Organizational Efforts for CZ Geochemical Data

S. L. Brantley, K. Lehnert, X. Niu, J. Williams, D. Sparks, J. Chorover, A. Aufdenkampe
Building Critical Zone Research Cyberinfrastructure

Critical Zone Exploration Network Data and Information Systems Workshop; State College, Pennsylvania, 17-18 September 2007

Biological, physical, and chemical processes transform bedrock and sediments into soil at the Earth’s surface. All terrestrial life on Earth is supported in the aptly named “critical zone” (CZ), where air, water, rock materials, and biota interact. The CZ is bounded at the top by the vegetative canopy and at the bottom by the lower limits of groundwater. Processes within this zone regulate the transformation of materials, solubilize nutrients for biota, buffer toxicants, create water pathways, and ultimately sculpt the landscape on which we live. Forty scientists from many disciplines attended a workshop recently at Pennsylvania State University to discuss needs for data and information systems to investigate the CZ.

This workshop grew from an international initiative for scientists interested in the CZ (Critical Zone Exploration Network, or CZEN). The U.S. National Science Foundation (NSF) has recognized the importance of CZ research with support for eight seed sites within CZEN (www.czen.org) and three proposed critical zone observatories (CZO).

Workshop attendees included representatives of the CZOs recommended for NSF funding, CZEN seed sites, the CZEN steering committee, NSF, cyberinfrastructure specialists, and CZEN students funded by NSF to work abroad. Workshop attendees primarily addressed the question, What measurements should be made at all CZ sites to allow cross-site comparison and better understanding of the CZ? In particular, the group considered measurements to address questions identified in earlier CZEN workshops:

1. How do processes in the critical zone control fluxes of carbon, particulates, and trace gases between land and atmosphere?

2. How do biogeochemical processes at critical zone interfaces govern long-term sustainability of soil and water resources?

3. How do processes in the critical zone that support and nourish ecosystems change over geologic and human timescales?

4. How do weathering processes affect the establishment of the critical zone, and how is this weathering engine perturbed by global environmental change?

Working groups each considered data needs for one of the science questions assuming a hypothetical site network receiving generous funding. The groups converged on a list of ~50 measurements spanning from chemical to meteorological to hydrological to biological (see list at http://www.agu.org/eos_elec or http://www.czen.org/17Sep07_results). Conducting all measurements at all sites is beyond the capabilities of the NSF CZOs and the CZEN seed sites. It remains for the community to prioritize and focus these measurements. Significantly, however, the measurement “wish lists” from the working groups that addressed different science questions showed substantial overlap.

The workshop was promoted by CZEN, a growing network of people, sites, tools, and ideas investigating dynamics of the critical zone. CZEN is developing an ontology for a consistent metadata and cyberinfrastructure system. Further information concerning CZEN and this workshop can be found at www.czen.org.

The full text of this meeting report can be found in the electronic supplement to this Eos issue (http://www.agu.org/eos_elec/).

—Michael Hopwood, Critical Zone Exploration Network, Ann Arbor, Mich. Email: mhop@umich.edu; Daniel Richter, Duke University, Durham, N.C.; Doug Miller and Susan Brantley, Pennsylvania State University, University Park.
Environmental scientists observe the natural system, develop sensor networks, share data sets, and model data to produce knowledge.

For example, in Critical Zone science we study earth’s surface today in order to project how it will look and act tomorrow.

Modelling the CZ

Interpreting the Record
At the first National CZO meeting, Shale Hills CZO presented our decision to “bag and tag” our data...in other words, we recognized that there are many, many types of data and we did not know how to organize it. So, we decided that each subdiscipline would “bag” their data up in some sort of rudimentary way and “tag” it in some sort of rudimentary way, so that we could get it online well before the cyberinfrastructure was built for each type of data.

EarthChem has recently decided that they want to put data into their EarthChem cyberinfrastructure but that this takes time...so, in the interim, before data goes into the cyberinfrastructure, they now accept datasets and registered them and place them on line (i.e. datasets are bagged and tagged)
CZO data types (trains) moving down different tracks at different rates

Hydrological time series

LIDAR data

Geochemical data

Geophysical data

Microbiological data

Geomorphological data

Biological data
We have worked on two different data models (more are needed)

Sensor-based data

- CUAHSI Hydrologic Information Systems
- Most of the HIS data is time series data
- CZ time series data should go into the HIS
- Pluses: wonderful team of workers in CUAHSI HIS + HydroDesktop and other tools already available
- Negatives: not really set up for geochemical data

