Preview of Award 1332257 - Annual Project Report

Cover

Federal Agency and Organization Element to Which Report is Submitted:	4900
Federal Grant or Other Identifying Number Assigned by Agency:	1332257
Project Title:	Integrated Data Management System for Critical Zone Observatories
PD/PI Name:	Anthony K Aufdenkampe, Principal Investigator
Submitting Official (if other than PD\PI):	N/A
Submission Date:	N/A
Recipient Organization:	Stroud Water Research Center
Project/Grant Period:	12/01/2012 - 03/31/2014
Reporting Period:	04/15/2012 - 03/31/2013
Signature of Submitting Official (signature shall be submitted in accordance with agency specific instructions)	N/A

Accomplishments

* What are the major goals of the project?

The "Integrated Data Management System for Critical Zone Observatories" (CZOData) Project was funded based on a solicited proposal submitted Aug. 1, 2011 with an originally proposed start date of Oct. 1, 2011. However due to funding delays at NSF, and then an unforeseen transfer of the award to a new PI and institution initiated in Aug. 2012 and followed by transfer delays, funding for this project was not officially awarded until Mar. 7, 2013. Despite this very significant delay in funding, all senior personnel and institutions involved in this project (with the exception of the original PI, M.W. Williams, Univ. Colorado at Boulder) remained highly committed to project goals, volunteering significant time and effort for over a year before NSF was able to re-award the project. However, without firm funding commitments in place before Mar. 7, 2013, most institutions would not allow the team to engage any technical staff toward product development. The project team back-dated the start date to Dec. 1, 2012 to allow the invoicing of any expenses incurred in autumn of 2012. Therefore, this report document both those accomplishments completed on on a volunteer basis between Oct. 1, 2011 and Dec. 1, 2012, and accomplishments since the revised official project start date.

The originally proposed goal of the CZOData project is to develop an integrated data management system for the NSF-funded Critical Zone Observatories (CZOs), to ensure preservation and open data sharing of data acquired by the CZOs. The proposed work on the CZOData system builds off of a prototype CZO data system developed in late 2009 to early 2011 with $260k from a $500k supplement to Boulder Creek CZO (the remaining $240k was distributed to the original three CZOs to support their local data management). The CZOData system is designed to leverage existing NSF investments in CI, and engage CZO community. "Community engagement and Project Management" has been considered an overarching Objective (i.e. Task 0) that permeates all other Objectives and tasks.

The major CZOData Tasks are:

Task 0: Community Engagement and Project Management. Instigate and support a cross-CZO Information Management Committee (IMC) through online and face-to-face meetings and workshops.

Task 1: Website. Create a CZO-wide information portal, as the main communication platform between the CZO program and the user community

Task 2: CZO Display File Format v2. Develop consistent metadata profiles for other CZ domains, by extending the CZO Display File Format beyond hydrologic metadata best practices developed during the CZOData prototype phase

Task 3: Data Publication Tools. Create data and metadata publication tools and templates, to streamline data publication into the portal

Task 4: Web CZchemDB. Web-based user interface to CZchemDB.

Task 5: Shared Vocabulary. Enhance the shared vocabulary system, to enable cross-CZO search based on parameter names

Task 6: Sample Tracking. Implement an International GeoSample Number (IGSN) Registration Agent for CZO samples, to tracking unique sample identifiers for CZO data

Task 7: CZO Data Manager Support. Support site data managers with data publication, as well as identifying, analyzing and prioritizing their information management needs

Task 8: Central Catalog. Develop the CZO Central metadata catalog, data discovery portal and harvester, extending it to additional CZO data types, and integrating it with the CZO-wide information portal

Task 9: Central Repository. Enhance the central data repository of time series and point data.

Task 10: Geospatial Data. Extend the CZOData system to include publication and sharing of spatial data and LiDAR.

Task 11: EarthChem Integration. Integrating CZchemDB data into the EarthChem system

Task 12: Visualization Tools. Developing a CZO Visualization system (CZO-VS)

Since early 2011, before this proposal was submitted, the CZOData project team recognized that solutions to most of these tasks would be constrained by the very different information models used by the different CZ domains and the existing cyber-infrastructures that we would be leveraging. An information model is an abstract representation of concepts, their properties and the relationships, constraints and operations between them. It implicitly or explicitly forms the conceptual foundation for each component of a data system. Therefore, the CZOData team realized that solutions to many of our challenges would be both more straightforward and more durable if we could develop a common information model for all CZOData components. However, this seemed outside the scope of the CZOData project when we submitted it. Six months later, the CZOData team submitted a separate proposal to address this need, and in Aug. 2012 the project was funded: "Developing a Community Information Model and Supporting Software to Extend Interoperability of Sensor and Sample Based Earth Observations" (NSF #1224638). The proposed Observations Data Model version 2.0 (ODM2) would become the foundation for all CZOData infrastructure and others (i.e. CUAHSI and IEDA/EarthChem). Since Sep. 2012, we have been co-managing the CZOData and ODM2 projects, and in many cases have prioritized some ODM2 tasks over CZOData tasks for the first year, given that ODM2 will be foundational to CZOData.

