

This an outline of ideas for an EarthCube Research Coordination Network (RCN) proposal with deadline on March 26, 2013.

EarthCube RCN: ESN – EarthCube Semantics Network

Peter Fox, Pascal Hitzler, Krzysztof Janowicz, Tim Finin and Mark Schildhauer

Key cybertechnological solutions required for the success of EarthCube have been outlined in a series of white paper and roadmap documents. Perhaps most needed are methods for dealing with information integration, interoperability, conceptual modeling, data-model intercomparison, data publishing, provenance, intelligent search, and the like. The ever present challenge is that they have to be effective at the scale of geoscience research activities and within a very diverse landscape of providers, use cases, and myriad existing tools and infrastructures.

The technological edge – the state of the art – proven capable of meeting the aforementioned challenges for EarthCube functions lay in the wide spectrum and maturing methods and tools collectively characterized as the *Semantic Web*. This field, well known to geosciences, develops cyberinfrastructure solutions based on computer encoding of meaning, e.g. shareable vocabularies and their inter-related data and services. At large, these *semantic methods* are currently in a rapid transition to become mainstream approaches in industry and broadly in academia. They have been adopted, e.g., by schema.org,¹ a joint initiative by major search engine providers like Google and Microsoft, to enhance the next generation web search engines. They are used by providers like nytimes.com to organize their extensive data collections which power their website.² They have been employed in developing IBM's Watson system,³ Apple's Siri⁴, and Google's Knowledge Graph⁵. These advances are very relevant to EarthCube and geosciences as the Internet/Web becomes the primary "computer" of choice for current and prospective researchers (including educated citizens).

From the outset, it has been part of the strategic plan of EarthCube to leverage the state of the art in semantic methods, in order to make use of the added value provided by the most up-to-date shared technology solutions. To guide implementation needs, roadmaps for different elements of EarthCube have been developed. However, it is not as simple as stating this need. As a whole the EarthCube audience, "all of geosciences" is mostly unfamiliar and often mis-informed of the capabilities, benefits and required investments in semantic approaches to advance *their* science, on a daily basis. To fully identify and formalize the required semantics to move the overall EarthCube vision forward, this RCN formalizes the transition between the initial conceptualizations and an active, relevant, collaborative community focused on the pursuit of the realization of semantic-cybertechnological solutions for EarthCube. Based on accumulated experiences in conducting such a semantic network (of researchers and practitioners) over the last ~7 years, the RCN would operate in two primary modes:

- Outreach - to find and engage members of the geoscience communities that are in need of, ready for, and welcome semantic solutions
- Inreach - to consolidate a cadre of semantics-savvy researchers and practitioners embedded in key geoscience projects

¹http://semanticweb.com/schema-org-adds-additional-type-property_b30861

²<http://data.nytimes.com/>

³<http://www.research.ibm.com/deepqa/faq.shtml#20>

⁴<http://blog.infomedia.com/business-technology/siri-and-the-semantic-web/>

⁵[http://blogs.gartner.com/darin-stewart/2012/05/17/googles-knowledge-graph-yeah-thats-the-semantic-web-sort-](http://blogs.gartner.com/darin-stewart/2012/05/17/googles-knowledge-graph-yeah-thats-the-semantic-web-sort-of/)

of/

Operations must include close partnership with EarthCube governance and Amendment II test-bed/ building block awards. Thus, initially it is envisioned that the EarthCube Semantic Network (ESN) RCN would focus on two next step issues, namely:

- documenting an *integrated* and semantically-enabled cybertechnological approach relevant to all successful EarthCube RCNs to underpin the needed semantic interoperability identified to date in other EarthCube awards (especially governance), and
- detailing a fine-grained set of implementation and research activities, with a near (fed into specific and relevant Amendment II awards), medium and long-term timeline for the realization of semantic-technology solutions.

In the two-year period ESN outcomes will be:

- a growing community network of geoscientists, computer scientists and CI practitioner who have either benefited from adopting semantic approaches and/or are substantially more informed about the needs, benefits, opportunities and effort involved,
- the formation of collaborative teams to address and implement near-term semantic-technology solutions, and
- the clear identification of pressing research issues around semantic-technology for geosciences

To achieve this, it is paramount that the development of overarching technological solutions must be massively informed by geoscience use-case requirements, the needs of end-users (*especially early career researchers*), and organized EarthCube groups. The main efforts in ESN will thus go towards functionally formulating these requirements by gathering the community and establishing continuous liaisons with EarthCube stakeholders. ESN will reach out to thousands of potential users and contributors to EarthCube through these alignments. Work in ESN will coordinate this community outreach and channel it in form of a concrete plan for empowering EarthCube through high priority needs for semantics, especially for shareable vocabularies, data, and services. We will therefor organized the ESN work along the following three lines.

- Continuous requirements assessment involving key stakeholders in EarthCube, spanning end-users all the way to organized EarthCube networks
- Development of a complete integrated conceptual plan and timeline for supporting EarthCube through the development and maintenance of semantics, shareable vocabularies, and related services
- Education of all EarthCube stakeholders about the capabilities and limitations of using semantic methods

ESN involves a significant number of Co-Investigators and Senior Personnel to drive these three efforts.

(1) Liaisons for continuous requirements assessment There will be appointed persons to act as liaisons with both end-user and technology groups in EarthCube. We will recruit liaisons with a background in both semantic methods and the topic of their group. Requirements assessment will be continuous and also involve continuous feedback on the conceptual plan being developed by the core technology planning group (2).

(2) Core Technology Planning Group This group will develop an integrated conceptual plan and timeline for supporting EarthCube through semantics and shareable vocabularies. They will be informed by the requirements assessment made through the liaisons (1), and by the state of the art in semantic methods. While they will incorporate approaches at the cutting edge of cur-

rently available methods, they will also be rather conservative with respect to technology forecasts, and shall strongly consider willingness of adoption of state-of-the-art technologies by EarthCube stakeholders.

(3) Consulting in Semantic Methods This group of ESN participants will disseminate foundational and hands-on knowledge to EarthCube stakeholders through tutorials and consulting. Initially, they focus on basic semantic methods, while later on it will be significantly informed by the outcomes of the Core Technology Planning Group, which is in turn informed by the Liaisons. The group will also consult end-user groups and other technology groups in EarthCube in how to realize an uptake of semantic methods for shareable vocabularies, data, and services.

Project duration: 2 years

Project budget: \$300,000

Participation:

User Group: Inclusive and open, with a variety of roles and levels of engagement possible

Policy Group: are steering committee members according to the Amendment I wording