NSF Earthcube White Paper

OneOceansGeology: Prospect of Serving the Integrated Data in a Visual Framework

Chris Jenkins [1], Mike Gurnis [2], Kerstin Lehnert and Suzanne Carbotte [3], Doug Fils [4], David Divins [4], Walt Snyder [5]

1. INSTAAR - U Colorado, Boulder CO (Chris.Jenkins@colorado.edu); 2. Caltech, Pasadena CA (gurnis@gps.caltech.edu); 3. LDEO - Columbia U, Palisades NY (lehnert@ldeo.columbia.edu; carbotte@ldeo.columbia.edu); 4. Consortium for Ocean Leadership, Washington DC (dfils@oceanleadership.org); 5. Boise State U, Boise ID (wsnyder@boisestate.edu)

BACKGROUND

An integration of the geological data of ocean basins is within reach.

Large-scale multinational resources have been devoted over the last few decades to direct-observations, deep-drilling and surface-sampling of the seabed. Many resulting discoveries have altered the world-view of humankind: the shifting tectonic plates, earthquake/volcanic belts, living deep- and hostile- environment communities, strange materials like gas hydrates and brine and asphalt lakes, our technological accessibility to events at the seafloor, and that the earth's climate is sensitive.

This progress will continue, but the science becomes more complicated as it advances and to stay efficient, it must be supported by data systems. One of those necessary systems will be concerned with the ocean-bottom: its materials, strengths, fabrics, processes and history. The US is able to take the lead on developing such a system, which would then become an attractor for data from many countries.

The design of OneOceansGeology is for a rich data -integration of the cored, drilled and observed in-situ and ex-situ observed data for the sediment and rock materials sampled from the basin floors by coring, drilling, or dredging, with information on their compositions, properties, features and histories. Fortunately, much of this is already in place but, for various reasons, is unconnected. A federating superstructure is called for, with interactive visuals serving the detailed scientific data. It will use a linked data architecture supported by algorithms for semantic 'cross-walk' between the vocabularies. The coordinate systems will reflect the dynamics of earth's history. Each one of these is a key to efficient data access because they give a knowledge context to the data.

PROPOSAL

Strategies

a. This is an Earthcube initiative with a group of software-tool components on the delivery side, and data institutions/serveries on the client side.

b. The proposal dovetails well with governmental data rich programs, e.g.: i. Law of the Sea efforts, including the US ECS program (http://continentalshelf.gov/); ii. One Geology (http://www.onegeology.org/); iii. Global Ocean Observing System (GOOS; http://www.ioc-goos.org/).

c. Interworkability with on-land systems is a key objective. The architectures, semantic cross-walks, and data topologies need to be tested/validated against each other. Therefore our proposal purposely includes entities with experience of data in landscape mapping, and geothermal/groundwater resources.

d. We will tackle more difficult issues such as the land-marine data disconnect, continental margins, geologic fluids, and data from tectonically disturbed regions. We may not completely solve the problems, but will make progress

e. A system of this type will naturally become a strong attractor for data, as experience shows with dbSEABED. That is because the benefits to contributors are immediately obvious.

Scope

Geographically: the ocean basins including continental margins. The depth of penetration will correspond to the IODP. First focus will be on the more tractable parts: the basins and those continental margins with simpler structure. As project techniques advance, more difficult terrains will be tackled.

Components

a. These *existing projects form the basis*: Integrated Ocean Program (IODP, Consortium for Ocean Leadership), the marine geoscience databases SedDB, PetDB, EarthChem and GeoMapApp (Columbia U), a large seabed database dbSEABED (U Colorado; Fig. 1), the geochemical data system EarthChem (marine & terrestrial), the sample registry SESAR, plate-motions from GPlates (Caltech; Fig. 2a), geologic timescales from CHRONOS (U Indiana), ontologies and linked data schemes from SEDIS (IODP-MI), linked data search technologies (Consortium for Ocean Leadership; COL; Fig. 3), National Geothermal Data System (NGDS; Boise U).

b. *International collaborations* are in the offing, especially with Germany, the UK, and Japan.

