EXECUTIVE SUMMARY: EARTHCUBE WORKSHOP RESULTS

(Kerstin Lehnert, Lamont-Doherty Earth Observatory of Columbia University: March 6 & 7, 2013)

Steering Committee: Chuck Connor, Elizabeth Cottrell, Rajdeep Dasgupta, Kerstin Lehnert, Abani Patra

Additional Report Authors: Maryjo Brouce, Adam Kent, Erik Hauri, Frances Jenner, Abani Patra, Julia Hammer, Christian Huber, Brendan McCormick, Amanda Clarke, Marion Le Voyer, Ken Rubin, Ben Andrews

Earth Cube Workshop Title: *Earthcube Domain End-User Workshop: Community-based Cyberinfrastructure for Petrology, Geochemistry, and Volcanology*

Introduction: More than 75 scientists, data managers, sample curators, and cyberinfrastructure specialists (on-site and remote participation) convened for 2 days at the National Museum of Natural History of the Smithsonian Institution in Washington, DC, to evaluate the status of cyberinfrastructure 'readiness' of petrology and, geochemistry, and articulate CI needs and requirements for these domains to contribute to the overall EarthCube architectural design phase. A late winter storm in the DC area obstructed travel and prevented more registered participants from attending on site.

Petrologists, geochemists, and volcanologists share the scientific interest in fundamental questions concerning the chemical and physical state of the Earth, the Moon and the other terrestrial planets; processes that have lead to physical and chemical differentiation and evolution of planetary interiors and environments through time; and the relationship between geologic processes and societal issues such as natural hazards and resource use. They generate data in the field by collecting samples or monitoring volcanic activities, in laboratories by performing chemical analyses or physical experiments, and by using these observational data to compute models.

The workshop identified important scientific drivers for advancing cyberinfrastructure in the domains, technical, data-related impediments to addressing scientific challenges, and resulted in a list of recommendations for next steps to realize the cyberinfrastructure vision for this community. With 28 science scenarios submitted in advance of the workshop and a record participation in the EarthCube Stakeholder survey, this community demonstrated a high level of engagement that continued throughout the discussions in plenary and breakout sessions.

SCIENCE ISSUES AND CHALLENGES

1. **KEY SCIENCE QUESTIONS or Important science drivers and challenges:** Participants identified

several high-priority science questions that will be the focus of and grand interdisciplinary efforts during the next 5-15 years:

- Understand the co-evolution of the geo- and biosphere
- Create a four-dimensional (space-time) description of the chemical and physical state of the Earth including the composition of and extent of fluxes between its major reservoirs -- core, mantle, crust, biosphere and hydrosphere.
- Understand the role of disequilibrium processes in the formation and evolution of planetary bodies.
- Integrate observations at volcanoes (e.g. seismic activity, ground deformation, emissions, magma chemistry and petrology, magma physical properties, plate tectonic parameters) in order to (1) forecast and mitigate natural hazards (2) understand and communicate volcanic impacts on society, and (3) to search for natural resources.
- Map the feedbacks between planetary evolution, plate tectonics, volcanic activity and climate on short and long timescales.
- Communicate the grand challenges of science to society.

2. **Current CHALLENGES to high-impact, interdisciplinary science:** Several categories of scientific challenges arose during the breakout sessions. Unless addressed, they will likely serve as significant impediments to conducting high impact, global scale research with a new Earth Cube infrastructure. These challenges can be broken into the following categories:

    - <u>Data challenges:</u> Limitations in accessible data types, diversity and extent of temporal and spatial scales of existing data sets (mostly collected for regional studies and with different goals), variations in meta data standards or lack thereof, differences in data documentation in different scientific community. All of these issues manifest mismatch between marine and terrestrial data resources, which will need to be merged and standardized to address global scale variations and patterns in magmatic phenomena.
    - <u>Geosample Strategies</u>: Most geologic terrains of interest to this end-user community do not have sufficient or even sample density through time and space. In addition, there is insufficient appreciation of this problem by many current geochemical/petrological practitioners and funding agencies.
    - <u>Sample Curation</u>: Poor and uneven access and management of sample collections, incomplete sample tracking and linking of samples to analyses in the literature or databases, discoverability of existing samples.
    - <u>Knowledge</u>: Lack of basic knowledge of limitations and uses of data and models across, within

and between disciplines.

