">
  <hasLocation>http://snap.ogs.trieste.it</hasLocation>
</Strike-Slip>
</rdf:RDF>
```

References

Duhem, P., & Wiener, P. P. (1954). La Théorie Physique: Son Objet et sa Structure [The Aim and Structure of Physical Theory]. Princeton University Press.

Hanson, N. R. (1958) Patterns of Discovery. Cambridge, UK: Cambridge University Press.

Polanyi, M. (1966). The Tacit Dimension. New York: Anchor Day Books.

Putnam, H. (1975). Mind, language and reality: Philosophical papers (Vol. 2). Cambridge, UK: Cambridge University Press. doi:10.1017/ CBO9780511625251

Suchman, L. A. (1987). Plans and situated actions: The problem of human-machine communication. Cambridge, UK: Cambridge University Press.