Science/Technology Impacts

a. *Future ocean exploration* will focus on using in-situ modules with high-capital value, and for which site conditions must be carefully assessed. This project will prepare and re-purpose existing datasets for that role.

b. *Cyberinfrastructure* advances, likely to be focused on: data topologies and sparseness; data semantics / semantic web, especially when trying to join the marine and offshore; data-in-knowledge types of visual display for locating and serving data; linked data architecture.

c. *New research possibilities* will be opened, for example: Sediment supply to the oceans from ancient continent positions and climate, Mineral history of the oceans, Ocean oxygenation and carbon - geographic resolution, Surface expressions of shallow seafloor igneous processes, Niche modelling over evolutionary timescales and geographies. Especially, the system will open new possibilities of planet-scale studies.

d. The *extent of uptake of the data products by related fields in science*, especially geophysics, biogeochemistry, numerical modelling will be a key metric of our success.

METHODS

Techniques

a. We will use data management methods adopted across the EarthCube community, having regard some idiosyncracies of marine data. We will build a software superstructure on existing systems, to achieve their interworkability.

b. Software-based semantic cross-walks for the various vocabularies of measurement parameters and seabed lithologies/features. These problems have largely been solved for the dbSEABED and IODP SEDIS projects. They replace manual writing of ontologies, with computed semantic nets, in the fashion of Wordnet (http://en.wikipedia.org/wiki/WordNet).

c. Overall, the project should be able to supply input to nearly any downstream application and vendor's software. We do want to interface with industry systems such as Petrel, a Norway-USA technology widely used for oil-gas resources. Academic systems such as CoreREF, Corelyzer and CoreNavigator (Fig 4b) will also be workable with the system. In the geologic-time domain, we look to interface with GPlates. Virtual globes (e.g. Google Earth, Virtual Ocean) have a role in the project.

Deliverables

a. Data systems supplying data in a federated way to queries, reports and visuals. Systems such as SPARQL are being evaluated for the purpose.

b. Visualizations, interactive for drawing and query, with 3D fence-diagram and seabed-surface geometries, and ready for the addition of geophysics (for instance as panels). The fences and surface will link the sites of actual data such as IODP and Images-Pages cores, or surficial samples.

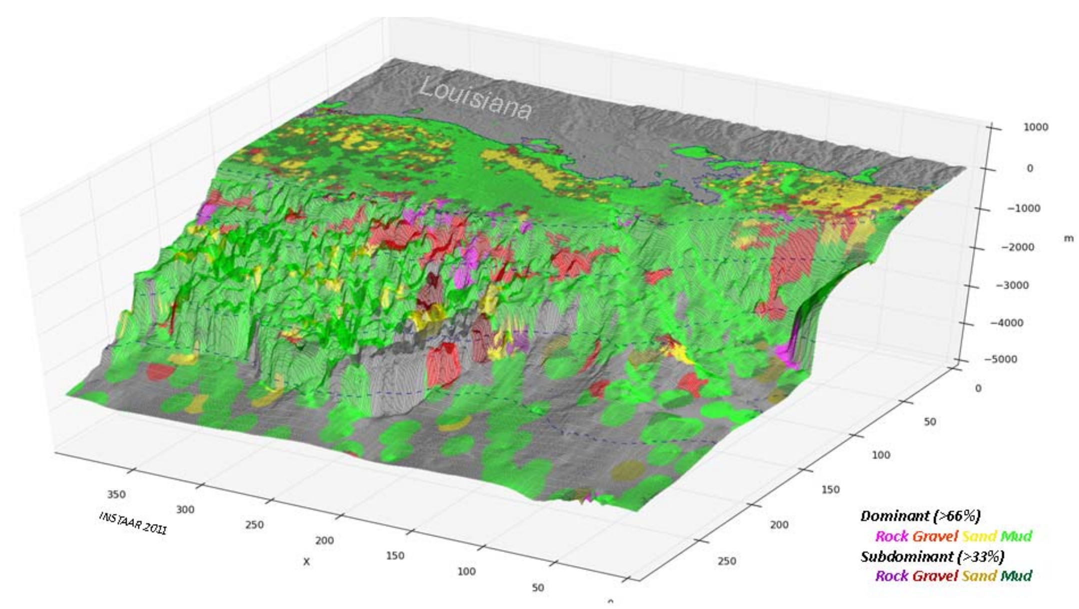


Fig 1. Data deliveries are more effective if they have a backdrop of the current state of knowledge.

