DRAFT

EXECUTIVE SUMMARY: EARTHCUBE WORKSHOP RESULTS

April 24-26, 2013. Millennium Harvest House Hotel, Boulder, CO

Conveners: Albert Kettner (Univ. Colorado) and Emilio Mayorga (Univ. Washington)
Organizing Committee: Anthony Aufdenkampe (Stroud Water Res. Center), Anne Carey (Ohio State Univ.), Basil Gomez (Univ. of Hawaii at Manoa / State of Hawaii Commission on Water Resource Manag.), Laodong Guo (Univ. Wisconsin Milwaukee), Sherri Johnson (Oregon State Univ. / US Forest Service), Bernhard Peucker-Ehrenbrink (WHOI), Peter Raymond (Yale)

Earth Cube Workshop Title: *An EarthCube Domain Workshop integrating the inland-waters geochemistry, biogeochemistry and fluvial sedimentology communities*

Introduction: This 2.5-day workshop brought together 55 diverse participants (with half being early-career: about 9 graduate students, 9 postdocs and 10 assistant-level faculty or research staff) and an additional 10-15 virtual participants to discuss the status, needs and opportunities regarding cyberinfrastructure impacts on the advancement of overlapping communities of scientists that address sources, composition, processes, fluxes and fates of constituents in terrestrial surface waters. 82 individuals registered for the workshop (including virtual), but several were unable to attend due to last-minute conflicts or health issues. Participant recruitment placed a special emphasis on inviting and attracting domain scientists, including ones with strong interests and activities in data integration and synthesis and cyberinfrastructure-enabled science. The workshop was structured to enable substantial exchange on scientific results and interests in order to both stimulate breakout EarthCube discussions and maintain strong participant engagement; these exchanges were facilitated through keynote presentations, posters, and many brief (2-5 min) "pop-up" presentations on topics spanning the range of disciplines, including existing cyberinfrastructure serving the represented disciplines. Participants were divided into breakout groups according to research approaches (scale, environments, disciplines), and re-mixed across breakouts to maximize the generation of new perspectives.

A goal of the workshop was to help define and form a community of aquatic scientists who have common ground and overlapping interests in their study of inland waters geochemistry, biogeochemistry and sedimentology, but may communicate rarely due to disciplinary fragmentation or regionally specific interests. The focus on a unifying set of environments or "hydroscapes" (dynamic water systems) and data needs and conceptual frameworks involved when addressing terrestrial, aquatic and atmospheric and anthropogenic influences on surface waters indeed proved to be an effective mechanism for identifying common data and cyberinfrastructure needs across disciplines.

Workshop participants identified important scientific drivers or grand challenges for advancing cyberinfrastructure capabilities benefiting their disciplines, as well as technical and other impediments to addressing these scientific drivers. They compiled an extensive list of data, software, tools and modeling resources used by the community, and created a list of recommendations and consensus on three unifying, grand-challenge use cases for advancing and leveraging cyberinfrastructure capabilities.

SCIENCE ISSUES AND CHALLENGES

1. Key science drivers and challenges: The study of dissolved and particulate matter is of relevance to geoscientists and ecologists and encompasses diverse landscape scales and types, element and material cycles, approaches, and data collection contexts. This broad community is highly interdisciplinary; two different breakout groups came to the similar conclusion that this interdisciplinary nature makes it very difficult to label data sets as being "within" vs "outside" the discipline. The sub-disciplines are complementary and interact with one another. Several unifying themes emerged, containing more detailed questions and challenges:

- We are in the era of anthropogenic changes.
  - Need for understanding current states in relation to historic mechanisms driving the systems. The role of *legacies:* Climate history, soil structure, past disturbance, past land use.
  - What is the magnitude of climate change impacts vs. direct human perturbations such as land use change, aquatic environment modifications, and hydraulic engineering?
  - What are the global trends in carbon export, concentrations, gas evasion fluxes, and burial?
  - How will climate change affect higher latitude changes/creation of wetlands?
  - Trajectories and impacts of wetland degradation and restoration.
  - Advancing understanding of water ecosystem services to address landscape management.
- Connectivity: Lateral linkages via water transport.
  - When and where do hillslope flowpaths connect and disconnect?
  - What is the impact of groundwater connectivity on stream processes?
  - How do we connect flowpaths and systems across scales?
- Temporal perspectives: Predicting time of response to climate change and disturbance across biogeochemical response variables and water body types.
  - Pulsed events, extreme events (hurricanes, land slides, …).
  - How do seasonality, magnitude and duration of events influence biotic responses?
  - How does temporal variability impact societal needs or benefits?
- Spatial perspectives: Predicting zones of conservation, transformation, propagation within inland water networks across range of response variables.
  - Defining and mapping the time-varying *hydroscape*, including small streams.
  - Upscaling different systems and fluxes to regional, continental, and global scales.
- Grand goal: Integrating and translating across spatial scales and forecasting in time. Needs:
  - Improving the mechanistic understanding of processes.
  - Increasing spatial and temporal extent and resolution of observations.
  - Dynamics of fluxes, process rates, and system scales.
  - Linking across different types of processes and forcings: physical, biological, chemical, geomorphic, and anthropogenic.
  - Linking understanding of quantity and composition of complex constituent mixtures.
  - Using fine-scale data and understanding to inform global scale understanding.
  - Determine how water, sediment and biogeochemical fluxes *throughout* a river basin are connected and affected by event magnitude, duration, sequencing and spatial extent.
- Other challenges:
  - How do relationships between discharge and concentration impact downstream ecosystem function, and how do food webs impact biogeochemical and even hydrologic responses?
  - Estimate global time series of monthly carbon burial fluxes in all aquatic depocenters, and carbon gas exchange fluxes across all water surfaces.
  - How do floodplains function geochemically and geomorphologically?
  - Understanding delta subsidence and retreat due to decreased sediment supplies.

TECHNICAL INFORMATION/ISSUES/CHALLENGES

Current challenges to high-impact, interdisciplinary science:

- Addressing the various types of data: At a point (across space, these will become); Spatial (local regional global); Temporal (minutes, days, weeks/months, annual, decadal, geologic scale); Field Samples; Modeled Samples
- Datasets that bridge measures of quantity and composition of complex constituent mixtures
- Disconnect between continuous measurements and concentration data.

- Challenge of observing and characterizing hot spots and hot moments (fluxes and processes that are highly concentrated spatially and temporally)
- Using, sharing, and coupling broader models (geochemistry, hydrology, etc)
- Challenges finding data.
  - Lots of free data, very different formatting, provenance, other data characteristics.
  - Zero order challenge is knowing what data is out there, knowing what resources provide access, etc. Data discovery is lacking.
  - Challenge in finding linked data - spatially and temporally
  - Downscaling and upscaling data
  - What are the core capabilities that end-user domain scientists need in terms of data management/cyber-infrastructure?
  - Can we filter content based on assumed associated data?
  - Community standards for data management would make our science practices more efficient. Low transaction cost is necessary for adoption.
  - Need benchmark data (such as for training/validating models). Model intercomparison portals, testing our model quality.
- Challenges using data.
  - Data quality varies, across soil types, DEM, sensors precision, accuracy
  - Enormous datasets that are difficult to use
  - Units are different, and metadata doesn't always provide clarity
  - Curation. Heterogeneous data quality. Lack of information about quality. Reviews of data sets? Summary of data quality and characteristics?
  - Clearinghouse function of Earthcube? Meta DataBase
  - Vocabulary variation. Semantic search.
- What are the pressure points for the community? AGU and Nature Geo saying "we won't accept this unless you link your publication to data". But...where does it go? Let's go to a couple of the heavy hitters and ask them to be the bad guys. Others will follow.
- Leverage other, related disciplines for their solutions to similar problems. Which atmospheric/ oceanic lessons would be translated to our domain?
- Need to change culture about code sharing. What are the incentives? Why do we have different rules for code vs data sharing?
- How do we decide the scale for making decisions? E.g., OGC deciding on standards vs grassroots community?

1. Desired tools, databases, etc. needed for pursuing key science questions.

- National data set from water treatment plants and sewage treatment plants (they generally are apprehensive and don't share but have great data)
- International data sets (some countries do not share)
- Smoothed county level data
- Improved and standardized statistical approaches for small systems
- High resolution data throughout the hydrological cycle, not just field season campaigns
- communication about current projects, research activities
- Tools for coupling models are important and useful.
- Rating datasets and models. Need to understand relative value of data and models faster.
- Need standards for data exchange and formats.
- Understanding and curating what has been done and what could be available.
- Digitizing the wealth of information that exists behind us (historical data).
- A large, comprehensive catalog? Consistent formatting. Need crosswalk for vocab.
- Central retrieval system; centralized searching, not necessarily hosting data physically

