EXECUTIVE SUMMARY: EARTHCUBE SEDIMENTARY GEOLOGY WORKSHOP

Marjorie Chan - University of Utah, and David A. Budd - University of Colorado, co-conveners;
March 25-26, 2013, University of Utah

Earth Cube Workshop Title: End-User Workshop For Sedimentary Geology

INTRODUCTION

The Sedimentary Geology Community (SGC) domain workshop brought together 57 geoscientists with expertise in modern and ancient sedimentology, stratigraphy, basin analysis, paleontology, paleoclimatology, sedimentary geochemistry, sedimentary petrology, petroleum geology, paleopedology, and geochronology. This community has historically focused on research questions related to the processes that form, shape and affect the Earth's sedimentary crust and distribute key resources such as hydrocarbons, coal, and water. Sedimentary geoscientists also use the sedimentary record to explore the continental crust's evolution, the dynamics of Earth's past climates and oceans, and the evolution of the biosphere. Sedimentary systems also form the framework for the research conducted in many other geoscience communities.

Prior to the workshop, participants provided statements on overarching science drivers in the field, challenges to integrating the community into EarthCube, and the research themes that could be pursued with an ideal EarthCube. Breakout sessions were held on scientific drivers, impediments to sharing and using data, current cyber resources, needed data sets and tools for the future, and potential impact of EarthCube on SGC teaching.

SCIENCE ISSUES AND CHALLENGES

A. **Important science drivers -** Three overarching societal issues were highlighted as drivers that will condition research within the SGC community over the next 5-15 years. Multiple scientific challenges were identified relative to each driver. The primary theme of SGC is to fully integrate our discipline with Earth, Atmospheric, Oceanic, Biologic, and quantitative sciences in addressing the sedimentary dynamics of Earth and planetary systems from the beginning of time, and the current role of human interactions into a sustainable future.

Driver #1 - Securing the energy and water resources needed for an increasing global population while balancing resources for a sustainable Earth.

Related research challenges within the SGC community:

1. Predicting lateral spatial heterogeneity in the geometry and physical properties of sedimentary rocks/bodies. This is a necessity for effectively predicting resource distribution, modeling fluid flow, and mitigating contaminant problems.
2. Improved understanding of organic-rich fine-grained sedimentary systems throughout geologic time, particularly their origin and the processes that generate sedimentologic and geochemical heterogeneity.

Driver #2 - Understanding the Earth as a system, the nature of global climate change, and its impact of climate change on life, the environment, and Earth resources.

Related research challenges within the SGC community:

3. The deep-time sedimentary record must be scrutinized to learn how the Earth's climate system operates in periods of stasis, rapid change, and greenhouse and icehouse conditions. This requires:
    a. Continued development of proxies for ancient climate, improvement in existing proxies, and reconciliation between proxies.
    b. Analyzing the sedimentary record to identify and understand the components of deep-time climate change (forcing factors, feedbacks, tipping points) and the resultant impact on the deep-time Earth system (changes to hydrologic cycle, weathering, denudation, sediment fluxes, nutrient runoff, ocean circulation, extinctions and originations of life, etc.).
4. Develop a deeper understanding of the interplay between life, the physical and chemical environment in Earth's past, climate, tectonics, environmental change, and sedimentary processes.

Driver #3 – Human activities influence, and in some cases dominate, many Earth surface processes. Understanding those anthropogenic influences will be necessary to minimize risks to society and insure environmental sustainability, particularly in deltaic, coastal zone, reefs, lake, and fluvial settings.

Related research challenges within the SGC community:

5. Development of morphodynamic models of how sedimentary environments and landscapes responded on daily to millennial scales to climate change, sea-level rise, sediment supply, induced subsidence, engineered structures, etc.
6. Determining how to use the sedimentary record to make predictions about future environmental changes, assess critical boundary conditions, quantifying parameters of environmental change, and evaluate rates of change.