Sample data

- EarthChem
- CZChem.db
- Most of the CZ geochemical data is spatial data (x,y and depth) and not time-series
- Pluses: set up for geochemical data; EarthChem group very very helpful and welcoming
- Negatives: we are still slogging through decisions about metadata and changes are still being made; only a small team is working on this

A few datasets have some of the characteristics of both, i.e. time series and spatial (x,y, and depth) – for example soil porewaters – and are not necessarily adequately handled by either data model. We are planning to put porewater data into CZChem.db
Major Components of CZchemDB

- **Main Data:**
  - Location/Site info – Geo-info, climate, landuse, etc.
  - Sampling info – Time, methods, treatment/preparation, etc.
  - Data – Chemical, physical and mineral properties and others

- **Meta Data**
  - Methods – Sampling, preparation and lab-analysis, etc.
  - Data quality – Precision, StDev, detection-limit, etc.
  - Source – Publication, projects, or contributor etc.
  - Lookup tables (controlled vocabulary) – Variables, units, standard etc.

- **Data sources**
  - CZO Observations;
  - Other published data of U.S.
  - Data from European Countries;
CZchemDB Data Entrance

- An Excel template file was developed for data entrance;
- Tools were developed to prepare, import, and append new data sets to the CZchemDB.

EXCEL Template
A database for soil and porewater chemistry. Online at Shale Hills website.

CZChem.db Template for Data Input (v.9.04)
Developed by X. Niu, J.Z. Williams, S.L. Brantley
(Penn State University, 2012)

Important Notes

1. You must enable the macros to utilize this template.
2. About the data list in the name box (upper left):
   - Click the Name-Box (uppermost left corner box above column A of worksheet). You will see a drop-down list of the list-names;
   - Click on any list-name (e.g. myUnitCode), you will be taken to the worksheet and you can scroll to see more information about the list.
3. About the drop-down lists:
   - If a cell has a drop-down list then you can only choose from that list;
   - If you don’t find your variables/phrases on the list, select “Other,” specify in Note column any info that is needed. You may also email jzw126@psu.edu to request an addition.
CZchemDB Application (1) – Data Searching

- Data searching
  - Location;
  - Geochemical element;
  - Project
- Data reporting
  - Formatted reports;
  - Excel file or text file

CZEN DB - dataCOMB

<table>
<thead>
<tr>
<th>COUNTER</th>
<th>STATE</th>
<th>LOCATION</th>
<th>PEDON_ID</th>
<th>DEPTH_TOP</th>
<th>DEPTH_BOT</th>
<th>HORIZON</th>
<th>TEXTURE</th>
<th>AIRO</th>
<th>CAO</th>
<th>MgO</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01</td>
<td>01</td>
<td>01</td>
<td>01</td>
<td>10</td>
<td>20</td>
<td>B1</td>
<td>B1</td>
<td>7.01</td>
<td>8.59</td>
<td>0.12</td>
</tr>
<tr>
<td>02</td>
<td>02</td>
<td>02</td>
<td>02</td>
<td>10</td>
<td>20</td>
<td>B2</td>
<td>B2</td>
<td>8.40</td>
<td>6.35</td>
<td>0.30</td>
</tr>
<tr>
<td>03</td>
<td>03</td>
<td>03</td>
<td>03</td>
<td>10</td>
<td>20</td>
<td>B3</td>
<td>B3</td>
<td>7.43</td>
<td>5.20</td>
<td>0.11</td>
</tr>
<tr>
<td>04</td>
<td>04</td>
<td>04</td>
<td>04</td>
<td>10</td>
<td>20</td>
<td>B4</td>
<td>B4</td>
<td>11.23</td>
<td>4.20</td>
<td>0.12</td>
</tr>
<tr>
<td>05</td>
<td>05</td>
<td>05</td>
<td>05</td>
<td>10</td>
<td>20</td>
<td>B5</td>
<td>B5</td>
<td>11.24</td>
<td>4.34</td>
<td>0.14</td>
</tr>
<tr>
<td>06</td>
<td>06</td>
<td>06</td>
<td>06</td>
<td>10</td>
<td>20</td>
<td>B6</td>
<td>B6</td>
<td>11.24</td>
<td>4.33</td>
<td>0.14</td>
</tr>
<tr>
<td>07</td>
<td>07</td>
<td>07</td>
<td>07</td>
<td>10</td>
<td>20</td>
<td>B7</td>
<td>B7</td>
<td>11.24</td>
<td>4.33</td>
<td>0.14</td>
</tr>
<tr>
<td>08</td>
<td>08</td>
<td>08</td>
<td>08</td>
<td>10</td>
<td>20</td>
<td>B8</td>
<td>B8</td>
<td>11.24</td>
<td>4.33</td>
<td>0.14</td>
</tr>
<tr>
<td>09</td>
<td>09</td>
<td>09</td>
<td>09</td>
<td>10</td>
<td>20</td>
<td>B9</td>
<td>B9</td>
<td>11.24</td>
<td>4.33</td>
<td>0.14</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td>B10</td>
<td>B10</td>
<td>11.24</td>
<td>4.33</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Data Report

