Progress and Timeline for Deliverables Report for the

Integrated Data Management System for Critical Zone Observatories (CZOData) Project (NSF EAR-1332257)

Submitted February 27, 2015 (v1)

Revised March 13, 2015 (v2)

Anthony Aufdenkampe (PI, Stroud Water Research Center, SWRC),
Jeff Horsburgh (Utah Center for Hydrologic Information and Computing, UCHIC)
Kerstin Lehnert (Interdisciplinary Earth Data Alliance, IEDA)
David Lubinski (University of Colorado at Boulder, CU-Boulder)
Emilio Mayorga (University of Washington Applied Physics Laboratory, APL)
Ilya Zaslavsky (San Diego Supercomputing Center, SDSC)

Introduction

The "Integrated Data Management System for Critical Zone Observatories" (CZOData) Project has the overall goal to develop a comprehensive, integrated data management system for the Critical Zone Observatory (CZO) program with capabilities to share, integrate, analyze and preserve the wide range of multi-disciplinary data generated within and across CZOs.

The CZOData Project was funded based on a proposal submitted Aug. 1, 2011 by lead PI Mark W. Williams, (Univ. Colorado at Boulder). The two-year award (#1153164) started on April 15, 2012, but was transferred to Anthony Aufdenkampe (Stroud Water Research Center) when Williams stepped down as PI in July 2012. The award transfer was finalized on March 7, 2013 (new award #1332257) with an effective end date of March 31, 2015. We have requested an extension to March 31, 2016.

A number of opportunities arose in the 20 months between proposal submission and final funding, in particular opportunities to align our development with the Observations Data Model 2.0 (ODM2) project, with the NSF EarthCube initiative, and with second generation cyber-infrasture for both the CUAHSI Water Data Center and the Interdisciplinary Earth Data Alliance (IEDA) data repository. In addition, the needs of the CZO network evolved, including the addition of four new CZOs in June 2013.

These opportunities and changing needs led to a number of changes to the project timeline and scope. In this progress report we provide details on each of the proposed tasks, including the original task description, any changes to scope, progress to date, and expected delivery date. We have requested a no cost extension to complete deliverables by March 31, 2016.

The 2011 CZOData Proposal listed 12 Tasks, which were expanded in the 2012 CZOData Proposal Addendum into **13 Tasks that each had 2-4 SubTasks**. We originally structured

our tasks according to whom the task was assigned to rather than on deliverables. To improve communication, here we restructure our tasks around major deliverables, but retain our Task numbering used in the 2012 CZOData Proposal Addendum. A CZOData Project Supplement was funded in Dec. 2013 to expand the CriticalZone.org website to host four new CZOs and to unify web browser presentation of CZO Datasets hosted at individual CZOs. In this report we also provide details on each of these new tasks.

The **Year 1 Project Report** detailed accomplishments from Oct. 1, 2011 to March 31, 2013. The prolonged project transfer process limited project spending during this first reporting period. We back-dated the start date to Dec. 1, 2012 to allow the invoicing of any expenses incurred in autumn of 2012 (such as the Dec. 2012 workshop). Therefore, the Year 1 report documents accomplishments completed on a volunteer basis between Oct. 1, 2011 and Dec. 1, 2012, and the funded accomplishments from Dec. 1, 2102 to March 31, 2013.

This **Year 2 Project Report** describes accomplishments for the first full year the project was actively funded, from March 31, 2013 to March 31, 2014. Because of funding delays, we were granted our first no-cost extension to March 31, 2015. Therefore, the project team considers that this Year 2 Project Report documents the midpoint of our project.

Table of Contents for Deliverables and Associated Proposed Tasks

Originally proposed tasks and subtasks have been grouped into the nine major deliverables. Task numbers have been maintained to match those in the 2012 CZOData Proposal Addendum. The following listing can be used as a hyper-linked table of contents to the detailed progress report below.

○ Task 10: Geospatial Data

Observations Data Model 2.0 (ODM2)

Since early 2011, before this proposal was submitted, the CZOData project team recognized that solutions to many of our data integration challenges would be more straightforward and more durable if we could develop a common information model that would work equally well for all CZOData system components and for nearly all CZ science disciplines. In Aug. 2012 the CZOData team was separately funded to do just that, to develop "A Community Information Model and Supporting Software to Extend Interoperability of Sensor and Sample Based Earth Observations" (NSF #1224638), which we named the Observations Data Model 2.0 (ODM2). Our project team, which is identical for both CZOData and ODM2 projects, succeeded beyond expectations with this ambitious goal. ODM2 facilitates data management and interoperability to such an extent that our partnering data centers, the CUAHSI Water Data Center and the Interdisciplinary Earth Data Alliance (IEDA), both began in 2014 to completely rebuild their cyber-infrastructures around ODM2. The enormous success of ODM2 provided an opportunity for the CZOData project, but to take advantage of ODM2 required that we put CZOData Tasks 2, 3, 4, 5 and 11 on temporary hold until April 2014 when ODM2 conceptual development was complete. The CZOData project team believes very strongly that these delays have been and will continue to be well worth the investment. This assessment is supported by our extensive list of invited presentations by the leadership of EarthCube, the Open Geospatial Consortium (OGC) and the Knowledge Network for Biocomplexity (KNB).

Progress on Deliverables and Associated Proposed Tasks

Community Engagement & Support

T0: Community Engagement and Project Management:

Proposed: This project is committed to engage in an ongoing, open and responsive dialog with the community, and especially with CZO investigators and Site Information Managers (SIMs), seeking their input and feedback on system capabilities, user interfaces, and content to ensure the system's utility and usability, and to maximize its value and impact for scientific research and education.

- **T0a: Information Management Committee**
  - **Proposed**: The IMC will identify critical issues, and will work with the project team to address CZOData challenges through the development of new information management approaches. Within the IMC, site scientists will be responsible for gathering feedback from their respective subdisciplines: Shale Hills for geochemistry; Boulder Creek for geophysics, Southern Sierras for spatial data, Jemez-Catalina for soils data, etc. The IMC will also be responsible for making decisions when conflicts arise around the shared vocabulary (SWRC, Q1-Q8).
  - **Changes to Scope**: In mid 2013, the IMC was expanded in size to include 10 CZOs rather than the proposed six. The increased costs of hosting a larger number of IMC members for annual workshops (Task 7b) and paying for a greater number of stipends

for faculty IMC members was offset by eliminating site visits to each CZO (Task 7a), as per agreement with NSF (phone call with Enriqueta Barerra on Nov. 12, 2013).

- **Progress to Date:** The IMC was formed in November 2012 and reformed for 10 CZOs in January 2014. IMC membership and communications have been organized and archived via the CZOData Information Management Committee (IMC) Google Group (https://groups.google.com/d/forum/czodata-imc). We have hosted two annual workshops (Task 7b) and an extensive number of web meetings and webinars (Task 7C). IMC subcommittees were formed at each of the two workshops to focus on feedback for various tasks (i.e. website and data page tasks in 2012-2013 and metadata tasks in 2014). All shared IMC documents, including workshop agendas, notes and materials have been publicly viewable within the CZOData-IMC Google Drive directory.
- **Expected Delivery Date:** The proposed task is completed as proposed, but engagement with the CZOData IMC will continue throughout the end of the project. In particular, we will use unexpended participant support costs to fund one to two additional IMC meetings (Task 7b) and an additional year of IMC investigator stipends to encourage active participation from faculty-level IMC members.

- **T0b: Web-based information events and workshops; mailing lists**
  - **Proposed**: The Information Management Committee, along with our executive team, will develop and support communications about data management, cybersecurity, and other timely and related computer and data security issues to CZO scientists, staff, and students through webcasts and associated web-articles developed for this purpose. Each year, the IMC will provide information including "how to" and "best practice" guidelines on network and computer security, as well as providing regular updates about progress and new system releases of CZOData via its web site and the CZEN mailing list. (SWRC, Q1-Q8)
  - **Changes to Scope**: None
  - **Progress to Date:** Communication with the IMC has included many phases of intensive engagement.
    - From April 2012 to August 2013, data and web managers from each CZO met for weekly WebEx meetings to help develop the CriticalZone.org website (4/2012 to 12/2012) and then the CriticalZone.org/Data section of the website (Task 1, 1/2013 to 8/2013). Over a dozen cyber-seminars were presented to CZO PIs to help make bigger decisions and to train CZO web managers and PIs to use the new CriticalZone.org Content Management System (CMS).
    - Video tutorials for how to use the CMS were recorded and posted to YouTube in a private channel for IMC members.
    - A CriticalZone.org YouTube channel (https://www.youtube.com/user/CriticalZoneOrg) was developed and maintained with funds from the CZOData project to consolidate and highlight video content from all CZOs on any topic.
    - From March to May 2014, we presented a series of seven CyberSeminars to the entire CZO network. These were recorded and are available at

> http://criticalzone.org/national/data/czo-data-cyber-seminars/ and at
> https://www.youtube.com/user/CriticalZoneOrg/playlists.