Here is an 'executive summary' visual of the seabed sediments and habitats for offshore Louisiana. It is used by fisheries habitat managers and numerical modeling programs. It is an effective background for point data display / query.

The integrated coverage is based on 100,000 data sites and is the result of using software tools for harmonizing over their varying data topologies, data sparseness, semantics, and uncertainties.

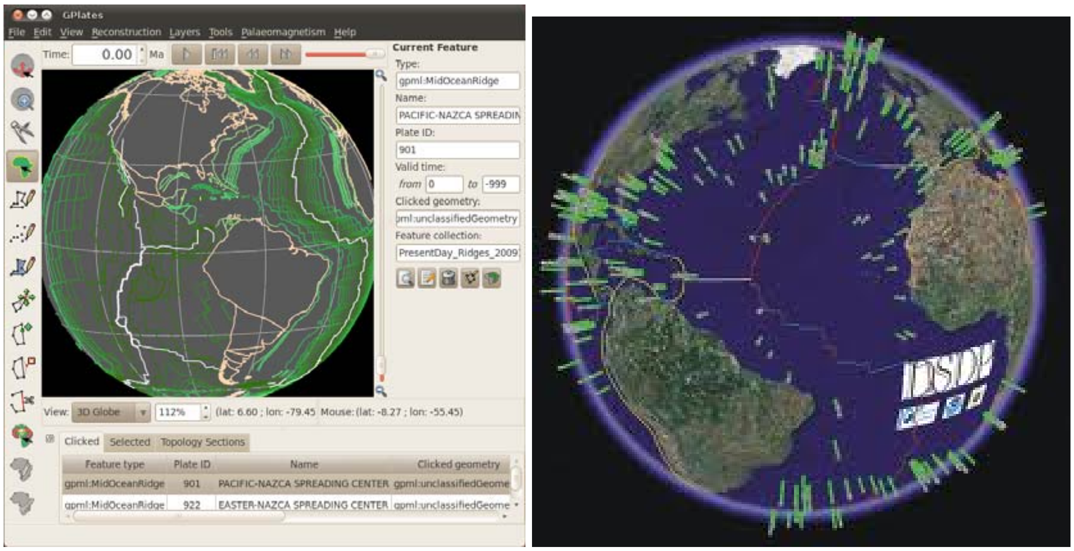


Fig. 2. GPlates, dbSEABED and IODP-SEDIS are natural partners. Downhole sections of ocean drilling cores can be apportioned to their true tectonic plate locations by geological age.

(a) An operator display from GPlates; (b) A plot of some ocean deep-drilling sites for which lithology, downhole logging, core image, and age-control data have been combined and are now inter-workable.

