Semantic Data Frameworks Come of Age: Semantic into the seams and then on to pragmatics (use).

Peter Fox, Tetherless World Constellation (RPI) and the SeSF team and collaborators.
pfox@cs.rpi.edu

Introduction

In the last 3-4 years the application of semantic web methodologies and technologies has been demonstrated to advance a number of geoscience discipline data needs. The blockage to widespread adoption is the myth that semantics in science settings are unproven and immature. Much evidence to the contrary exists and is reinforced by government and commercial sector advancements as well. This whitepaper indicates the opportunities for geosciences in capitalising on advances in semantic data frameworks.

Progress

In the last ~ 5 years, there has been a move toward framework development (in contrast to systems). A loose working definition of each is:

- Systems have very well-define entry and exit points. A user tends to know when they are using one. Options for extensions are limited and usually require engineering.
- Frameworks have many entry and use points. A user often does not know when they are using one. Extension points are part of the design.

Frameworks are coming of age because: methods of designing and developing them have matured, as has the required technology, and there are now many adopters. Data Frameworks respond to the need to account for technology change, flexible use of standards and changing and diverse vocabularies and data storage and exchange formats. Additionally, much of this 'openness' or open-world approach rests on the adoption of the Web and the Internet as a primary and ever increasing means of data exchange for geosciences (and other disciplines). Here too the Semantic Web is a common framework allowing data to be found, shared and used across discipline and application boundaries, again resting on open standards of such as [URI]s, [RDF], [SPARQL] and [OWL], as well as maturing tools and applications such as triple stores, and reasoners.

The important maturity is in the availability of a broad spectrum of semantic encoding, ranging from the simplest 'linked data' [LD] to highly developed ontologies such as those applied in projects such as the Virtual Solar-Terrestrial Observatory [VSTO1, VSTO2]. However there are now many examples in between such as the Earth System Grid [ESG] supporting the IPCC climate model output community and emerging needs in knowledge discovery for climate [CKD]. Numerous other examples exist [4th] supporting the new paradigm of data science.

What has emerged from these efforts is that a schematic such as that in Fig. 1 (developed for the Semantic eScience Framework [SeSF] project) is now within reach of broad science communities.