  - ○ From November 2014 to January 2015, we presented four additional CyberSeminars available at http://criticalzone.org/national/data/czo-data-cyber-seminars/ and at https://www.youtube.com/user/CriticalZoneOrg/playlists.
  - ○ **Expected Delivery Date:** Engagement with the CZOData IMC will continue throughout the end of the project, with extensive engagement planned throughout the last year of the project from the present to March 31, 2016.

- ● **T0c: Synthesis working group**
  - ○ **Proposed**: A synthesis working group will meet annually to discuss a topic of interest that is cross-CZO. The goal of the workshop is a manuscript on that topic that uses all available data and serves as a state-of-the-art synthesis on that critical zone topic. In turn, participants will provide the IMC with feedback on what data was available, what data was not available, metadata concerns, additional needs, etc. (CU-Boulder, SWRC, Q3 and Q7).
  - ○ **Changes to Scope**: Synthesis workshops were eliminated in cooperation with NSF Program Officer Enriqueta Barrera during rescoping when the project was transferred to SWRC in early 2013, in order to adequately fund the other five workshops that were originally under-budgeted, and to provide travel funds for additional CZOData Team members to attend those workshops (i.e. Jeff Horsburgh and Emilio Mayorga did not originally have travel budgets for workshops). In addition, the reduced scope acknowledged that seven workshops in two years was unnecessary and might lead to workshop fatigue by the project team and participants, especially in light of the intense workshop schedule from the ODM2 project (3 workshops from 2012-2014) and from EarthCube (dozens of workshops starting in Oct. 2012).
    - ● It made sense to eliminate the two synthesis workshops rather than the IMC and Sub-discipline workshops because of the intensive participation of the Critical Zone community in the CZ-Domain EarthCube workshop in January 2013, which resulted in valuable advances in the vision of CZ science and how it might benefit from a new cyber-infrastructure (see the EarthCube workshop exec summary-CriticalZone-FINAL Report).
  - ○ **Progress to Date:** N/A
  - ○ **Expected Delivery Date:** Eliminated from scope (see above).

- ● **T0d: Subdiscipline workshops**
  - ○ **Proposed**: IMC participants will convene additional workshops on subdisciplines to enhance skill sharing and communication among data managers and scientists. Three workshops: Geochemistry; GIS Data Management; Hydrologic Data Management. (SWRC, IEDA, Q2, Q5, Q8)
  - ○ **Changes to Scope**: These workshops were delayed to take advantage of development of deliverables around ODM2, in order to more constructively guide feedback. Also, it was decided that the specific topics of each workshop would be determined by the CZOData Project team and the IMC to best meet present needs.

- ○ **Progress to Date:** One of three workshops have been held to date:
  - ● A CZO Data Template & "Display File v2" workshop was held Feb. 9-11, 2015 at IEDA, to work with the IMC Display File sub-committee to finalize a specification for CZO Display Files that would allow for both hydrological time series datasets and geochemical specimen measurement datasets to be represented in a unified format. The 2015-2-9_CZODisplayfile-WorkshopAgendaAndNotes Google Doc is publicly viewable.
- ○ **Expected Delivery Date:** The two remaining workshops will be completed by summer of 2015 and winter of 2016 respectively.

T7: Support for Site Information Managers:

Proposed: We will support Site Information Managers (SIMS, including both data managers and web managers) as follows: 1) biweekly teleconferences with SIMs under the supervision of SDSC; 2) a document sharing web site; 3) all SIMs will be members of and participants in the annual IMC meeting and monthly telecons; and 4) individual 2.5 day site visits will be made to each CZO. These visits will let us formulate user requirements, prioritize development, and find areas of synergy between local scientists, SIMs and the cross-CZO data sharing efforts. In addition, it will allow us to jointly design, develop, install and troubleshoot local components of the data publication system, including forms and templates for creating metadata and software for generating display files in the uniform format defined by the project. The results of the site visits will be formalized in a report on data management and data sharing needs of CZO project partners and will be used to inform further CZOData design. (SDSC, UCHIC, CU-Boulder)

- ● **T7a: Conduct site visits to all CZOs**
  - ○ **Proposed**: Conduct site visits to all CZO sites, to understand specific science requirements and identify and develop recommendations for shared technological approaches and unified data publication protocols. (SDSC, UCHIC, CU-Boulder, Q1-Q2)
  - ○ **Changes to Scope**: Synthesis workshops were eliminated during rescoping in conversations with NSF Program Officer Enriqueta Barrera in Autumn of 2013, to allow for fully funding participants from the expanded CZO network (10 vs. 6 CZOs) to participate in the IMC and attend the other workshops.
  - ○ **Progress to Date:** NA
  - ○ **Expected Delivery Date:** Eliminated from scope (see above).
- ● **T7b: Conduct workshops for SIMS**
  - ○ **Proposed**: Conduct workshops for data managers, and establish continuous exchange of ideas, technologies and innovations through regular phone/web conferences and one-on-one communication. (SDSC, UCHIC, CU-Boulder, Q3-Q8)
  - ○ **Changes to Scope**: None
  - ○ **Progress to Date:** The two proposed annual IMC workshops have both been held:
    - ● The first IMC Workshop was held on Dec. 10-12, 2012, at SWRC and via WebEx. The 2012-12-11_IMC-WorkshopAgenda, presentations and other workshop

documents are publicly viewable in the 2012-12-11_CZOData-IMC-Workshop Google Drive folder.

- ● The first IMC Workshop was held on May 28-30, 2014, at SWRC and via WebEx. The 2014-5-29_IMC-WorkshopAgenda, presentations and other workshop documents are publicly viewable in the 2014-5-29_CZOData-IMC-Workshop2 Google Drive folder.

  - ○ **Expected Delivery Date:** The proposed task is completed as proposed, however the costs of the first two IMC Workshops, and associated stipends for faculty-level IMC members, were significantly lower than budgeted. This provides an opportunity to use unexpended participant support costs to fund an additional one or two IMC meetings and an additional year of IMC investigator stipends to encourage active participation from faculty-level IMC members.

- ● **T7c: Bi-Weekly VTC with SIMS**
  - ○ **Proposed**: Host bi-weekly video teleconferences (VTCs) with site data managers. (SDSC, Q1-Q8)
  - ○ **Changes to Scope**: None.
  - ○ **Progress to Date:** We maintained an active schedule of hosting VTCs, via WebEx, with data managers and web managers from the previous phase of the project to the present, as described in detail above for Task 0b.
  - ○ **Expected Delivery Date:** The proposed task is completed as proposed, but engagement with the CZOData IMC will continue throughout the end of the project.

- ● **T7d: Maintain file sharing portal**
  - ○ **Proposed**: Maintain a document and file-sharing portal for data managers (password protected). (SDSC, Q1-Q8)
  - ○ **Changes to Scope**: None.
  - ○ **Progress to Date:** The CZOData project has maintained several file sharing portals.
    - ● The first was developed primarily with prior funding to SDSC, and was used to share files among data managers as we converged on display file specification and ironed out bugs in formatting. The password-protected display file submission site (central.criticalzone.org) is still operational, and continues to process time series data submitted by data managers. In 2013, the front-end components of this system have been moved to the new CriticalZone.org Content Management System (CMS).
    - ● With the development of the CriticalZone.org website and Content Management System (CMS), CZOs gained an ability to share files of any sort with the public.
    - ● The CZOData Team and IMC members have shared files for collaborative work within the publicly viewable CZOData-IMC Google Drive directory.
  - ○ **Expected Delivery Date:** Completed development and continuing operational function.

CriticalZone.org Website

T1: CZO-Wide Information Portal:

Proposed: Develop, maintain, and improve the fully-integrated web system at CriticalZone.org, which is powered by a Content Management System (CMS). Improvements include integrating the CMS with the CZO Data Discovery [Search] Portal (see Task 8). Work on the CMS will be led by CU-Boulder, SWRC, and SDSC in cooperation with all CZOs.

Changes to Scope: Two major tasks were added to the scope of the project at different times after the start date: 1) expanding the role of the CMS to include Data & Dataset pages and 2) adding four new CZOs and their content to the website. Both tasks were made possible with additional support via a supplement for website work, as well as the formation of a CZO National Office. The National Office began contributing to maintenance of the CriticalZone.org website in mid 2014.