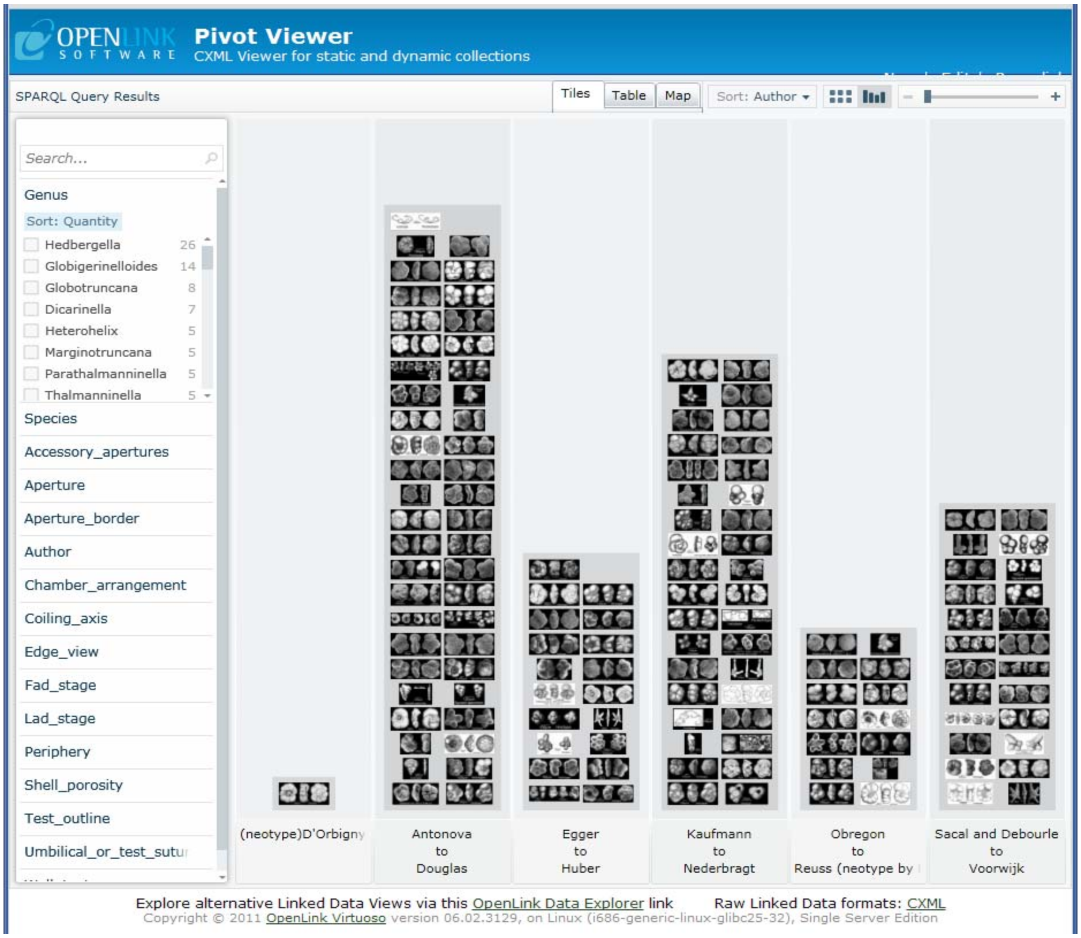
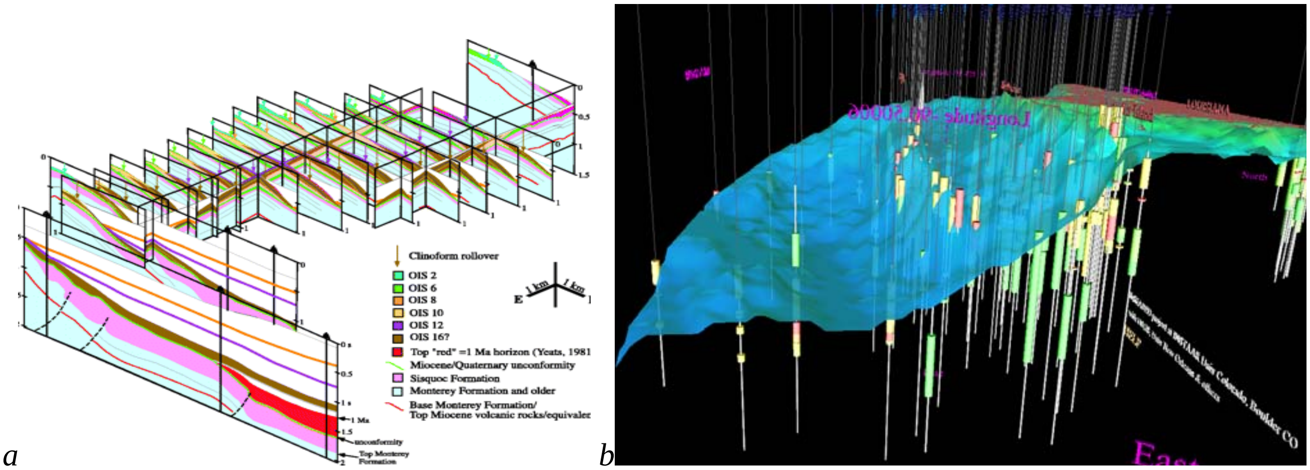


Fig. 3. Here is a linked-data facet-search Pivot Viewer, built by Consortium for Ocean Leadership to show a species identification key. This technology is highly applicable in the geosciences for serving query results - especially for interactive visualizations.



a *b*

Fig. 4. These examples of modes of visualization that would be used in OneOceansGeology. They can be displayed either on a virtual globe or flat map-area. Panels and tracks of geophysical data could easily be inserted. (a) fence-stratigraphic diagram (not this project); (b) an interactive virtual-reality display from CoreNavigator of a complex of core stratigraphies (Ship Shoal, Louisiana).

17 Oct 2011