- Interdisciplinary Conceptual Framework: Missing conceptual models, gaps in understanding, for instance of process rates, insufficient geochronometers over the full time range.
- Community: Barriers to collaboration, incomplete shared knowledge of data resources, expertise, specialized skills, toolsets, lab capabilities, etc. The community needs a "facebook" style networking site for geochemists and aligned scientists as a means to bridge this barrier.

3. **Technical Information / Issues / Challenges:** The workshop participants identified an initial set of desired capabilities (tools and databases) needed for pursuing key science questions:
   - Discoverability: Improved infrastructure (search engines, catalogs) is needed that facilitates discovery of data, samples, models, and management tools.
   - Interoperability: Software packages utilized during data acquisition should be transparent to data analysis and visualization softwares. Tools should support analytical thinking and numericisms.
   - Compliance: Data and metadata should be captured at the point of acquisition in a way that they can seamlessly be managed throughout their life cycle, including upload to repositories in order to satisfy data management requirements.
   - Format standards: Standards need to be established within the existent data repositories.
   - Sample tracking: Systems should be in place to promote spatial contextualization of analysis through sample registration, imagery, and links between samples (hand samples, thin sections, splits, etc.) and analytical data.
   - Archiving: Absent repositories, databases, and heterogeneous media should be identified and recovered.
   - Metadata: Ancillary contextual information such as science objectives, data provenance, and uncertainty estimates at each step in workflow needs to be included with the data.

4. **Community next steps:** Participants agreed on a number of steps that could be taken in the near future, many of which are 'low-hanging fruit'.
   - Inventory: Create an online list of resources, a list of metadata for each technique/method, and an online list of science scenarios which fall under the "geochemistry/petrology" umbrella. These should be hosted on the EarthCube website and should be editable by registered users, but moderated and administered centrally. Individuals should start to populate the list of existing resources that they know of (data, samples, models, visualization tools, educational tools, expertise pool) and the list the metadata they use on a daily basis.
   - Participation: Get started! Pitch in! The following simple tasks should be taken on at the

individual level:

  - register on the Earthcube website and populate your own profile with picture/information
  - register your own samples (SESAR)
  - refer to the list of existing resource, start using them fully in your research and teaching, and encourage your students and colleagues to use them.

If everyone is doing a little bit of all of this, we will make HUGE progress rapidly! We need to encourage a change of mindset ("cultural shift") so that our community begins contributing to grassroots cyberinfrastructure.

- <u>Social networking</u>: Create a social network and expertise pool for early career and other scientists to exchange and build up the future EarthCube collaboration network/trust.
- <u>Sample Curation</u>: Consider samples as resources: we are not going to create a new HUGE sample repository, but just record and publicize which samples are where, who are the owners/ people responsible for each specimen, which analyses have already performed, how to access/borrow each specimen. A new library can be created, initially compiling existing online searchable sample catalogues and also encouraging collection managers and curators to digitize their own collection catalogues.
- <u>Databases</u>: Initiate work on immediate needs/low hanging fruits:
  - geospatial model of surface heat flux (e.g. Shapiro & Ritzwoller, 2006), plate boundary motions (e.g. Bird, 2003), crustal thickness (Bassin, 2000)
  - tephrachronology database
  - gas emission database
  - comprehensive, well standardized, error-documented volcanic rock composition database (marine and terrestrial)
  - getting historical data properly represented in EarthChem
  - GeoPRISMS as a venue to publish data (private and in grey literature)
- <u>Community building</u>: Self-organize in small working groups made up of people with similar science goals. Attempt to articulate specific science drivers and identify data formatting and cataloguing challenges. For example,
  - Forming an IAVCEI Commission of Explosive Volcanism working group to assess the need for a database of semi-quantitative data (video) of explosive eruptions and analogue experiments.