- **T1a: Develop the CZO online content management system (CMS)**
  - **Proposed**: Help complete the CMS and its initial content.
  - **Changes to Scope**: None.
  - **Progress to Date:** The CMS was mostly built through prior funding, but the present phase of CZOData project funding helped support the management and oversight of its final completion and website launch at the AGU National Meeting in December 2012. Work included coding of the CMS as well as assistance to individual CZOs for loading their initial content.
  - **Expected Delivery Date:** Completed in December 2012.
- **Supplemental Task A: Add Data and Datasets to the CZO CMS**
  - **Proposed**: This supplemental task was not in the original proposal, but requested by the National CZO Steering Committee and NSF in mid 2012 and funded in December 2013 via a CZOData Project Supplement and by reducing Task 1b.
  - **Changes to Scope**: The CMS at CriticalZone.org was originally designed to promote and display all CZO content except that related to data. The data content and data values were instead to be handled by individual CZOs on their own web servers at different web addresses (URLs). Some of those data values (i.e. time series data) and metadata for other datasets were to be ingested into a centralized system (central catalog and repository) and made visible on a separate CZO data discovery portal. However, the CZOs had received feedback through the National CZO Steering Committee and NSF 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. Moreover, it was clear that the majority of the data was not being centrally ingested and thus, was not as easily discovered. For example, some geochemical data in excel formats could not be ingested at that time, so could not be displayed in a consistent, discoverable manner across CZOs. 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. 2012 IMC meeting (see Task 7b) to use the website content management system (CMS) as the primary means for logging files into the CZO Central Catalog (not the repository). The actual data files would remain either on the websites of each CZO (i.e. Shale Hills CZO) or at an appropriate data repository (i.e. OpenTopography.org for LiDAR data).

- ○ **Progress to Date:** From January to August 2013, we collaborated with CZO web staff and investigators, via weekly WebEx meetings and extensive email exchanges (see Task 0b), to develop the back-end CMS database and forms and the front-end user interface of the CriticalZone.org Dataset pages. We also developed data sharing and data use agreements. We continued to work on improving the front-end interface and Dataset content until Aug. 2013. These systems are fully functional; CZO data managers are able to continually add new Datasets into the CMS system and the public can freely browse and download those datasets at http://criticalzone.org. We started working with the four new CZOs in Fall 2014 to input their content, but need to nudge them in Spring 2015. We are also working on redesigning the Data lander pages to be more useful.
  - ○ **Expected Delivery Date:** Main functionality was completed in August 2013.
    - ● We expect the four new CZOs to input content by Fall 2015. By that time, we also anticipate that we will launch a new layout for the Data lander pages.
- ● **Supplemental Task B: Add four new CZOs (Calhoun, Eel, IML, Reynolds) and their content to CZO CMS**
  - ○ **Proposed**: This supplemental task was not in the original proposal, but requested by NSF in autumn 2013 and funded in December 2013 via a CZOData Project Supplement and by reducing Task 1b.
  - ○ **Changes to Scope**: Four new CZOs were funded by NSF in summer of 2013, two years after the CZOData proposal was written. These CZOs were ramping up their efforts in the Fall of 2013, with much coordination at the AGU Meeting in December 2013. It was essential that all four be well represented on the CriticalZone.org website as soon as possible.
  - ○ **Progress to Date:** In January 2014 we began working with CZO web managers from the new CZOs, and in April 2014 we launched the CriticalZone.org websites for the four new CZOs and an updated new navigation system for the entire website that better enables navigation between now 10 different CZOs. Helping the four new CZOs develop and load content into the CriticalZone.org web system took three months of near weekly web-meetings and frequent email exchanges. As it was with the original six CZOs, getting complete content is the primary issue with preparing for launch.
  - ○ **Expected Delivery Date:** Main functionality and training completed in April 2014.
    - ● We will nudge some of the new CZOs to fill in content gaps by Fall 2015, however, adding content to CriticalZone.org web pages is the responsibility of individual CZOs, and now the CZO National Office.
- ● **T1b: Further iterate on the CZO CMS, add content**
  - ○ **Proposed**: Great websites are iterative. We recognize that nobody gets everything right the first time. Moreover, the web and the way we use it continually changes (i.e., the exploding use of the mobile and social web), and CZO itself continues to evolve. Given all this change, the best approach is to test the web site's effectiveness over time and continually make adaptations. Potential changes include making content easier to update, adding new features, and altering the wording of "calls to action" that we want the site visitors to follow.
  - ○ **Changes to Scope**: The addition of two substantial Supplemental Tasks to the proposed website work (adding datasets and expanding to include four new CZOs, as

9

described above) required that we decrease efforts elsewhere (even with the supplement).

- ○ **Progress to Date:** Most of our efforts centered on improving website speed, which was essential with the expansion of the website to encompass dataset pages as well as all the content for four new CZOs. Speed was improved in two major ways. First was by setting up a caching system and writing code that substantially reduced the number of database queries for complex pages. Second was that our web host at SDSC upgraded their infrastructure and hardware. Speed was improved by summer 2014. We have also begun working on improving the user experience across the main CZO web resources for finding and accessing data: 1) dataset listings at criticalzone.org, 2) data search portal at search.criticalzone.org, and 3) data visualization portal at viz.criticalzone.org.
- ○ **Expected Delivery Date:** Speed improvements were completed in September 2014.
  - ● Going forward, we will also improve some interface elements and functionality across the three data-related websites by Fall 2015.
  - ● The CZO National Office, which in mid 2014 assumed the role of maintaining the non-data portions of CriticalZone.org, will fund additional general website improvements in 2015 (i.e. connections with czen.org, site-wide search box, new home page, updated navigation).

- ● **T1c: Host CZO online content management system, and integrate it with the CZO data discovery portal**
  - ○ **Proposed**: Key aspects of hosting are maintenance and support. A smoothly running website helps the people visiting the website as well as the CZO personnel updating it. Maintenance includes software updates to ensure site security, speed, and reliability. Support includes helping CZO personnel with their questions, and fixing bugs. The CZO CMS needs to work well with the CZO data discovery [search] portal (see Task 8), with consistency across the interface.
  - ○ **Changes to Scope**: The original intention was for the CZO Website Content Management System (CMS, at http://criticalzone.org/) and the CZO Data Search Portal (http://search.criticalzone.org/) to be loosely integrated via links and by sharing similar visual elements to provide better consistency for visitors to both websites. But it became clear that the integration needed to be much tighter. We needed to directly share the CMS metadata widely in order to enhance data discovery and interoperability with other national data systems. This required connecting the CMS to a data portal via the increasingly used ISO 19115 Geographic Metadata Standard (recommended by the Federal Geographic Data Committee and adopted by NOAA, the US Geological Information Network and many other national organizations).
  - ○ **Progress to Date:**
    - ● The CMS software and its foundational technology stack (database, webserver, etc.) has been maintained to deal with any critical security issues and a number of physical and structural changes to the SDSC web hosting infrastructure over the years. In general, maintenance and support have proceeded as expected with one exception. Adapting to numerous hardware and software changes made by our website host at SDSC took substantially more effort than predicted.

- In Dec. 2013 we began the process of developing automated metadata harvesters from the CriticalZone.org website CMS database into the CZO Data Search system (http://search.criticalzone.org/) that is built upon ESRI Geoportal that allows metadata discovery and interoperability with other national data systems. We originally developed the CMS web form to follow a very enhanced version of the "Dublin Core" metadata standard, but since Dec. 2013 we have chosen to map the CMS database entries to the much "richer" ISO 19115 Geographic Metadata standard in order to fully export all of the metadata contained in Dataset pages. This approach is fundamentally so much better than what we had previously loaded into the CZO Data Search Portal. We put the effort on hold during the first half of 2014 in order to launch the four new CZOs at CriticalZone.org. We renewed our efforts in the summer of 2014 and set up basic one-way syncing from the CMS to the data search portal before the CZO All-Hands Meeting in September 2014. Additional functionality and improvements were made in the subsequent months. More minor modifications remain.
  - **Expected Delivery Date:** Major integration completed.
    - Maintenance and support were essentially "completed" by mid 2014 when responsibility for much of the basic, ongoing website maintenance shifted from the CZOData project to the CZO National Office.
    - Main functionality for the integration with the data discovery portal was completed in November 2014.
    - Minor functionality enhancements for the integration (i.e. user interface adjustments, Google Analytics tracking) is expected to be complete by Fall 2015.