Research Challenges Common to Drivers #1-3 above:

7. What controls stratigraphic architecture and landscape dynamics? Revisiting the respective roles at varying temporal and spatial scales of autogenic (intrinsic feedback loops) vs. allogenic (climate, tectonics, eustacy) controls.
8. Development of geochronological tools that provide more precise and accurate timing of critical events in Earth's history are necessary to meet all other research challenges. Geochronology must address: (i) the timing, duration, and rates of ancient climate change; (ii) how rapidly life responds to environmental change; (iii) the rapidity of geochemical changes in the Earth system; and (iv) the recurrence & magnitude of natural hazard events preserved in the sedimentary record.

b. Current challenges to high-impact, interdisciplinary science - The workshop specifically focused on challenges and impediments to sharing and using data and other cyber resources. Results are loosely grouped with respect to data bases, cyber structure, and people issues. ***Those that maybe unique to this community are in bold italics.***

<u>Data</u>

1. Lack of knowledge regarding what database and tools exists.
2. Access to data, particularly subsurface data in the private sector and legacy data in physical collections, theses and dissertations, and gray literature, etc.

3. ***Inadequate documentation of data (lack of location data, meaning of symbols on graphical data, unstated uncertainty and reproducibility, no stratigraphic/facies context, incomplete age information, is it raw or corrected data, how was it corrected).***
4. Coordinate systems in metadata are not uniform.
5. Lack of specified methodologies (unified data paradigm) – are different data sets really comparable?
6. Concerns about quality and authenticity of data, particularly older data.
7. ***Uncertainty in whether errors have been removed and data updated (changes applicable to dates, stratigraphic nomenclature, taxonomy).***
8. Data discovery across organizations is difficult (impossible).
9. Coupling diverse data sets is hard.
10. ***Uncertainty in observational data like measured stratigraphic sections – measuring scale (resolution) typically unstated, unclear if lack of features is due to absence or failure to record; what is recorded is subject to interpretation and expertise; definitions/classification of features may vary; no consistent format for representing features/data.***
11. Inability to search/query both observational and interpretive data.

<u>*Cyber structures – they do not mimic science workflows and thought processes*</u>

12. Lack of catalogues as to what is held by organizations; no portal that provides all the need connections.
13. Lack of interoperability between data sources due to vendor's proprietary data formats.
14. Steep learning curves to use cyber resources (not user friendly due to overly specialized formatting, processing, unintuitive interfaces, difficulties in uploading, etc.).
15. Lack of uniform formatting and standard; too rigid a data entry form to capture what is needed.
16. Existing resources not easily searchable, especially for information by place/area, time interval, type of object (e.g., type of facies, environment, sedimentary structure).
17. ***No easy way to integrating subsurface and surface data.***

<u>*People*</u>

18. Lack of incentives to data share.
19. Reluctance to share pre-publication of data, interpretations, and implications (cannot get a citable DOI for just a data set).
20. Lack of training needed to use cyber resources.
21. Lack of time and resources to maintain a website, data base, or tool.
22. ***Concern about unethical uses of one's data (e.g., GPS coordinates of fossil and mineral localities makes poaching and theft for commercial purposes possible).***

TECHNICAL INFORMATION/ISSUES/CHALLENGES

A. Existing tools, databases, etc. needed for pursuing key science questions - The Paleogeoscience domain workshop previously identified ~140 cyber databases, repositories, or tools, with particular focus on paleobiology, marine sediments, geochronology, and paleoclimate. The Sedimentary Geology workshop added 83 additional cyber resources to the compilation – 38 databases, 17 repositories, and 28 tools. These additions particularly focused on LiDAR, map resources, and tools for use in sedimentary geology and subsurface analysis. Of particular note is that there are very few databases for onshore sedimentary geology, most repositories of subsurface data are state or federal agencies, and the most thorough software tools are commercial.

B. Desired tools, databases, etc. needed for pursuing key science questions - To forge new ground and develop richer comprehensions of complex problems and systems, sedimentary geology research requires multidisciplinary approaches, easy access to large volumes of geologic and geophysical data, better integration of that data and legacy data, and increasingly sophisticated numerical modeling of sedimentary systems and stratigraphic architecture.

A Google Earth-like interface is envisioned with topography, surface, and subsurface geology. The interface would (i) allow a wide range of queries, (ii) compile and visualize a variety of data for different time intervals and geographic locations, and (iii) have the ability to create cross sections from designated line paths and make maps for designated areas and time/depth intervals.