* What was accomplished under these goals (you must provide information for at least one of the 4 categories below)?

Major Activities: **Task 0: Community Engagement and Project Management.**

In Feb. 2012, PI Aufdenkampe presented a cyber-seminar to the CZO community, providing a "Data Management 101" and an overview of the pending CZOData proposal.

In April 2012, PI Aufdenkampe began leading, before having any official role in CZOData, an effort to restart and complete work on the National CZO Website (http://criticalzone.org). This began a year-long series of weekly web conference

(WebEx) meetings with 1-2 data managers from each of the six CZOs, coupled with monthly feedback from CZO PIs. Although website development was the initial motivator, these weekly meetings included discussion of a wide range of topics and issues relevant to the day-to-day data management by each CZO. These meetings continued to be weekly until June 2013.

In September 2012, PI Aufdenkampe started leading weekly WebEx meetings with the CZOData Developer Team (i.e. all senior personnel on the award), beginning about a month after receiving notice from the original PI, M.W. Williams, of his intent to step down and transfer the award. These meetings focused on making high-level decisions on project priorities and design approaches and technologies. These meetings continue to be held weekly to the present and will continue to the end of the project.

In Oct. 2012, PI Aufdenkampe secured NSF EarthCube funding for a Critical Zone Domain workshop (proposal submitted July 2012, with workshop scheduled) to further enhance the development of community needs and community understanding of possible solutions. This began the use of EarthCube as platform for communication and feedback.

In Nov. 2012, PI Aufdenkampe formed, with input from CZO PIs, the Information Management Committee (IMC), which includes the CZOData development team, 1-2 data management staff from each CZO, and 1-2 faculty investigators from each CZO who can represent the cyber-needs of different CZ science disciplines. The inclusion of investigators -- who are not necessarily the PIs -- was designed to enable a more integrated partnership.

On Dec. 11-12, 2012, the CZOData team held the first face-to-face IMC workshop, hosted by Aufdenkampe at SWRC (http://criticalzone.org/national/events/event/2012-11-13-czodata-imc-meeting/). 27 peopled attended; 16 in person and 11 remotely via WebEx. The workshop included a series of presentations by the CZOData developer team to introduce the IMC to data management concepts, the existing CZOData prototype, the proposed CZOData project phase 2, and the cyber-infrastructures and technologies that CZOData would be leveraging. The second day of the meeting was devoted to discussions toward making decisions on: (1) an approach to uniformly present our data files on our new criticalzone.org website; and (2) developing subcommittees to explore adaptation of existing metadata standards relevant to each CZO discipline. Meeting agenda, notes and participates are attached.

On Jan. 21-23, 2013, the CZOData team convened and hosted the CZ-EarthCube Domain workshop at the University of Delaware. The detailed agend and links to reports can be found at http://criticalzone.org/national/events/event/2013-01-21-earthcube-czo-workshop/.

Task 1: Website.

On Nov. 30, 2012 the new integrated CZO website was launched (http://criticalzone.org). This website contained separate subsections for each CZO and the national program page, allowing for consistent structures and theming and for lateral navigation between CZOs on the same topic. The design of the website structure, organization and topic headings was very carefully thought-out and reviewed by a committee of participants from all the CZOs through weekly WebEx meetings devoted to Website development (see Task 0) The new website has been developed on the Expression Engine content management system, which enabled a

high degree of integration between different website components.

On June 3, 2013 we launched the data pages section of the integrated CZO website (i.e. http://criticalzone.org/christina/data/ etc.). The original scope for the integrated website was to exclude data pages, allowing each CZO to present their data files on their own website as they chose, as long as the data were ingested into the CZO Central Catalog and Repository. However, the CZOs had received feedback that enough people still searched for data via "old-fashioned" web browsing that it was important to have a uniform web layout of these pages to facilitate this searching. To avoid duplication of efforts that might require data managers to fill out metadata forms for both the website and the CZO Central Catalog, the IMC developed a plan at the Dec. 11-12 meeting to use the website content management system (CMS) database as the primary means for logging files into the CZO Central Catalog. We thus developed the CMS web form to follow an enhanced version of the "Dublin Core" metadata standard and are presently working on automatic export of that information to the CZO Central Catalog.