Search.CriticalZone.org

T8: CZO Central metadata catalog, data discovery portal and harvester:

Proposed: The initial version of the CZO data discovery system developed as part of the CZO data management prototype will be significantly expanded to support search across different types of CZO data from all sites. We will develop software to automatically register and index the harvested Display Files of different types to the central metadata catalog, populating metadata fields with information extracted from the Display Files. This will enable browsing and querying of different types of Display Files by title, keywords, contributor, spatial location, thematic category and similar fields. In addition, we will develop three key extensions of the data discovery portal: 1) to allow download of the discovered collections of Display Files via a simple user interface; 2) to let CZO users visualize the discovered data in a viewer specific for this type of data; and 3) to implement the necessary web service interfaces for CZO Central to become a DataONE Member Node. As part of this task, we will also extend the Display File harvester capabilities so that it can initiate a data publication workflow specific to the data type of the Display File and route it to the appropriate CZOData sub-system. (SDSC)

- **T8a: Develop means of registering non-timeseries data in portal**

- ○ **Proposed**: Develop solutions for registering CZO-collected data of types other than time series with the central portal (primarily, geochemical samples and spatial data layers) (SDSC, IEDA, Q5-Q7).
- ○ **Changes to Scope**: None.
- ○ **Progress to Date:** As described above (Supplemental Task A and Task 1C), we developed Dataset pages on CriticalZone.org to serve two purposes, first to serve as a traditional way to "browse" datasets, but secondly as a means to harness the CMS to produce [ISO-19115 Geographic Metadata](#) documents that could be ingested into the CZO Data Search Portal. As a result, data of any resource types that are added by data managers to the CMS, are automatically registered in the search portal and made available for search. All conversions (between the CMS metadata model and standard ISO-19115) are handled behind the scenes. The metadata then becomes available via standard formats and protocols through the search portal.
- ○ **Expected Delivery Date:** Completed November 2014.
- ● **T8b: Develop web-services for geochemical data**
  - ○ **Proposed**: Assist the IEDA team in development of standards-based web services for geochemical data and registering them to the CZO data portal (SDSC, Q5-Q8)
  - ○ **Changes to Scope**: The development of web services for IEDA's major data systems was put on hold in 2014 while IEDA completely rebuilds its cyber-infrastructure from the ground up around ODM2.
  - ○ **Progress to Date:**
    - ● The new [CZO Data Search Portal](#) provides a number of web services (http://search.criticalzone.org/api/) to enable discovery of the entire federated CZO data catalog, including data that have been registered by CZOs at CriticalZone.org but reside in the EarthChem Library.
    - ● IEDA's [System for Earth Sample Registration (SESAR)](#) began in autumn 2014 to develop its first web services based on ODM2 for geochemical sampling features (i.e. specimens and sites). Several initial REST web services over ODM2 - for sites, variables, actions and values - have been developed (http://cybergis.sdsc.edu:7000/docs/). These services are being used (and tested) by the [CZO Data Visualization Portal](#) (Task 12).
  - ○ **Expected Delivery Date:** Will be fully completed by Summer 2015.
- ● **T8c: Enhance online CZO catalog**
  - ○ **Proposed**: Enhance the online CZO catalog application and data portal to make dataset discovery more user-friendly and intuitive. As part of this task, enable dataset download, for appropriate data types, directly from the data portal. (SDSC, Q3-Q6)
  - ○ **Changes to Scope**: None.
  - ○ **Progress to Date:** A lot of effort was put into making data discovery more user friendly and intuitive. A completely redesigned user inferface "front end" for the [CZO Data Search Portal](#) was setup at http://search.criticalzone.org to support data discovery, faceted search, data download, and provenance tracking, and works on both desktop and mobile platforms. In addition, the "backend" of the CZO Data catalog, based on the [ESRI Geoportal Server](#) platform, was also substantially updated

with a number of enhanced features, including SOLr free text search and conversion of the default metadata schema from the basic Dublin Core Metadata Standard to the much more expressive [ISO 19115 Geographic Metadata Standard](#). The information in the portal is now also accessible programmatically, via the [Catalog Services for the Web (CSW)](#) service interface.

- ○ **Expected Delivery Date:** Completed Sept 2014.

- **T8d: Develop code to register and download display files**
  - ○ **Proposed**: Develop code to automatically register the harvested display files to the online catalog, and make them downloadable via a simple user interface (SDSC, Q5-Q8).
  - ○ **Changes to Scope**: With the decision to redesign CZO Display Files around ODM2, rather than simply extending the specification to new disciplines, we created a need to develop completely new display file harvesters for CZO Display File v2, in addition to maintaining harvesters for CZO Display File v1.
  - ○ **Progress to Date:**
    - Display files (both versions 1 & 2) were exposed on the new CZO Data Search Portal in September 2014 as a separate resource type in order to discover and download them in a standard manner.
    - Harvesters and validators for CZO Display File v1 continue to be maintained.
    - Harvesters and validators for CZO Display File v2 are being actively developed in parallel with the development of the CZO Display File v2 specification (Tasks 2 & 3).
  - ○ **Expected Delivery Date:** Harvesters and validators for CZO Display File v2 will be developed by June 30, 2015 and refined throughout 2015 as the CZO Display File v2 specification gets tested and refined.

- **T8e: Enhance CZO online data registration**
  - ○ **Proposed**: Enhance the CZO online system that assists data managers in registering and managing their observation networks (password protected) (SDSC, Q1-Q6).
  - ○ **Changes to Scope**: In Phase 1 of CZOData (2009-2011), CZO datasets were manually registered into the CZO Catalog by CZO Data Managers via http://central.criticalzone.org. For Phase 2, we decided to leverage the CZO Dataset Listings at CriticalZone.org via the CMS (Supplemental Task A and Task 1c) as a means of data registration.
  - ○ **Progress to Date:** The scope of this task increased substantially. It resulted in much more comprehensive enhancements than planned, because of the move to the streamlined CMS data registration, harvesting and synchronization via the search portal. The earlier registration portal functionality was integrated with the CMS data registration front-end. The registration backend continued to support data validation against controlled vocabularies, automatic loading of validated data into an ODM1 database (for time series display files), and update of web services.
  - ○ **Expected Delivery Date:** Completed April 2014

- **T8f: Add DataOne Member Node interfaces**

- ○ **Proposed**: Add DataONE Member Node interfaces to CZO Central (SDSC, Q6-Q8).
- ○ **Changes to Scope**: DataONE is a distributed cyber-infrastructure system that enables interoperability across a network of data centers, or Member Nodes. DataONE was developed using the data cataloging approaches adopted by the biological community, including the LTER network. However, the DataONE approach is different from the approaches used by the geosciences community, including CUAHSI and IEDA. In 2011, when this proposal was written, CUAHSI and DataONE were beginning discussions on how to bridge this fundamental incompatibility. Development of a CZOData Member Node for DataONE relied completely on the success of those efforts. The CZOData project postponed this task until such an interoperability capability is developed. Second, because CZOData itself is not a data center, this project cannot directly become a DataONE Member Node. Rather, we would assist CUAHSI and IEDA in becoming Member Nodes and ensuring that CZOData is properly exchanged within the DataONE network.
- ○ **Progress to Date:**
  - ● We have been involved in conversations between DataONE and CUAHSI since the beginning of this project.
  - ● The CZO Data Search Portal was developed to use Catalog Services for the Web (CSW) service interface to enable federated data exchanges between data systems. DataOne is presently planning to develop support for the CSW service API: this support has been discussed by the DataOne team since at least 2012, and is currently showing as in progress (https://redmine.dataone.org/issues/3229). Once that is done the CZO system will be compatible with DataOne at the metadata level.
  - ● The implementation of ODM2 by CUAHSI and IEDA also moves us closer down the path toward becoming a member node. Compatibility at the data level will be accomplished as the new CZO data infrastructure based on ODM2 is rolled out. Currently, several REST service interfaces for ODM2 are already developed (to expose measurement sites, variables, actions, values) as mentioned in Task 8b. Once these services expose content of ODM2 more fully, it will be easy to wrap them as DataOne service interfaces.
- ○ **Expected Delivery Date:** End of 2015, depending on collaborative efforts between CUAHSI and DataONE.

T9: Central data repository of time series and point data

Proposed: The CZOData project will greatly expand the hydrologic time series publishing mechanism that we initially explored in the data management prototype. In particular, we will develop a versioning and archiving system for CZO-collected time series, to ensure data integrity and accessibility. (SDSC)

- ● T9a: Wrangle data from SIMS
  - ○ **Proposed**: Work with data managers to publish additional time series data from CZO sites. Ensure that the configured display files are successfully harvested into the central system, and troubleshoot harvesting problems when required. (SDSC, Q1-Q8)
  - ○ **Changes to Scope**: None

