EXECUTIVE SUMMARY: EARTHCUBE WORKSHOP RESULTS

Convened by A.K. Aufdenkampe, C.J. Duffy, G.E. Tucker
Univ. of Delaware: Jan. 21-23, 2013

Earth Cube Workshop Title:
Engaging the Critical Zone community to bridge long tail science with big data.
Short Title: Critical Zone EarthCube Domain Workshop

Introduction:

Critical Zone (CZ) scientists take as their charge the effort to integrate theory, models and data from the multitude of earth science disciplines collectively studying processes on the Earth's surface - from the atmosphere at the vegetation's canopy to the lower boundary of actively cycling groundwaters. Sixteen CZ disciplines were represented at the workshop with experiences that span the range from Big-Data to Long-Tail science. The national Critical Zone Observatory (CZO) network has begun the process of building a community cyberinfrastructure that the workshop organizing committee feels can serve as **a pilot for the EarthCube endevor** of engaging of a diverse community of Earth System scientists to embrace and co-develop a shared data and modeling system.

The Critical Zone EarthCube Domain Workshop had 103 registered participants, with 68 participating in person, 40 participating virtually, and several on-site visitors. 28 were early career (6 graduate students, 11 post-docs, 11 assistant level faculty). Most participants were self-described as representing more than one of the 16 CZ disciplines:

- Biogeochemistry 30
- Biology / Ecology 15
- Biology / Molecular 3
- Climatology / Meteorology 15
- Data Management / CyberInfrastructure 46
- Engineering / Method Development 8
- Geochemistry/Mineralogy 13
- Geology / Chronology 14
- Geomorphology 15
- Geophysics 8
- GIS / Remote Sensing 31
- Hydrology 46
- Modeling / Computational Science 36
- Outreach / Education Research 7
- Soil Science / Pedology 16
- Water Chemistry 14

Workshop participants self-divided into four breakout groups, developing extensive responses to workshop outcomes over five breakout sessions. We have here distilled those breakout notes, which are available at https://drive.google.com/#folders/0B_VW4kvIBAzQSE91SEdkVzIGYkE.

SCIENCE QUESTIONS, CYBER CHALLENGES AND NEEDS

1. **KEY SCIENCE QUESTIONS:** Participants identified several high-priority science questions that will be the focus of interdisciplinary efforts during the next 5-15 years (compiled from all breakout groups, sessions 1 & 5):

The central scientific challenge of the critical zone science community is to **develop a "grand unifying theory" of the critical zone through a theory-model-data fusion approach.** This concept expands on the classical notion of Hans Jenny's state equation for soil formation -- $S = f(cl,o,r,p,t,\ldots.)$, where *S* is for soil, *cl* represents climate, *o* organisms including humans, *r* relief, *p* parent material (lithology), and *t* time -- into a 4D landscape-scale model of coupled physical/ chemical /biological processes that frame the critical zone's evolution, function, and response to change. Developing such a grand unifying theory requires answering three broad questions:

1. How do tectonics, lithology, climate and biology co-determine the evolution of critical zone structure and function?
    - "structure" = 3D arrangement of the remnants of physical and chemical weathering from surface to bedrock, and associated spatial patterns in biological communities. It includes properties such as topography, chemical composition, porosity, permeability, and physical structure (cohesion, fracture density, shear strength, and similar properties), as well as biological communities both above and below the surface.
    - "function" = the processes of transforming and transporting energy and materials. CZ function includes all "ecosystem services", including water routing, storage and filtration; biogeochemical transformations such as nutrient, carbon or greenhouse gas uptake/storage/release; sediment flux; and others.

2. What are the drivers of energy and material fluxes (i.e. water, sediment, carbon, nutrients, solutes, etc) moving through the critical zone?

3. How will critical zone structure, function and evolution respond to human and natural disturbances and over various time and spatial scales?

A second, yet equally important, challenge is whether a unified theory of the Critical Zone can create the necessary knowledge base to evaluate the complex issues of supporting sustainable landscapes. Several specific high priority questions were identified to provide detailed examples of the applications of the broader questions above. These were considered high priority in large part because of their immediate relevance to human and ecological sustainability issues.