Although we will continue to tweak and enhance the national CZO website system, we consider this task (i.e. development of the website *system*) to be largely completed. Data managers and CZO investigators will of course continue to actively feed new content into the system.

Specific Objectives: **Task 2: CZO Display File Format v2.**

The present CZO Display File Specification (v1) is available at https://criticalzone.org/images/national/associated-files/1National/CZO_Display_File_Format_Specifications_2011-01-31.pdf. The CZOData Team and IMC has been compiling a list of requirements for the next version of the CZO Display File format. However, moving forward on finalizing that format requires that we first finalize ODM2 design, which will provide a consistent metadata profile for managing both time series and sample data. We expect to finalize the ODM2 conceptual design by Sept. 2013.

Task 3: Data Publication Tools.

This task is also pending a final ODM2 design, expected Sept. 2013.

Task 4: Web CZchemDB.

The CZchemDB database has been finalized by the Penn State team who developed it and sent to the IEDA (Columbia) team to translate from the prototype desktop implementation in Microsoft Access to an enterprise server implementation on a Linux/Appache web server and using a PostgreSQL Relational Database Management System. The translation process is actively underway at Columbia.

Task 5: Shared Vocabulary.

In 2012, the existing CZO Shared Vocabulary system (http://sv.criticalzone.org/) was moved by USU to a new set of servers along with the switch of criticalzone.org to SDSC. However, major enhancing the existing CZO Shared Vocabulary system (http://sv.criticalzone.org/) has been set to be completed in year 2, after final development of ODM2.

Task 6: Sample Tracking.

Developing an IGSN registration agent within the CZO website has been set to be completed in year 2, after final development of the CZO data website (launched in June 2013).

Task 7: CZO Data Manager Support.

Weekly online meetings held with site data managers, to discuss issues of data portal and catalog construction, as well as needs and priorities of the CZO sites. Support CZO data managers in preparing for CZO renewals in February 2013. In december 2012, the CZOData team organized a meeting with CZO data managers at the Stroud Water Research Center, to create a joint coordinated development strategy for CZOData and individual CZO data management systems. The outcome of this meeting was a renewed effort to load data into the CZO central system by each of the data managers.

Tasks 8-9: Central Catalog & Repository.

The SDSC team worked with data managers to load 111 new data files into the CZO central system within the past 8 months. The new data files were automatically registered in the CZO central metadata catalog. The search.criticalzone.org/geoportal site now has 468 records, from all sites.

Coordination between the central metadata catalog and the CZOData information portal has been one of the development priorities of the project. The SDSC and UCBoulder teams worked with together to export metadata from the central metadata catalog, implemented as Geoportal Opensource system, into a format compatible with the new content management system used as the foundation for the CZO information portal. Strategies are now in place for the future synchronization between the CZO information portal and the geoportal, as additional display files are ingested harvested and ingested into the system.

Task 10: Geospatial Data.

The SDSC team coordinated with OpenTopography developers to make LiDAR datasets discoverable through the CZO catalog search at http://search.criticalzone.org. This added 155 additional records to the CZOdata catalog.

Task 11: EarthChem Integration.

Integrating EarthChem tools into CZOData has been set to be completed in year 2, after final development of the CZO data website (launched in June 2013) and ODM2 (to be completed in Sept. 2013).

Task 12: Visualization Tools.

In 2012 the UW-APL team worked closely with the Christina River Basin CZO team to develop a prototype CZO visualization system (http://wikiwatershed-vs.org) based on the Northwest Association of Networked Ocean Observing Systems (NANOOS) Visualization System (NVS, http://nvs.nanoos.org/Explorer). The major development demonstrated by this prototype was the use of CUAHSI Hydrological Information System (HIS) Web Services to feed USGS and Delaware Environmental Observing System (DEOS) into a unique NVS instance. This was the first known web-client of CUAHSI HIS Web Services.

In 2013, UW-APL has developed a framework within the recently launched NVS 3.0

to allow NVS to be deployed in many instances with CZO rebranding. These significant upgrades will be test launched for Christina River Basin CZO, including their streaming sensor network data, in Aug. or Sept. 2013.

Significant Results:

National CZO Website (http://criticalzone.org)

Key outcomes or
Other achievements:

National CZO Website data pages (i.e. http://criticalzone.org/christina/data/ etc.).

CZO Display File Specification (v1, https://criticalzone.org/images/national/associated-files/1National/CZO_Display_File_Format_Specifications_2011-01-31.pdf)

CZO catalog search at http://search.criticalzone.org

CZO visualization system (http://wikiwatershed-vs.org)

CZ-EarthCube Domain workshop agenda and reports, http://criticalzone.org/national/events/event/2013-01-21-earthcube-czo-workshop/.