- ○ **Progress to Date:**
  - ■ From mid-2011 to mid-2013, after phase 1 funding ended, the CZOData team worked closely with CZO data managers to publish data series as CZO Display Files v1, and helped fix issues with files. These files were parsed, validated and ingested into six different HydroServer database systems that SDSC maintained on behalf of each CZO to enable full functionality of the CUAHSI HIS system for CZO time series data. This was an intensive effort as the six CZOs at that time prepared for renewal proposals in Feb. 2013.
  - ■ From Jan. to Dec. 2013, the CZOData team worked intensively with CZO data managers to develop and populate the new Dataset pages at CriticalZone.org, which could accept data files of any type. This effort was then put on hold for 7 months while CriticalZone.org was being rebuilt to accommodate four new CZOs.
  - ■ From July 2014 to the present, the CZOData team has trained and retrained new and old CZOs to load Datasets onto CriticalZone.org.
  - ■ In early 2014, SDSC set up four new HydroServer database systems, one for each of the new CZOs, to network time series data into the CUAHSI HIS network.
  - ○ **Expected Delivery Date:** A continuous effort.
- ● T9b: Develop archiving and versioning system.
  - ○ **Proposed**: Develop an archiving and versioning system at the central CZO repository, which ensures that latest time series versions are available via web service calls while previous versions of the time series harvested by CZO sites are still managed in the central database. (SDSC, Q2-Q4)
  - ○ **Changes to Scope**: None.
  - ○ **Progress to Date:** All harvested display files are archived and regularly backed-up at SDSC. The latest version of the data (time series) is available via service calls.
  - ○ **Expected Delivery Date:** Completed.
- ● T9c: Develop hardware/software for central data repository
  - ○ **Proposed**: Develop reliable hardware and software environment for the central data repository (redundant mirrored databases and web servers). (SDSC, Q3-Q5)
  - ○ **Changes to Scope**: None.
  - ○ **Progress to Date:** All HydroServer instances for each CZO are mirrored on two servers, and their availability is continuously monitored using the r-u-on infrastructure. The search portal is also continuously monitored. Messages about system outages are sent to the maintenance group at SDSC.
  - ○ **Expected Delivery Date:** Completed April 2014.
- ● T9d: Ensure data availability via UI
  - ○ **Proposed**: Ensure that the harvested data become available via standards-based service interfaces, and update the interfaces when new standards are adopted (in particular, WaterML2 via SOS). (SDSC, Q5-Q7)
  - ○ **Changes to Scope:** None.
  - ○ **Progress to Date:** Data and metadata are available in standard formats and via standard protocols, including OGC CSW, OAI-MPH, OGC WMS and WFS where

applicable. CZO time series data are available in WaterML 2 via REST service interfaces. They are not yet available via SOS because the Hydrology SOS profile was approved only recently as OGC Best Practice (in late October 2014). Comprehensive WaterML2/SOS update will be coordinated with CUAHSI Water Data Center. The SDSC group participated and co-organized development of these hydrologic standards at the OGC (Zaslavsky is a co-chair of the OGC/WMO Hydrology Domain Working Group; Valentine is a co-chair of the WaterML 2 Standard Working Group).

- ○ **Expected Delivery Date:** Primarily completed in Dec. 2014. The SOS2 Hydrology profile will be completed in fall 2015.

Viz.CriticalZone.org

T12: Visualization Tools:

Proposed: The CZO Visualization System (CZO-VS) will be developed by the APL team at the University of Washington as an adaptation and enhancement to NVS. The CZO-VS will focus on visualization and analysis of time series data generated by CZO sites and distributed via the SDSC-hosted system using CUAHSI HIS-based web services and OGC standard web services. The CZO-VS will present a U.S. national view showing the six CZO sites. Upon site selection or map zooming, all monitoring locations ("assets") from the selected CZO site will become visible and accessible for browsing and filtering. NVS will be enhanced to present an expanded set of asset metadata based on a core sub-set of standardized CZO metadata and controlled vocabularies. Likewise, current NVS asset filtering and selection will be enhanced to reflect a core sub-set of the standardized CZO shared vocabularies. The near-real-time focus of NVS will be expanded from the current 30 days to 60 days. More importantly, a new, generalized, user-friendly temporal search and display capability will be developed to allow more unfettered access to CZO data. In addition to CZO data, the CZO-VS will present other assets relevant to CZO sites if they are available from the CUAHSI HIS Central catalog, including those from USGS, EPA and other national providers; however, the data search, exploration and visualization capabilities for these assets will be limited compared to CZO assets, as they are not part of the CZO integration efforts. Core GIS layers relevant to CZO sites will also be presented, enabled via data coordination with the CZO Central metadata catalog and that build on site spatial data sets. (APL)

- T12a: Enhance Vizer time-handling components
  - ○ **Proposed**: Enhance NVS framework with generic time-handling components. (APL, Q1-Q4)
  - ○ **Changes to Scope**: None.
  - ○ **Progress to Date:** The NVS framework (generalized and renamed to the "Vizer" framework in 2014) has been expanded to support multiple components of time handling functionality:
    - ■ The "near-real-time" time window for time series visualization plots was expanded from 30 days to 60 days (late 2013). This is currently visible in the Christina River Basin CZO App (http://viz.criticalzone.org/Christina).
    - ■ A rich timeline interface was first added in June 2013, and subsequently enhanced throughout 2014. It supports interactive time selection; ability to follow along with

real time; time stepping and animation controls; timeline auto displays and hides, and auto scaling; and support for long-range time series and climatologies. This tool currently has some functional limitations and is limited to time-dependent map overlays (e.g., time-dependent maps of climate variables derived from models or satellites), which are not part of the target deliverables for the CZO Data Visualization Portal.

- ○ **Expected Delivery Date:** Time-handling capabilities will be enhanced and hooked up to all CZO hydrological time series (derived from the SDSC-hosted, CZO Central CUAHSI HIS WaterOneFlow 1.1 web services, ultimately from CZO Display Files created by each CZO) by extending existing Vizer timeline interface and data harvesting capabilities, providing plots for the entire length of a site's time series. This feature will be available at the CZO Data Visualization Portal as an initial limited release by Mar 31, 2015, and completed by Apr 30, 2015.

- ● T12b: Create visualizations of current CZO data
  - ○ **Proposed**: Access and visualization of time series data focused on each CZO site, including: initial national-map view of the 6 CZO sites; CZO visual branding; and incorporation of CZO metadata vocabularies in data searching and dynamic filtering; incorporation of core GIS layers. (APL, Q1-Q6)

  - ○ **Changes to Scope**: The 4 new CZO sites (Calhoun, Eel River, IML and Reynolds) needed to be accommodated in the Visualization tools, data harvesting, and coordination with individual CZO data management teams.

  - ○ **Progress to Date:** The initial version of the CZO Visualization System, now labeled "CZO Data Visualization Portal", was released early Summer 2014 under http://viz.criticalzone.org. This system leveraged developments with the underlying "Vizer" framework (previously described as the NVS Framework) over the last 18 months. The Visualization Portal features a central page with visual access points for each of the ten CZO's. Each CZO has its own "App" focusing on its region and data. The portal has been customized to the general CZO web visual branding, and individual Apps include customizations provided by individual CZO's. GIS layers representing CZO boundaries have been added to each CZO App (Jan 2015), based on files provided by individual CZO Data Managers. Sites from all CZO hydrological time series currently available via the SDSC-hosted, CZO Central CUAHSI HIS WaterOneFlow 1.1 web services have been loaded into each CZO App (Jan 2015) with an initial version of "variables" measured at each site (Mar 12, 2015); listing of variables is visible for browsing, but time series visualizations are not accessible yet (see Task 12a). Progress has been made in assessing and processing the CZO metadata vocabularies (Nov 2014 - Jan 2015).

  - ○ **Expected Delivery Date:**
    - ■ General refinements to user interface: Mar 31 and Apr 15, 2015.
    - ■ Incorporate additional GIS layers delineating CZO sub-regions provided by CZO Data Managers, as well as sub-region bookmarks for map navigation: Mar 23, 2015.
    - ■ The handling of CZO metadata vocabularies is being finalized. Final version of "variables" measured accessible on time series sites pop-ups will be deployed by

Mar 31, 2015. Filtering capabilities for "variables" measured will be initially released by Mar 31, 2015, and refined by Apr 30, 2015.

- ■ Initial time series visualizations for CZO sites: Mar 31, 2015. See Task 12a for more information.

- ● T12c: Optimize data access via Vizer

  - ○ **Proposed**: Work with SDSC to develop and deploy data-access optimization schemes, and to enable incorporation of select NVS components into the CZO data portal. (APL, SDSC, Q2-Q7)

  - ○ **Changes to Scope**: None

  - ○ **Progress to Date:** Data-access optimization schemes for CZO time series are being investigated. Targeted linkages from the CZO Data Search Portal to the CZO Data Visualization Portal are also being assessed.

  - ○ **Expected Delivery Date:** Apr 30, 2015

- ● T12d: Implement cross-CZO capabilities

  - ○ **Proposed**: Implement search, access and visualization capabilities that integrate across CZO sites; enhance data selection and download tools. (APL, Q5-Q8)

  - ○ **Changes to Scope**: None.

  - ○ **Progress to Date:** The development of the CZO Data Visualization Portal has already entailed ingesting CZO site time series from the SDSC-hosted, CZO Central CUAHSI HIS WaterOneFlow 1.1 web services. While the current presentation is based on CZO-specific Apps, a next step will involve the creation of a new app that incorporates all such CZO site time series.

  - ○ **Expected Delivery Date:**

    - ■ Implement search, access and visualization capabilities that integrate across CZO sites via a new, cross-CZO vizer App: Apr 30, 2015.