- What is the impact of human-induced changes to the nitrogen cycle on the land, air, water, and ecosystem of the critical zone across the scales where science-based management decisions and actions are made (individual land parcels to basin scales)?
- What is the current distribution of soil carbon at global, regional and landscape scales, and given the drivers of these distributions how will soil carbon stocks change in the next 50-100 years?
- What essential biodiversity and other biological variables are most relevant for characterizing the biological processes that co-determine critical zone structure and function? At what scale are these variables best measured?

2. KEY [CYBER] CHALLENGES:

2. KEY [CYBER] CHALLENGES: Several themes emerged as consistent challenges faced within/across the involved disciplines (compiled from all breakout groups, sessions 2 & 5):

General cyber-challenges include:

- CZ data is diverse and much of it is “dark” . There is no one-stop shop for even knowing what is available, let alone accessing it.
- One constraint that limits community access to Essential Terrestrial Variables (ETVs) for watershed modeling is that the data sit on many servers, with multiple (and heterogeneous) formats, very large files, and complex security, making it difficult for scientists or students to use the data. A second challenge is that even if the above problems were fixed, the scale of the data and the tools necessary for data mining, fusion, and visualization are not yet readily available or usable by scientists. The problem of accessing and sharing real-time data collected by CZO scientists is a theme in this challenge
- Modeling, computation, and numerical prediction is carried out in an ad hoc manner with limited cross-domain collaboration (water-bio-rock) and without the benefit of close interaction with cyber scientists and numerical analysts. An outcome is that such results both challenging to obtain and are not easily reproducible.

Specific scientific challenges that require cyber-solutions:

- Understanding diverse scientific worksflows by CZ scientists and applying appropriate tools to promote shared discovery requires a fundamentally new approach to how the scientific process will evolve from experimental data, to interpretation and models, to the creation of knowledge and wisdom.
- Uncertainty and variability are fundamental to all CZ use cases. Across a range of activities -- from field experimentation where sensors are impacted by environmental noise, to issues of communication in wireless sensor networks, to real-time data assimilation in nonlinear spatially distributed models, to data and model analytics, visualization and computational steering -- uncertainty and variability must be addressed. Although these areas are effectively dealt with in individual CZ disciplines, there is not at present a general framework to efficiently deal with this specific challenge.
- Closed technologies such as WSN’s (wireless sensor networks) have evolved as proprietary products that are not yet useful for the Critical Zone problems where low-power, integrated, heterogeneous, co-located systems of research-grade sensors are necessary to resolve multi-state, multi-process discovery within fully coupled bio-geo-chemical hydrological systems. In particular, research-grade, low-power bio- and chemo-sensors are particularly missing in the integrated measurements at CZO’s.

3. **KEY [CYBER] NEEDS** needed for pursuing key science questions with brief elaboration (compiled from all breakout groups, sessions 4 & 5):

Each breakout group independently envisioned a future cyber-infrastructure that might enable seamless 4D visual exploration of the knowledge (data, model outputs and interpolations) of critical zone structure and function, similar to today’s ability to easily explore historical imagery of the earth’s surface using Google Earth. This map-based visualization system would allow a user to zoom above or below the Earth’s surface to view:

- point locations with sensor-based or sample-based time series observations, and direct access to that data

- profiles from soil pits and boreholes with sample-based data, and direct access to that data
- 2D satellite imagery and GIS data coverages from many different agencies and sources, with time sliders to explore historical images and view differences in time
- 2D & 3D images of CZ structure obtained for the subsurface via geophysical approaches or for the surface obtained via LiDAR and other geospatial imaging approaches.
- Depth to groundwater and depth to bedrock
- Modeling results, visualized in 2D, 3D and 4D
- 2D and 3D interactive visualizations of select datasets

A number of immediate needs were identified:

- Much more data in discoverable repositories with full metadata (i.e. too much of CZ data is “dark data”).
- Easy-to-use web application suites for integrated data discovery, access, visualization and publication. This would lower the activation energy to get more CZ scientists to use the existing cyber-infrastructure.
- A power users “toolbox” in an easy-to-install and easy-to-teach cross-platform package to enable cyber-savy CZ scientists and data managers to more easily manage local data, publish their data to repositories and directly access existing data resources using web services APIs.
- Training and support to increase the overall computational and data handling skills of the CZ science community at all levels, from taking the first steps beyond spreadsheets to contributing to open-source scientific software projects.
- Central data catalogs that allow single searches of multiple repositories from many CZ disciplines and domains.