C2CHEMDB
CZchemDB Application (2) – Data mining and modeling

- Data application
  - Tau calculations;
  - Other calculations;
- Future applications:
  - Data mining;
  - CZ modeling;
First publication describing CZChem.db

CZChemDB and EarthChem: Advancing management and access of critical zone geochemical data

Xianzeng Niu, Kerstin A. Lehnert, Jennifer Williams, Susan L. Brantley

Earth and Environmental Systems Institute, Pennsylvania State University, University Park, PA 16802, USA
Lamont-Doherty Earth Observatory, Columbia University, 61 Route SW, Palisades, NY 10964, USA

Article history:
Available online xxxx

ABSTRACT

Multiple Critical Zone Observatories (CZO) have been established in recent years in the USA and other international settings to conduct collaborative research on processes that occur at and near Earth’s surface, also known as the Critical Zone (CZ). Data documentation and data sharing are two persistent problems facing the CZOs that impede the ability for cross-site comparisons and integrated analysis. In this study, a relational database was developed for CZ rock and regolith geochemical data – CZChemDB. There are a total of 24 interrelated tables in the database, each representing different aspects of CZ features. The main data group includes tables of locations, sites, samples, sub-samples, preparation/treatments, laboratory-analysis and data values. The meta-data group includes tables of methods, references, and data quality. Lookup tables (variables, units, methods, etc.) contain list of “controlled” vocabularies. The CZChemDB is currently implemented in the MS Access database management system but the authors are moving to incorporate it into the EarthChem data system for broader online accessibility and usability. This integration also complements the EarthChem’s global geochemical database with CZ regolith data. The structure of the CZChemDB is simple, straightforward, and flexible so that it has potential to accommodate other chemical data collected from CZOs, such as pore fluid data. Furthermore, the development of CZChemDB represents the first attempt toward the standardization of geochemical data documentation and data sharing among CZOs. This effort will establish a model to bridge the connections between data acquisition, data management, data sharing, and data searching/discovering that are all essential but weak in terms of linkages within most geoscience research projects.
Toward standardized measurements of regolith-water interaction in the vadose zone at Critical Zone Observatories: The Use of Lysimeters to Measure Soil Porewaters

A paper in preparation to be submitted to Chemical Geology

(Alphabetical Order):

Introduction

To understand the Critical Zone requires measurements across gradients in environmental variables. Toward that end, several Critical Zone Observatories have been set up around the world. However, little effort has focused upon choosing standardized tools or approaches for measurement or sample collection. One reason for this is that several porewater-sampling methods are used and each method has utility for different systems. One way to promote standardization is to first elucidate the techniques that are currently in use in research watersheds. In this paper, we summarize the porewater sampling methodologies that are used at Critical Zone Observatories and satellite sites. We also collate a subset of the porewater chemistry data into simple plots that show variations in porewater chemistry as a function of climate, rock type, and ecosystem.
Current status of CZChem.db

CZchemDB is currently implemented as a MS Access database with no web-based interactions. Data submitted to CZchemDB are currently ingested into the database by data managers at Penn State. As of July 31 2011, 21 investigators from 11 institutions have contributed data to CZchemDB for 46 different field locations of 3 CZOs (Shale Hills, Luquillo, Jemez River Basin – Santa Catalina Mountains) and one international site (Plynmimon, Wales, UK). Total contributions thus far represent 255 cores collected and 2237 samples analyzed with a total of 33,142 analytical values. Ten additional template contributions are awaiting ingestion into CZchemDB.
Conclusions

• The different types of CZ data require different data models and a lot of work to organize

• Each subdiscipline needs to do the work for their data type, and the different subdisciplines are progressing at different rates

• This is hard work...it is also a lot of work for cyber folks to build the cyberinfrastructure (and it is cool), but the real difficulty is getting the disciplinary scientists to do the work to agree and set up their data models

• CZChem.db is well along the track: the goal is to put the database into EarthChem