- We need mobile science apps to make field work more efficient. Improved models of concentration discharge relationships (*how do we share models?)*
- Maps of built infrastructure information and data (tile drainage, pipeline, sewage treatment outlets, past land use)
- Ground water chemistry database
- Continuous categories of soil maps and soil chemistry
- Fertilizer use data
- Watershed activity for research
- Historical maps of land use, lead deposition etc
- Species distributions of fish, invertebrates, amphibians, native, invasive species
- Hyporheic flow paths
- Soil moisture maps

COMMUNITY NEXT STEPS

1. List of what your community needs to do next to move forward, and how it can use EarthCube to achieve those goals.

<u>EarthCube activities.</u>

- Sign up and participate in EarthCube Ning site group. Send email to all participants, all who registered for workshop, and otheres who expressed interest, to join this EarthCube Ning group.
- Make the Workshop Google Drive accessible from web site via link that gives read-only access; and send invitation to participants to give individuals edit access.
- Participate in second RCN call, if there is one.

<u>Community building</u>. Maintain discussions and momentum

- Continued at least informal gatherings at conferences, including AGU; and later sponsor a special session at the Joint Aquatic Sciences Meeting in May 2014
- Identify potential sources of support for subsequent meetings, and general alignment
    - US: SESYNC / Water Science Software Institute; CUAHSI, CZO, LTER StreamChemDB, USGS CIDA, Global Rivers Observatory
    - International: GEOSS/GEO Water Quality Community of Practice, IAEA, GLORICH, GEMS/Water, RECCAP
    - Relevant observatories: GLEON, NEON, CZO, LTER/USFS
- Prioritize focal areas - ask specific questions (eg, Eutrophication; Human influence on reservoirs, lakes and wetlands; Mechanisms that control gas exchange). Success of past efforts has been in creating a database that handled narrow sets of data.
- Identify low-hanging-fruit cyberinfrastructure steps and products, with current funding:
    - Increased integration with terrestrial work
    - Model sharing
    - Explain relevance to societal problems.
    - Get universities involved. Graduate training - consider a piece of the planet; do a synthesis of the available data; visit, tour, and meet with researchers. Model this after the field-course program of the Duke-university based Organization for Tropical Studies.
    - Database building
        - Divide "wish list" into efforts that we need other communities to tackle vs. efforts that we need to take on
        - Create an interactive wiki with data sources, searchable per region, use/parameter group
        - Online training webinars
            - Loading/using data in CUAHSI HIS and Water Data Center
            - IEDA System for Earth Sample Registration (SESAR), http://www.geosamples.org/
            - IEDA EarthChem Data Library, http://www.earthchem.org/library

    - Begin loading our dark data in these systems, even if not perfect for our community.
    - Assemble a priority list of dark data (e.g., Hans Eugster and Peter Kilham saline-lakes datasets).
    - Develop/distribute checklist for desirable metadata, templates for data
- Organize and sponsor Software Carpentry “boot camps” for this community
- Write white Paper, for Eos or peer-reviewed journal

Advance the three Consensus Use Cases, both with EarthCube and more broadly.

1. **Title:** *GoogleEarth-like "H2O" (Headwater to Ocean) Data/Model Access and Visualization.* **Goal:** Create a portal that provides access to constituents transported by rivers (from headwater to the coastal ocean) to better constrain fluxes and the understanding of processes that determine fluxes. This includes sensor data (in-stream, remote sensing), legacy data (links to existing data repositories, but also mining/rescue of dark data), model output, integrated spatial data characterizing watersheds, and careful quantification and propagation of uncertainties.
2. **Title:** *The role of "events" on water, temperature, sediments, solutes and ecology: comparing case studies of ENSO impacts ("IMENSO", IMpact of ENSO) on the flux of sediments and aquatic biogeochemistry.* **Goal:** Examine the role of ENSO climate variability on sediment/ carbon/ nutrient fluxes through case studies of data rich and data poor watersheds around the world (e.g. California, New Zealand, Amazonia, Ethiopia, Australia).
3. **Title:** *Role of inland waters in historic and contemporary global biogeochemical cycles (GBC).* **Goal:** Develop an understanding of inland waters role in GBC, based on an understanding of water quantity (fluxes, storage, residence times) as a foundation to understanding carbon and nutrient fluxes, with an emphasis on greenhouse gases. Begin with a contemporary, process-based global baseline understanding, then predict past and future.