    - ■ Enhance data selection and download tools: Other priority developments and CZO project decisions have resulted in delays with these capabilities. They will be available by summer 2015.

- ● T12e: Add UI/Visualizations based on requests from SIMS

  - ○ **Proposed**: Assess and design additional user-interface and visualization features based on CZO user priorities; includes assessment of limited search and visualization of geochemical sample sites and time series data from non-CZO data streams available through CUAHSI HIS. (APL, Q4-Q8)

  - ○ **Changes to Scope**: None

  - ○ **Progress to Date:**

    - ■ The base map was changed to accommodate the need to be able to zoom in to fine spatial scales (Dec 2014).

    - ■ SESAR/IGSN sample sites can now be ingested and shown on the CZO Data Visualization Portal. Current presentation is limited by the availability of CZO samples registered with SESAR (see Task 8b). IGSN sample sites for currently SESAR-registered CZO's are accessible for the Susquehanna Shale Hills CZO

App (http://viz.criticalzone.org/ShaleHills) (Jan 2015) and the Luquillo CZO App (http://viz.criticalzone.org/Luquillo) (Mar 12, 2015).

- ■ Near-real-time time series data and visualizations are available for some non-CZO data streams accessible through CUAHSI HIS, currently only for the Christina River Basin CZO (http://viz.criticalzone.org/Christina) (Sep 2014).
- ○ **Expected Delivery Date:**
  - ■ SESAR/IGSN sample sites will be expanded to Apps from all CZO's that have registered samples with SESAR (which currently is limited to Susquehanna Shale Hills and Luquillo): Updates will occur on Mar 31 and Apr 30, 2015, if available.
  - ■ A site clustering scheme is also being developed to improve map-based search and access for closely spaced sites: Apr 15, 2015.
  - ■ Non-CZO time series data streams from select national monitoring agencies (particularly USGS) available through CUAHSI HIS will be made available to all CZO Apps: Apr 30, 2015.

CZchemDB Online

T4: Develop a web-based user interface to CZchemDB:

Proposed: Currently, CZchemDB is not web-enabled and can only be accessed locally at PSU. We will move CZchemDB into a more robust relational database management system and will develop an interactive, web-based interface to CZchemDB, similar to the interfaces that we developed for other IEDA databases that allow users to access the databases online and retrieve data of their choice. The interface will provide users with tools to search for geochemical data using a variety of criteria that will be defined in consultation with CZO investigators and based on the metadata that are stored in CZchemDB. Users will be able to create customized subsets of the data, integrated into a single data table that they can view online or download for further analysis, and access the complete metadata about samples and analytical procedure. (IEDA)

Changes to Scope: The initial evaluation of the CZchemDB prototype, implemented by the Penn State team in MS Access, concluded that it would need to be significantly modified to enable an enterprise server implementation that would be interoperable with the CyberInfrastructure at IEDA. Furthermore IEDA's cyberinfrastructure also began a major foundational transition starting in late 2013; IEDA had decided to adopt ODM2 as its underlying information model to enable a number of new capabilities. Therefore, to not leave CZChemDB "in the dust" as an unmaintainable web-application that would be disconnected from the new EarthChem system, our team decided to fully implement CZchemDB as a profile of ODM2 within IEDA's new ODM2-based EarthChem2 cyber-infrastructure, which will be rolled out in pieces over the course of 2015.

- ● T4a: Transfer CZChemDB to new RDBMS
  - ○ **Proposed**: Transfer CZchemDB to new relational database management system (RDBMS) (IEDA, Q1-Q2).
  - ○ **Changes to Scope**: None.

- ○ **Progress to Date:** The CZchemDB database has been finalized by the Penn State team that developed it and sent to the IEDA (Columbia) team. The IEDA team translated it from the prototype desktop implementation in Microsoft Access to an enterprise server implementation on a Linux/Apache web server and using a PostgreSQL Relational Database Management System.
  - ○ **Expected Delivery Date:** Completed in late 2013.
- ● T4b: Develop specifications for CZchemDB User Interface
  - ○ **Proposed**: Develop specifications for CZchemDB User Interface (IEDA, Q3-Q4).
  - ○ **Changes to Scope**: None.
  - ○ **Progress to Date:** Although CZChemDB has been "mapped" to ODM2 as the backend database system, we intend that the front-end data loaders and other functions have the feel of the original CZchemDB. The user interface also includes search, access and visualization tools. IEDA has been actively partnering with Exaptive (https://exaptive.com/), a Data Driven Web Application development firm in Boston, to develop IEDA's new EarthChem User Interface for search and visualization. As such, IEDA has been actively developing EarthChem2 UI specifications, and the CZchemDB web application UI will directly adopt those design elements.
  - ○ **Expected Delivery Date:** Initial specifications are complete.
- ● T4c: Develop web application for CZChemDB UI
  - ○ **Proposed**: Develop web application for CZchemDB UI.
  - ○ (IEDA, Q5-Q8)
  - ○ **Changes to Scope**: None.
  - ○ **Progress to Date:** Demos of the new EarthChem UI (developed by Exaptive) are now available for private testing.
  - ○ **Expected Delivery Date:** The public web application will be available by the end of 2015, depending on the development of IEDA's new ODM2-based cyber-infrastructure.

T11: Integrate CZchemDB data into the EarthChem system:

Proposed: We will work with the Penn State group to develop scripts and procedures to convert the CZchemDB data into XML documents, so that the rock plus regolith chemistry data can be seamlessly incorporated into the EarthChem system. We will develop web services that will harvest the XML encoded data from CZchemDB for inclusion in the EarthChem Portal database so that CZ geochemical data can be searched seamlessly together with other EarthChem partner databases. (IEDA)

- ● T11a: Convert CZChemDB into XML
  - ○ **Proposed**: Develop scripts and procedures to convert the CZchemDB data into XML documents. (IEDA, Q1-Q3)
  - ○ **Changes to Scope**: None.
  - ○ **Progress to Date:** All data existing within the MS Access version of CZchemDB have been ingested into a PostgreSQL clone of the original CZchemDB and from there mapped and transferred into an ODM2 version of CZchemDB.

- ○ **Expected Delivery Date:** Completed spring-2014.
- ● T11b: Harvest CZOChemDB XML into EarthChem
  - ○ **Proposed**: Develop web services that will harvest the XML encoded data from CZchemDB for inclusion in the EarthChem Portal database. (IEDA, Q4-Q6)
  - ○ **Changes to Scope**: As described above, because EarthChem is being developed around ODM2, CZchemDB also needed to be developed around ODM2.
  - ○ **Progress to Date:** IEDA's new ODM2-based EarthChem2 cyber-infrastructure is still under development. Because CZchemDB has already been converted to ODM2 (Task 11a), when the new ODM2-based EarthChem2 is ready, CZchemDB will be ready to ingest into the new EarthChem.
  - ○ **Expected Delivery Date:** By the end of 2015.

Controlled Vocabularies

T5: Shared vocabulary system:

Proposed: The prototype Shared Vocabulary system will be upgraded to support mapping of variable names used by CZOs to terms in a common variable ontology to enable cross-CZO search by variables. Translation between domain terminologies to establish a CZO-wide shared vocabulary system will also be explored. We will develop vocabulary systems that will serve domain vocabularies for point time series data and for geochemical data in a standard format (e.g., Simple Knowledge Organization System, or SKOS), which would enable standards-based mapping between terms from different vocabularies, and their synthesis in a comprehensive, dynamically-updated CZO vocabulary management system. We anticipate that this system will allow us to manage controlled vocabularies and higher-level structures (polytaxonomys, thesauri and ontologies) and interface with vocabularies used by other environmental communities, such as the TemaTres system used by LTER. (UCHIC, SDSC)

Changes to Scope: Base on ODM2 CVs.

- ● T5a: Extend the existing vocabulary system
  - ○ **Proposed**: Extend the existing vocabulary system to variables and other terms used by CZOs. (UCHIC, Q1-Q8)
  - ○ **Changes to Scope**: See discussion in progress below.
  - ○ **Progress to Date:**
    - ■ After discussion with the CZOData Information Managers Committee, we decided in May 2014 to decommission the prototype CZO Shared Vocabulary system at sv.criticalzone.org. This decision was based on several factors, including: 1) we worked with the CUAHSI Water Data Center to move the existing CUAHSI HIS CV system from USU to the CUAHSI Water Data Center in efforts to make it more sustainable; 2) the functionality at sv.criticalzone.org was based on and nearly identical to the original CUAHSI HIS system - we wanted to avoid duplication and confusion; and 3) the CUAHSI system had many CV terms that had been contributed by CZO/hydrology community members that the CZO system did not have for describing hydrologic time series.