Databases[1] *(Geo-referenced; can also be catalog information, not just data)*

1. Geologic maps, cross sections, seismic and GPR lines, LiDAR data, macrostratigraphy.
2. Distribution of fossil organisms through space and time (e.g., Paleobiology Database).
3. A better compilation and integration of the available paleoclimatic data.
4. Drill hole that integrates or links across state boundaries and includes locations, formation tops, geophysical logs, cored intervals, core photographs, poro-perm data, thin section imagery, total organic carbon values, thermal data (e.g., vitrinite reflectance).
5. Measured sections of outcrop and core (both referenced by midpoint of section or line in dipping units). Include scanned images of cores, lithologies, sedimentary structures, grain sizes, textures, fabrics, contacts, trace fossils, thin section imagery, poro-perm data, mineralogy and whole-rock geochemical data (e.g., stable isotopes).
6. Sedimentary rock imagery: stratal geometries, sedimentary structures, photomicrographs, etc.
7. Data on age constraints of stratigraphic units, including source and basis of age.
8. Hub for coordinating databases.

[1] *The SGC recognizes the need to develop protocols for metadata, as well as protocols and formats for core and measured section databases.*

Search Capabilities

9. Multi-tiered search engines to access and search different databases.
10. Searchable map-based areas of interest by time, space, stratigraphic unit or topic.
11. Spatial querying for published work.
12. Filtering tools for searching (search engine and tagger).
13. Ability to search by example - an image of the object or a verbal description - and the query system finds things that are similar (fuzzy query for dark data?).

Tools *(must enable range of data formats and conversions)*

14. Template or checklist tool for metadata format.
15. A suite of tools to easily sort/analyze data using available metadata.
16. Ability to map (with contours) all types of quantitative data.
17. A set of tools that will correlate between sections/core.
18. Tools for compilation and correlation of biostratigraphic ranges for different index taxa.
19. Basic sedimentary interpretation tools (e.g., of depositional environment) involving guided questions that direct interpretation process.
20. Capability of determining sediment volumes/thickness/accumulation rates/fluxes from measured sections, logs, seismic data, etc.
21. Open source visualization software for well, seismic, and LiDAR data.
22. Open source visualization software for stratigraphic columns, timescales and other data (biostratigraphic, chronological, geochemical, petrographic, etc.).
23. Higher resolution paleoclimate climate models.
24. Interoperability with the CSDMS (Community Surface Dynamics Modeling System) suite of modeling tools.

25. Ability to track users of particular features to help organize conferences and workshops of people with common interests.

<u>Other</u>

26. Ability to enter the data as it is collected.
27. Continued development of GeoDeepDive techniques - machine learning data-mining to extract info from PDFs and convert it to a database that can be directly queried.
28. Training modules for database creation and entry, search tools, analysis and visualization.

C. Potential impacts on education and workforce preparation:

The SGC community recognizes that EarthCube can revolutionize how we teach by focusing on research themes and rich real-world data. But educational applications will need to have an ease of use and shallow learning curve so that student can use the data and tools with limited training. EarthCube can be a teaching tool that will supports project-based courses, help build critical thinking skills, promote student inquiry at all levels, and demonstrate how to do research. Participants saw great potential in using EarthCube to explore the spatial and temporal scales of investigation unique to the Earth Sciences. For the non-STEM student, it can change the way they think about how science is done.

Two proposed examples of how the SGC community might use EarthCube in the classroom:

- In a lecture topic on sand size in rivers, students could use the GIS-module and zoom to multiple localities. They access information on the grain sizes and average month discharge (proxy for average flow velocity) at various points in numerous rivers. Students then explore patterns, distance from the source, relation to flow velocity, slope of the land surface, etc. in order to assess controls on grain size (do they decrease with distance from the source as the lecture asserted?).
- In an assignment to analyze the paleogeography in the Cretaceous of the western U.S., students choose sites (GIS module) and retrieve stratigraphy, biostratigraphy, detailed measured sections that show lithology, grain size, sedimentary structures, biofacies, etc. They then use a cross-section building tool and a paleogeographic mapping tool to make their own reconstructions through time.

