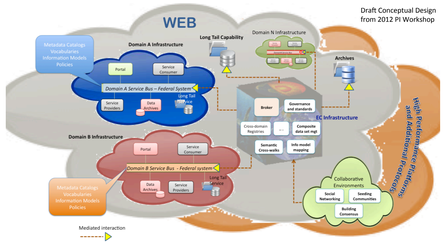
NSF EarthCube and Data Interoperability

EarthCube

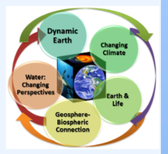
groups events documents workshops funding community my page contact



Imagine a world...with easy, unlimited access to scientific data from any field, where you can easily plot data from any source and visualize it any way you want, and you can easily model your results and explore ideas from your desktop, lab, or in the field. EarthCube aims to make this vision a reality!

What is Earthcube?

"EarthCube is a bold new NSF activity to create a data and knowledge management system for the 21st Century. The Directorate of Geosciences (GEO) and the Division of Advanced Cyberinfrastrictire (ACI) of the US National Science Foundation are the sponsors of this effort. The scientific objective is to develop a framework over the next decade to understand and predict the Earth system from the sun to the center of the Earth. NSF is facilitating a community dialog to define the research and educational framework for this new effort and with the expectation that it will transform the conduct of research and education for the geosciences and related fields with the intended outcome of accelerating progress towards the scientific objective. EarthCube has the potential to:



- Transform the conduct of data-enabled
- Create effective community-driven cyberinfrastructure
- Allow global data discovery and knowledge management
- Achieve interoperability and data integration within and across disciplines

To achieve its full potential, EarthCube must be connected to and guided by the geoscience community that are actively engaged in research that often involve collecting and using data, simulating processes, and visualizing complex interacting systems. To do this, there needs to be close interaction between the end-users and those who will create the software and make sure cyberinfrastructure is responsive to changing user needs as well as accommodating the opportunity to capture innovations that is part of the ever changing tools and landscape in computer science and computing hardware.

EarthCube is fostering a dialog among a breadth of disciplines in the geo-, bio-, and cyber-sciences to create a framework to understand and predict the Earth System and to provide an milieu that greatly increases the productive and capability of research and educator to undertake the science required to tackle the pressing problems in the geosciences and for society. As such, we have started conversations on important themes that will form part of the foundation of what EarthCube will become. These are represented by the community groups you see on this website (governance; data discovery, mining, and access; workflows; and semantics/ontologies) We have also funded a number of concept awards that will be testing out various approaches and components that may be needed to build what an EarthCube might be...

EarthCube is an outcome and a process → EarthCube is based on broad community and NSF participation

...We anticipate the structure and functionality of EarthCube will emerge over the next few years. During this time, NSF will facilitate dialog between geo- and cyber-scientists and continually seek end-user input and feedback and consensus on the way forward. Right now is your chance to help shape EarthCube and influence what it can do and what it can be. Please help us make this work for you and your science.

If you are coming to this website and want to share your thoughts, but do not know where or how you can do that, please contact the NSF EarthCube team at earthcube@nsf.gov. They will be able to help you and find a way that you can make your thoughts and ideas known. We welcome you to EarthCube and look forward to hearing from you," (US National Science Foundation, 2011, "What is EarthCube").

Why EarthCube?

- Nature does not recognize separate disciplines
- EarthCube will democratize access to data
- EarthCube will increase research time by reducing time need to find, access, and analyze data
- EarthCube will enable more interdisciplinary research and the pursuit of new questions
- EarthCube will accelerate the pace of discovery
- EarthCube will give scientists the same chance of making major contributions regardless of institution size or endowment

What are the Goals of EarthCube?

The goal of EarthCube is to transform the conduct of research by supporting the development of community-guided cyberinfrastructure to integrate data and information for knowledge management across the Geosciences. Specific goals of EarthCube are to:

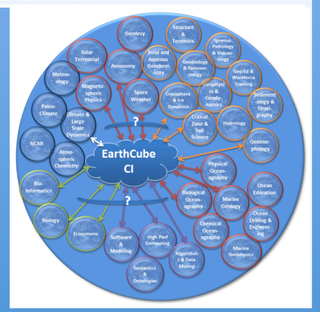
- Transform research and data management practices within the geosciences community over the next decade.
- Provide unprecedented new capabilities, including access to data and visualization tools, to researchers and educators.
- Vastly improve the productivity of the geosciences community.
- Accelerate research on the Earth system.
- Provide a knowledge management framework for the geosciences

Who is EarthCube?

EarthCube is a collaborative endeavor between NSF and the community of geoscientists, computer scientists, information scientists, social scientists, educators and more. Community engagement and contributions are enabled by awards, a serie of end-user workshops, and general outreach efforts, including the website at http://earthcube.ning.com, Facebook, Twitter, YouTube and LinkedIn, which allow NSF to reach out to the US geosciences community to determine their needs and facilitate building an infrastructure for the future. It is our hope that the lessons learned in Earthcube can be used to replicate the process within the other NSF Directorates and Offices, and potentially, internationally by peer organizations.

Who is EarthCube for?

EarthCube is for Earth, atmosphere, ocean, computer, information and social scientists, IT experts, educators, policymakers, data managers, and much more! EarthCube participants thus far include:



Where is EarthCube?

Earthcube is a virtual organization. The center of gravity of the program is the community website hosted by the community engagement group at http://earthcube.ning.com. Earthcube is carried out on the discussion lists, in the blogs, on the website, on Facebook, Twitter, LinkedIn and YouTube, and at the end-user workshops across the country.