- In 2014, work was completed to compare the CV terms at sv.criticalzone.org to those available via the CUAHSI HIS system. Any unique terms from the CZO system were moved into the CUAHSI HIS system. CZO Data managers were then encouraged to use the CUAHSI HIS CV system for choosing metadata terms to describe their hydrologic time series data.

- In 2014 it was also recognized by the CZOData team that the existing CUAHSI HIS CV system would be adequate for describing hydrologic time series data, but that it did not contain additional vocabularies needed to describe datasets derived from specimens (e.g., geochemical data, etc.). Given that the CZOData systems surrounding data derived from specimens are largely adopting the ODM2 information model, in late 2014/early 2015 the CZOData team began work on an additional system for managing ODM2 controlled vocabularies. This system is under development now (http://vocabulary.odm2.org/), with expected completion in March 2015.

- The following tasks have been completed with respect to the new ODM2 CV system: 1) design and implementation of the underlying database is complete; 2) the ODM2 CV terms for finished CVs developed by the ODM2 team have been migrated to the new database; 3) the structure of the ODM2 CVs has been mapped to terms in the SKOS namespace for RDF/XML encoding of the vocabularies; 4) a REST web service has been implemented for publishing the CV terms for programmatic access in SKOS format; 5) the main user forms for a web user interface have been created for displaying the ODM2 CV terms and enabling users to submit new terms and modifications to existing terms. Remaining work includes: 1) modifying the REST web service to serialize responses in SKOS RDF/XML format; 2) connecting the main user forms to the populated database of CV terms; and 3) finalizing the moderation system for terms submitted by community members.

- **Expected Delivery Date:**
  - CUAHSI HIS CV system for time series metadata - available now.
  - Initial release of ODM2 CV System with CV listings - March 31 2015.
  - Updated ODM2 CV system with system moderation and live update capabilities for SIMS to edit/add terms - May 2015
  - Maintenance of the CV system through the end of the project.

- T5b: Support mapping of variables to common ontology
  - **Proposed**: Support mapping of variable names used by CZOs to terms in a common parameter ontology, to enable cross-CZO search by parameters. Ensure that harvested data conform with shared vocabularies as managed by the USU team. (SDSC, Q1-Q8).

  - **Changes to Scope**: None

  - **Progress to Date:** The common set of vocabularies used for parameter names has been maintained as sv.criticalzone.org. As described above the system was moved to the CUAHSI Water Data Center (WDC). On submission of hydrologic display files, parameter names are now being verified against vocabulary web services from CUAHSI WDC, and feedback is provided to data managers if parameter names do not

match. This part has been operational and maintained as part of time series display file ingestion and registration process. Once the ODM2 CV system is operational, validation will be extended to parameter names maintained in that system.

- ○ **Expected Delivery Date:** Parameter ontology completed for CUAHSI HIS based on ODM1.1. ODM2 vocabulary ontologies will be completed by late-2015.
- T5c: Setup SKOS for controlled vocabularies
  - ○ **Proposed**: Setup a simple knowledge organization system (SKOS) for CZO controlled vocabularies. (SDSC, Q4-Q8)
  - ○ **Changes to Scope**: None.
  - ○ **Progress to Date:**
    - The new ODM2 controlled vocabulary system described above with initial version expected to be completed in March 2015 will expose all ODM2 controlled vocabularies and individual terms using SKOS. This will make it easy for programmers and other systems to access the vocabularies and will help with interoperability. We have completed mapping of the ODM2 CV concepts and the ODM1/CUAHSI HIS concepts and their attributes to SKOS (required for encoding in SKOS). We have also developed a REST web service interface that will make it easy to retrieve XML/RDF SKOS representations of both controlled vocabulary lists and individual terms from each vocabulary. We plan to use this within the CZOData data entry templates to provide data managers with access to the CV terms directly within the templates. Once complete for the ODM2 vocabularies, we can work with the CUAHSI Water Data Center personnel to expose the ODM1/CUAHSI HIS terms for hydrologic time series using the same system.
    - There is also ongoing work within IEDA to expose all of the IEDA vocabularies as SKOS.
    - At the same time several components of the CZO controlled vocabularies have been organized as SKOS, including the CUAHSI parameter vocabulary published by our partners at http://edscvs.ccny.cuny.edu/. ODM 1.0 controlled vocabularies are also exposed through the same server. Several additional efforts to setup SKOS services within this project, have been undertaken, via publishing ODM2 vocabularies in MMI, and registering these vocabularies in the experimental SciCrunch API at SDSC. The latter allows simple web service calls to annotate texts with terms found in the vocabularies, providing term references, which can be SKOS-formatted.
  - ○ **Expected Delivery Date:**
    - Initial version of ODM2 Controlled Vocabularies published as SKOS (http://vocabulary.odm2.org/) - March 31, 2015.
    - ODM1/CUAHSI HIS Controlled Vocabularies published as SKOS - dependent on collaboration with Water Data Center personnel.
    - IEDA Controlled Vocabularies published as SKOS - Summer 2015.

Other related tasks:

- T11c: Update EarthChem Vocabularies

- ○ **Proposed**: Update EarthChem vocabularies to include terminology used by CZchemDB, and integrate with CZO shared vocabulary system. (IEDA, Q5-Q8)
- ○ **Changes to Scope**:
- ○ **Progress to Date:** To the extent possible, the EarthChem vocabularies (and others used by the IEDA systems) have been harmonized with the ODM2 controlled vocabularies. Our team has examined the terms and created crosswalks where appropriate. Many of the IEDA terms have been included in the ODM2 controlled vocabularies, and, when encoded in SKOS, the ODM2 CVs will provide links to equivalent terms in the IEDA vocabularies. However, our plan moving forward is to validate submitted datasets from the CZOs against multiple vocabulary systems and not limit CZO data managers to a single system. The idea is that metadata terms used in describing CZO datasets should come from an established community vocabulary, but that CZO data managers will need some choice and flexibility about which one to use (e.g., the CUAHSI HIS vocabulary for hydrologic time series and the ODM2 and/or CZChemDB vocabularies for datasets derived from specimens).
- ○ **Expected Delivery Date:**
  - ■ Overlapping IEDA Vocabularies integrated with ODM2 vocabularies - March 31 2015.

CZODisplayFile2 & Publication Tools

T2: Consistent Metadata:

Proposed: We will expand the current metadata standards developed for hydrologic time series as part of our prototype development to support additional data types created by CZOs. This will involve continuous collaboration with SIMs and research scientists from each CZO site through the IMC to provide sufficient metadata descriptions for these data to be interpreted and useable for cross-CZO comparison and modeling. We will extend the Display File format design and metadata to present geochemical sample information at the individual CZO web sites in a human- and computer-readable form. For the geochemical extensions to the Display File format, we will use the metadata schema that has been in use for several geochemical databases (PetDB, SedDB, GEOROC, NAVDAT, EarthChem) and includes documentation of samples (geospatial information, description and classification, collection method, etc.), sample preparation, analyzed material, data quality (reference material measurements, blank values, analytical precision), analytical technique, and data reduction procedures.

- ● T2a: Extend CZO display file format to geochemical data
  - ○ **Proposed**: (IEDA, UCHIC, SDSC; Q2-Q4)
  - ○ **Changes to Scope**: None.
  - ○ **Progress to Date:** The CZOData team had a 2.5-day meeting in February 2015 at IEDA to the CZO Display File v2 format specification to support geochemical data. At this meeting we also revisited the existing display file format for hydrologic time series and addressed several issues that have been brought to our attention by CZO data managers. We invited several participants from the CZO information managers committee who self identified themselves to help work on the display file specifications

to attend this meeting. The result of this meeting has been: 1) updated display file format for hydrologic time series; 2) display file format for geochemical data; 3) scoping and plans for software tools surrounding the display file formats.

- ○ **Expected Delivery Date:**
  - ■ Major decisions completed and documented publically at https://github.com/CZOData/CZODisplayFile2, along with examples.
  - ■ The specification document is being drafted and reviewed in a private Google Document and will be released by the end of April 2015. We will incorporate feedback from CZO Data Managers as they use the CZODisplayFile2 specification throughout 2015.

- **T2b: Extend CZO display file format to spatial data**
  - ○ **Proposed**: For the geochemical extensions to the Display File format, we will use the metadata schema that has been in use for several geochemical databases (PetDB, SedDB, GEOROC, NAVDAT, EarthChem) and includes documentation of samples (geospatial information, description and classification, collection method, etc.), sample preparation, analyzed material, data quality (reference material measurements, blank values, analytical precision), analytical technique, and data reduction procedures. (CU-Boulder, SDSC, UCHIC; Q3-Q5)

  - ○ **Changes to Scope**: Since 2011, the architecture of the CZOData system developed to take advantage of the CZOData content management system (CMS). CMS is the primary interface for data managers to interact with the system, while the CZOData Central has been configured for background communication and synchronization with the CMS. Therefore, spatial data is better handled via the CMS in their native formats (i.e. shapfiles, etc.) rather than via CZO Display Files.