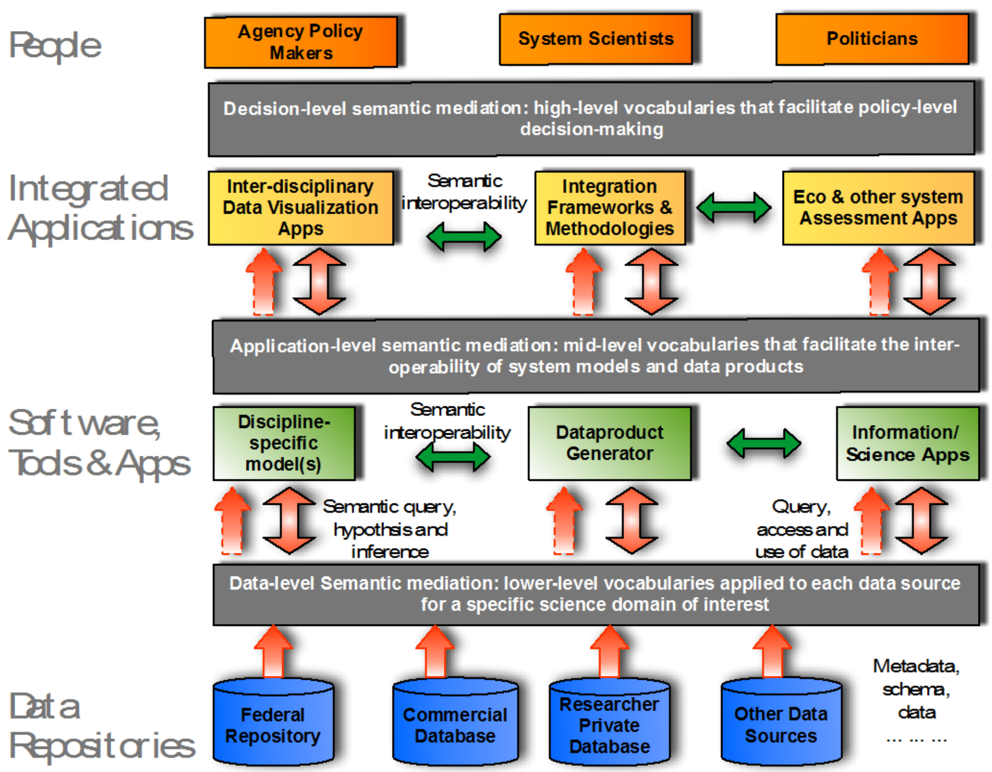

Figure 1. Generalized semantic data framework layered architecture. Grey rectangles highlight the interface use of semantics, beginning with data-level (common), moving to application-level (new) and decision-level (talked about a lot but not really implemented much).

To fully capitalize on the 'use' approach, we adopt the informatics use case development methodology (developed out of the VSTO project and refined at RPI in the Tetherless World Constellation). Figure 2 shows a depiction of that methodology. It is not the purpose of this document to fully explain the detail therein but only to emphasize that determining the level of semantic representation and suitable technology/ implementation choices are a fundamental part and outcome of the method.

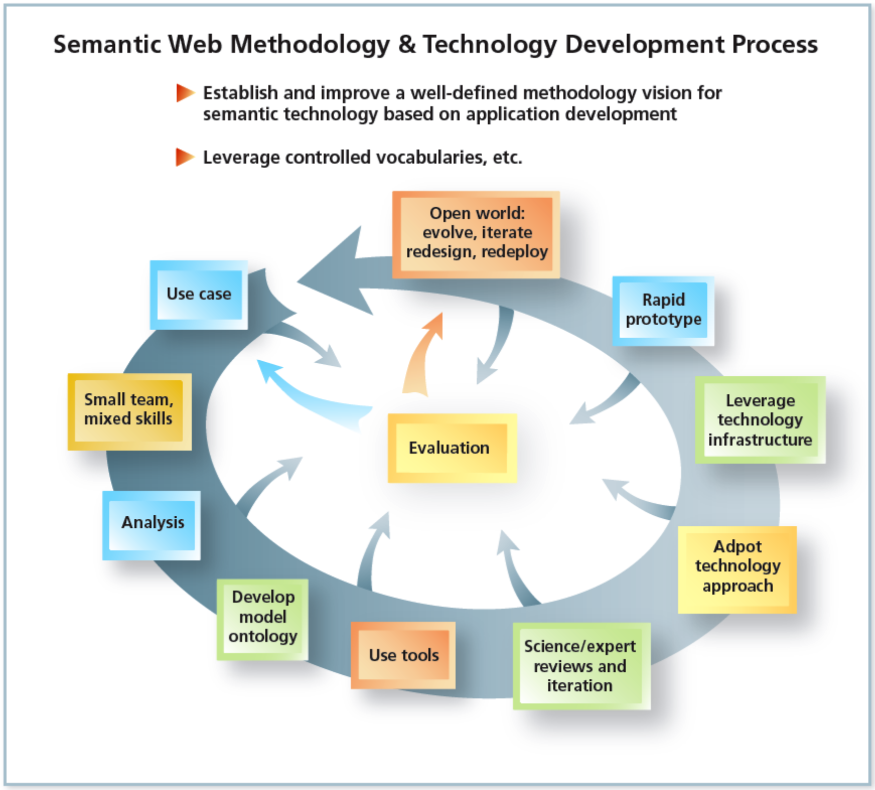

Figure 2. RPI/TWC Informatics Development Methodology.