Participate in EarthCube!

To get engaged, you can become a member (it's free and open to all), join the virtual discussion groups, find collaborators via EarthCube Member Connections, follow EarthCube on social media or apply to the NSF EarthCube program solicitation. You can also tell NSF what you want EarthCube to do and be by taking the EarthCube Survey.

You also have the opportunity to participate in upcoming EarthCube end-user workshops and influence what EarthCube will do and be. Several workshops have already occurred. For details on workshop outcomes, please see Executive Summaries and Notes.

As of mid-2013, two-thirds of the two dozen workshops had already occurred:

- Geochemistry
- Meetings of Young Researchers in Earth Science (MYRES) V: Sedimentary Record
- Early Career
- Structure and Tectonics
- EarthScope
- Experimental Stratigraphy
- Advances in Data Assimilation & Ensemble Prediction
- Critical Zone
- Envisioning a Digital Crust
- Paleogeoscience
- Education & Workforce Training
- Petrology & Geochemistry
- Sedimentary Geology
- Community Modeling
- Integrating Inland Waters, Geochemistry, Biogeochemistry and Fluvial Sedimentology Communities
- Deep Sea Floor Processes and Dynamics
- Integrating Real-time Data into the EarthCube Framework

Several more workshops have been funded and are planned:

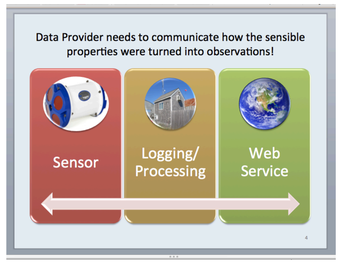
- Ocean 'Omics
- Bringing Geochronology into the EarthCube Framework
- Articulating Cyberinfrastructure Needs of the Ocean Ecosystem Dynamics Community
- Marine Geophysics
- Rock Deformation and Mineral Physics Research

From the EarthCube website http://eartcube.ning.com

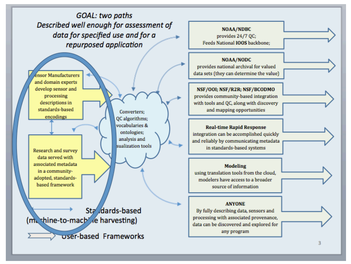
Challenges & Opportunities

Sharing knowledge is a challenge. Standards are important but .... which standards should one use and .... often the use of standards creates barriers to sharing. Creating complete descriptions of information is time-consuming and often unfunded. The integration of encoded knowledge from the beginning – when an observation is made – and building upon it throughout the data life-cycle will help alleviate the burden.

With the advent of freely available, community developed standards (e.g., W3C and the OGC®), we are at a crossroads in how we describe data sets. Sensors and process lineage can be well-described, encoded and registered, providing the ability to create relationships or ontologies in collaborative environments. By encoding the knoweldge in standards-based encodings, we can automate harvesting of information about how an observation came to be and enable the translation of meaning across domains and agencies.



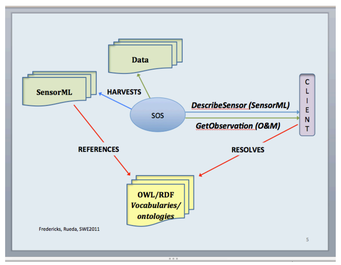
With each step in data management, information is lost. A sensor has limitations and a particular configuration may change the result of how an observable property is reported as an observation.



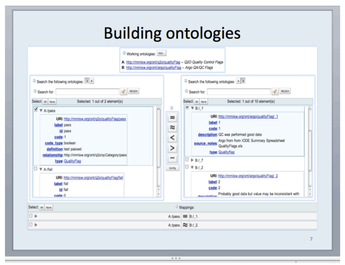
Knowledge is hidden in PDF manuals and log books. Data providers need the ability to fully-describe this knowledge in a way that it can be shared in one standards-based framework, yet made available for the world through catalogues and translators.

Methodology

An approach to integrating semantically-enabled vocabularies into web services of the Open Geospatial Consortium (OGC®) is described below. Well-described terms are encoded in standards-based tags (SKOS) and registered at the Marine Metatdata Interoperability Project (MMI) Ontology Registry and Repository (mmisw.org). The terms provide knowledge about sensors and processing, as well as observable properties and observations. For example, a methodology for computing waves can be linked to an authoritative reference, describing precisely how a property was derived. The model fully-describes the process lineage by developing a series of SensorML files describing each component: an original equipment file (OEM) describes the sensor; the CONDEP file describes the configuration and deployment information including event lists; the processes are described and data quality tests are also described, with input, output and parameters for each component.



Through the use of OGC Sensor Web Enablement Sensor Observation Service (SOS), a client can specify what data is desired (geo-spatial, temporal queries, as well as QC limitations). Upon request, data are wrapped in O&M encodings, fully-describing the data returned by the GetObservation request. DescribeSensor requests provide descriptions of sensor(s), deployments and processing. The encodings include links to the URLS (SKOS-encoded terms at the MMI ORR).



Above is an example showing how a relationship between QC flags can be created using the MMI Vine Tool, using the information provided in the response to the GetObservation. This enables an aggregator to harvest data with described QC flags and have the knowledge to relate terms with different names and values and then to define an action relating to the concept not the name or the value.

Woods Hole Oceanographic INSTITUTION