* What opportunities for training and professional development has the project provided?

The CZO IMC specifically targeted early-career cyber-savy faculty to serve as investigator representatives of the data needs for their CZO. Likewise, the CZ-EarthCube workshop targeted early career participants. These were intentional choices to provide a training and professional development opportunities.

* How have the results been disseminated to communities of interest?

Results have been disseminated through cyber-seminars, web conferences, websites, workshops and workshop reports. See Accomplishments above.

* What do you plan to do during the next reporting period to accomplish the goals?

See the task-by-task descriptions of accomplishments for mention of tasks set for completion in project year 2.

Products

Journals

Books

Book Chapters

Thesis/Dissertations

Conference Papers and Presentations
Zaslavsky, I., T. Whitenack, M. Williams, D. G. Tarboton, K. Schreuders, and A. Aufdenkampe (9/28/11). *The Initial Design of Data Sharing Infrastructure for the Critical Zone Observatory*. Environmental Information Management Conference, EIM 2011. Santa Barbara, CA.

Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Other Publications

Technologies or Techniques
Nothing to report.

Patents
Nothing to report.

Inventions
Nothing to report.

Licenses
Nothing to report.

Websites

Title: National CZO Website

URL: http://criticalzone.org

Description: National CZO Website data pages

Title: http://criticalzone.org/christina/data/

URL: An example of the new integrated data pages for each CZO.

Description: CZO catalog search

Title: http://search.criticalzone.org

URL: CZO Shared Vocabulary system

Description: http://sv.criticalzone.org/

Title: CZO visualization system prototype

URL: http://wikiwatershed-vs.org

Description:

Title:

URL:

Description:

Other Products
Nothing to report.

Participants

Research Experience for Undergraduates (REU) funding

What individuals have worked on the project?

Name	Most Senior Project Role	Nearest Person Month Worked
Anthony K Aufdenkampe	PD/PI	2
Ilya Zaslavsky	Co-Investigator	1
Kerstin Lehnert	Co-Investigator	1
Jeff Horsburgh	Co-Investigator	1
David Tarboton	Co-Investigator	1
Emilio Mayorga	Co-Investigator	1

David Valentine	Co-Investigator	1
Thomas Whitenack	Other Professional	1
David Lubinski	Other Professional	4
Heather Brooks	Other Professional	2

What other organizations have been involved as partners?

Name	Location
CUAHSI	Boston, MA

Have other collaborators or contacts been involved? Y

Impacts

What is the impact on the development of the principal discipline(s) of the project?

No other project that we know of has yet integrated data from such a broad range of disciplines represented by the CZO community. As a result, we proposed to the EarthCube community -- which has such a goal -- that the CZOData project might represent a pilot of the EarthCube endeavor. See Aufdenkampe's blog post at: http://earthcube.ning.com/profiles/blogs/critical-zone-domain-pilot-for-earthcube?xg_source=activity.

What is the impact on other disciplines?

CZ science itself represents 16 geoscience disciplines, and by leading by example within the larger EarthCube community we can have an even broader impact. CZOData project team members have given over a dozen invited talks on experiences and accomplishments from CZOData, which is a testimony to our broad impact on other disciplines.

What is the impact on the development of human resources?

We have focused significant efforts in training data managers and investigators from all six CZOs, and have placed a significant effort on engaging early career scientists. Participants in all our activities have provided very positive feedback on the value and effectiveness of our project and it's outreach activities.

What is the impact on physical resources that form infrastructure?
Nothing to report.

What is the impact on institutional resources that form infrastructure?

The intense effort to develop a complex, modern, integrated website has elevated the web-development capabilities of the Stroud Water Research Center.

What is the impact on information resources that form infrastructure?

The purpose of this project is to develop information infrastructure that can be a resource to all CZOs and their partner institutions.

What is the impact on technology transfer?

A secondary purpose of this project is to develop within the CZO community the skills to use the technology and

information infrastructure that we will develop.

What is the impact on society beyond science and technology?

Sustainability of our natural resources requires effective and efficient information and data sharing, not only between scientists but also for use by resource managers, citizen-scientist volunteers, educators, students, the media and the general public.

Changes

Changes in approach and reason for change
Nothing to report.

Actual or Anticipated problems or delays and actions or plans to resolve them
Nothing to report.

Changes that have a significant impact on expenditures
Nothing to report.

Significant changes in use or care of human subjects
Nothing to report.

Significant changes in use or care of vertebrate animals
Nothing to report.

Significant changes in use or care of biohazards
Nothing to report.