  - ○ **Progress to Date:** The changes to scope have resulted in a better mechanism for registering spatial data. Advantages include:
    - ■ All data managers communicate with the system via a centrally managed CMS.
    - ■ No need to have a separate step of preparing resource-specific display files.
    - ■ Better reflects the current organization of CZOData: autonomous generation of display files for spatial data could have been justified for a smaller number of CZOs with relatively autonomous data systems; once the number of CZOs increased, and increased their reliance on a centralized metadata management, it became clear that a different approach was needed
    - ■ Robustness of the entire system increased as data managers communicated with the system via a set of metadata forms; then the metadata descriptions are retrieved, converted to the standard ISO 19115/19139 format, and automatically synchronized to the CZOData search engine (geoportal/SOLR). After that, the registered spatial data are retrieved and placed on a spatial data server at SDSC, so that it becomes available via standard OGC service interfaces (currently, via WMS and WFS).
    - ■ Currently, separately registered spatial data (shapefiles, GML files, geotiff images, etc.) are registered in search.criticalzone.org.

  - ○ **Expected Delivery Date:** Completed.

T3: Data and metadata publication tools and templates:

Proposed: A major need is software tools and templates to help with moving data from notebooks and data loggers to their web sites. We will develop software that includes a suite of functions for metadata-based analysis, quality control, transformation, and management of CZ data sets. For example, we propose to adapt the GCE LTER Data Toolbox that can directly parse data from data sources such as Campbell Scientific table and array-based data loggers, which are commonly used at CZO sites to collect sensor data. Benefits for site data managers include application of QA/QC rules, flag generation, parameter calculation, reformatting, visualization to allow for additional QA/QC, and ASCII output for direct upload to site web pages. Similarly, we will use templates developed by LDEO for several different types of geochemical analyses as detailed above to develop a suite of new, CZO specific templates for geochemical data. (UCHIC, SDSC, IEDA)

- T3a: Develop display files tools for time-series data
    - **Proposed**: Develop metadata forms and tools to support automatic generation of display files from local CZO data management systems for time series. (UCHIC, SDSC, Q2-Q4)
    - **Changes to Scope**: As part of the CZO Display File v2 format effort, we have begun to develop a number of MS Excel data loading templates for each major data type profile. These templates will let users fill out simple forms that have drop-down lists for controlled vocabularies and will automatically generate a CZO Display File v2 in the final sheet.
    - **Progress to Date:** Prototype data loading templates are available at https://github.com/CZOData/CZODisplayFile2.
    - **Expected Delivery Date:**
        - Final CZO Display File v2 data loading templates for hydrological times series will be completed by the end of March 2015.
- T3b: Develop display file tools for geochemical data
    - **Proposed**: Develop metadata forms and tools to support automatic generation of display files from local CZO data management systems for geochemical data submission. (IEDA, UCHIC, Q3-Q5)
    - **Changes to Scope**: See response to Task 3a, above.
    - **Progress to Date:** See response to Task 3a, above.
    - **Expected Delivery Date:** Final CZO Display File v2 data loading templates for geochemistry and soils (CZchemDB) will be completed by late spring 2015.
- T3c: Develop display file tools for spatial data
    - **Proposed**: Develop metadata forms and tools to support automatic generation of display files from local CZO data management systems for spatial data submission. (SDSC; Q4-Q6)
    - **Changes to Scope**: Implementing a display file exclusively for spatial data has been dropped. See Task 2b, above.

- ○ **Progress to Date:** See the description of the task T2b above: instead of generation of display files from local data system (and a likely burden on data managers to maintain the display file generation code), the system was configured to take advantage of the CZOData CMS, where the same metadata forms are filled out centrally. This streamlined system operation. The additional work that we had to do involved mapping and automatic translation of metadata fields into standard ISO XML documents, for automatic ingestion into the central metadata catalog. As a result, the same functionality is accomplished with a more robust solution
  - ○ **Expected Delivery Date:** N/A

IGSNs for CZOs

T6: Unique sample identifier for CZO data:

Proposed: Tracking individual samples and daughter samples and analyses has been identified by the CZO user community as an important component to the success of integration and synthesis of CZO data products. In order to address the specific requirements for CZO samples, we will build the CZO Registration Agent in the IGSN system to provide registration and metadata management services for CZO samples as requested and defined by the CZO community. We will work with the CZO community to review existing metadata profiles for various object types that can be registered with IGSNs and add new ones if necessary that implement shared vocabularies as needed. Use of the IGSN will be voluntary except as mandated by sub-disciplines. (IEDA)

- T6a: Develop specification for CZO object types and metadata profiles
  - ○ **Proposed**: Develop specification for CZO object types and metadata profiles, (IEDA, Q1-Q2).
  - ○ **Changes to Scope**: none
  - ○ **Progress to Date:** Through a series of discussions in workshops, webinars, and email groups over the past year, we collected requirements for CZO sample types from data managers and researchers. We have compiled the desired metadata fields that were not yet in SESAR, the most requested being UTM support for geospatial coordinates, new materials such as "ice" and "synthetic", and ability to record more comments about sample preparation. We have developed the work plan to incorporate these new requirements and will investigate the pros and cons of creating new CZO object types versus using existing object types for seamless integration into the existing system.
  - ○ **Expected Delivery Date:** The CZO-specific terms for SESAR have been compiled into the ODM2 vocabulary lists (Task 5a) and will be released via the ODM2 Vocabulary SKOS server (Task 5c) by March 31, 2015. The ODM2 SKOS system will allow for the CZO community to suggest new terms, edit old terms and otherwise manage the term lists to the end of the project (March 31, 2016).
- T6b: Build the CZO IGSN Registration Agent
  - ○ **Proposed**: Build the CZO IGSN Registration Agent. (IEDA, Q3-Q6)
  - ○ **Changes to Scope**: After discussions emphasizing the importance of integrating the CZO samples with other global samples, we decided to use the existing SESAR

back-end, while implementing the features asked for by the CZO community. Steps will be taken so that all CZO samples may be searched and accessed separately if needed.

- ○ **Progress to Date:** We are in the process of implementing the requests of the CZO community into the SESAR Registration Agent. Several CZOs already use the SESAR Registration Agent for their samples.
- ○ **Expected Delivery Date:** Summer 2015.

LiDAR and Geospatial Data

T10: Publication and sharing of spatial data and LiDAR:

Proposed: Besides geochemical and hydrologic data, this phase of the CZOData project will provide integrated access to CZO spatial data and LiDAR products. As part of this project, we will register CZO-related LiDAR data through the CZO Central data portal, and make them easy to discover and acquire along with other types of data registered at the portal. This will be accomplished by interfacing the CZO portal catalog with the registry of LiDAR datasets maintained by OpenTopography, and by incorporating elements of the OpenTopography portal (visualization, in particular) in the CZO portal. (SDSC)

- T10a: Work with OpenTopography to process and archive LiDAR
    - ○ **Proposed**: Continue working with the OpenTopography project to utilize the LiDAR project resources for archiving and processing of CZO LiDAR data. (SDSC, Q5-Q8)
    - ○ **Changes to Scope**:CZO PIs signed a Memorandum of Understanding (MOU) with Open Topography in Dec. 2011 and the CRB-CZO has already posted all their lidar data (see http://www.opentopography.org/index.php/news/detail/ncalm_and_critical_zone_observatories_data_available/).
    - ○ **Progress to Date:** We worked with the OpenTopo team to ensure that CZO Lidar data are processed and made available via standards-based services.
    - ○ **Expected Delivery Date:** Completed in 2012.
- T10b: Register LiDAR in CZO Data Portal
    - ○ **Proposed**: Register available CZO LiDAR data managed by OpenTopography, via the CZO data portal. (SDSC, Q5-Q8)
    - ○ **Changes to Scope**: None.
    - ○ **Progress to Date:** We worked with the OpenTopo team to ensure that the LiDAR catalog is compatible with the catalog behind the CZO Data Search Portal, and thus were able to federate CZOData search with LiDAR OpenTopograhy search. Both catalogs support OGC CSW interface. All LiDAR data available via OpenTopography, are now also available through the CZO Data Search Portal.
    - ○ **Expected Delivery Date:** Completed in summer of 2014.
- T10c: Manage spatial data in CZO Central repository
    - ○ **Proposed**: Replicate and manage CZO spatial data at the CZO central data repository at SDSC. (SDSC, Q3-Q8)

- **Changes to Scope**: None.
- **Progress to Date:** Spatial data contributed by CZO data managers through the CMS, have been retrieved at SDSC, registered to the [CZO Data Search Portal](#), and re-published as standards-compliant web services, following OGC specifications. The new service data endpoints are also registered at the portal.
- **Expected Delivery Date:** Functionality was completed in Autumn 2014.