There was also strong enthusiasm for EarthCube to be a vehicle for virtual field trips via 3D rendering of outcrops and fly-over tools. Field geology requires students to think in four dimensions, evaluate multiple working hypotheses, work at multiple spatial scales, and draw conclusions from incomplete data sets. Virtual trips in EarthCube have the potential to achieve the same results, but also be more comprehensive in that the field component can be supplemented by ready access to appropriate analytical data, thus making the “trip” a comprehensive analysis and investigation. Such opportunities make it possible for all student geoscientists, including limited mobility students and non-traditional students, to visit and “work” any location.

COMMUNITY NEXT STEPS

List of what your community needs to do next to move forward how it can use EarthCube to achieve those goals:

- In order to improve communication with all members of the community, a listserve was immediately established after the workshop (SedGeoNet listserve). It is anticipated that this listserve will be turned over to the new STEEPE coordinating office (*Sedimentary Geology, Time, Environment, Paleontology, Paleoclimatology, and Energy,* http://www.steppe.org/).

- The STEEPE coordinating office will be encouraged to develop structures (web-based dialogs, workshops, eNewsletters, virtual "idea" fairs) that will move the community forward on EarthCube related activities

- Four potential Research Coordination Networks (RCN) initiatives and one Building Block initiative were identified for development (see below). Each would include geoscientists from other domains (particularly structure, tectonics, igneous/metamorphic petrology, geochemistry, hydrogeosciences, geophysics, geomorphology).

RCN Initiative - Geoscience images

An RCN focused on imagery will bring researchers and educators together to: identify the range of scientific images (fields of view that span kilometers to nanometers), converge on common understanding as to what information should be extractable for images in EarthCube, and begin a dialog on how to create interoperable image databases that can be searched and exploited.

RCN Initiative - Framework and user interface for collecting field data

Field data is typically viewed as point data (objects in space and time), but for sedimentary geologist the measured section can also be viewed as streaming data – multiple tracks of different data collected progressively up the outcrop. The challenge is capturing this data digitally in real time so that in can be retrieved and used for varied applications. This issue needs to be pursued in coordination with other field-based geoscience domains and must integrate with cyberinfrastructure experts from the start so the field workflow can be captured and integrated.

RCN Initiative – Subsurface data integration

Tremendous amounts of subsurface data already exist. These data are very diverse. Some is streaming, some is point data. It includes geophysical and geochemical data; images, actual samples/cores, analytical compilations, and derivatives (maps, cross sections, etc.). Some are digital others are analog. Much is already is disparate databases (particularly those of state & federal agencies), but none are interoperable. Georeferencing is varied. Data discover and access is random.

The goal of this RCN would be to what data needs to be made more accessible, how it might be linked, and how it needs to be visualized. The RCN would have to include cyberinfrastructure experts so as to make the connection between how the data can be managed versus how geoscientists want to access and use this data.

RCN Initiative - 3D Geodata and its visualization

There are many groups across geosciences that are interested in 3D visualization of surface and subsurface data. An RCN amongst these groups would focus on issues surrounding spatial resolution, temporal resolution, metadata standards, and the development of tool that better integrate surface data (2D maps) to subsurface data. Exploitation of existing tools/databases would be included. A goal would be to assess issues such as what types of data, what types of visualizations, and workflows. The long-range target is a portal with multiple visualization tools, demonstrations on each tools capabilities and applications, and training modules that make each tool broadly accessible.

Building Block Initiative – Mini EarthCube - Mesozoic Geology of the Colorado Plateau

A "mini EarthCube" project will integrate current data sets to build an architectural, geospatial visualization model (a location-based "Google Earth"-style search engine) of the Mesozoic sedimentary geology on the Colorado Plateau (Utah, Colorado, Arizona, New Mexico). This will be a proof of concept resulting in a geoinformation framework with a portal that allows access and visualization of fused/ tiered/multi-scaled geological layers. It will serve as a test case for the future, grander EarthCube. This "mini EarthCube" initiative requires partnership among various entities within and beyond the SGC, and integration with GIS cyberinfrastructure experts. We propose building a model patterned after the successful, existing Lunar Mapping and Modeling Project (LMMP) – www.lmmp.nasa.gov, in collaboration with the cyberinfrastructure team that designed and delivered the lunar model to NASA for use by the lunar science community, educators, and the general public. This collaborative approach will leverage on NASA and NSF-funded capabilities and approaches that can be the springboard to quickly move to a focused geologic project that will have high impact.