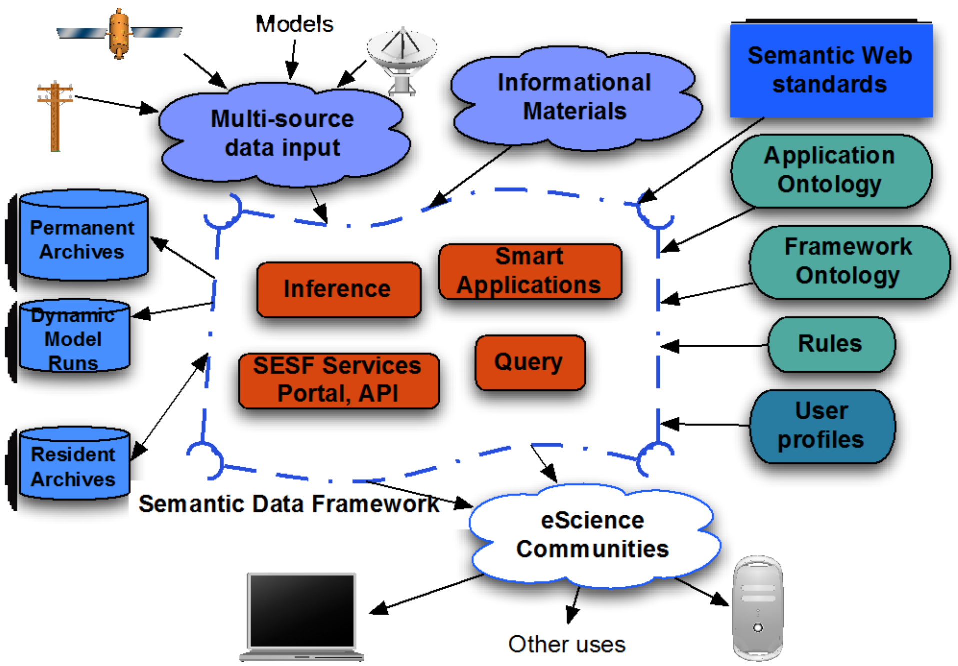

Figure 3. Schematic of a Semantic Data Framework and its relation to what connects to it: data, information, semantics, applications, services, etc. in service to eScience communities.

Finally, Figure 3 shows a schematic of the relation between key resources needed to achieve geoscience results and an example semantic data framework (in this case, [SeSF]).

Outlook

Significant opportunities now exist to design and implement evolvable semantic data framework infrastructures for geosciences. Many are in place or are ready. The broadest possible community must be served by them. Aligning the frameworks with the web and internet standards and protocols, and allowing implementation and technology evolution to be a core part of the capability, are key design elements for future e-infrastructures. The level of semantics to be adopted (or developed) is a significantly easier choice or task, and the languages and tools are increasingly mature. Go for it – see the Call to Action in [4th].

References

[4th] Fox, P., Hendler, J. 2009, Semantic eScience: Encoding Meaning in Next-Generation Digitally Enhanced Science, in The Fourth Paradigm: Data Intensive Scientific Discovery, Eds. Tony Hey, Stewart Tansley and Kristin Tolle, Microsoft External Research, pp. 145-150.

[CKD] Climate Knowledge Discovery Workshop Report, https://verc.enes.org/collaboration/attachments/198/Ludwig.pdf

[ESG] D.N. Williams, R. Ananthakrishnan, et al., The Earth System Grid: Enabling Access to Multi-Model Climate Simulation Data. In: *Bulletin of the American Meteorological Society*, 90(2): 195-205, 2009.

[LOD] E.g. see Linked Open Government Data at http://logd.tw.rpi.edu

[OWL] D.L. McGuinness, F. Van Harmelen, OWL web ontology language overview, W3C recommendation, 2004.

[RDF] Resource description framework (RDF) model and syntax. W3 Consortium, http://www. w3. org/TR/WD-rdf-syntax

[SeSF] Semantic eScience Framework, http://tw.rpi.edu/web/project/SeSF

[SPARQL] SPARQL Query Language for RDF, W3C Recommendation 15 January 2008, http://www.w3.org/TR/rdf-sparql-query/

[URI], http://www.w3.org/Addressing/#background

[VSTO1] D. McGuinness, P. Fox, L. Cinquini, P. West, J. Garcia, J. L. Benedict, and D. Middleton, The Virtual Solar-Terrestrial Observatory: A Deployed Semantic Web Application Case Study for Scientific Research, *AI Magazine*, Vol. 29, 2007.

[VSTO2] P. Fox, D. McGuinness, L. Cinquini, P. West, J. Garcia, and J. Benedict, Ontology-supported scientific data frameworks: The virtual solar-terrestrial observatory experience, *Computing in Geoscience*, Vol. 35, 2009.