Response to Reviews of

NSF Proposal 1153164: Integrated Data Management System for Critical Zone Observatories

January 13, 2012

In this response, we first articulate the overall goal and structure of our proposal to develop the CZO Integrated Data Management System (CZOData), followed by an enhanced description of how each of the proposed components will be integrated and of our project management approach to effectively meet our goals. We then provide a more detailed list and timeline of major tasks and sub-tasks, indicating project collaborator responsibilities for each and including a new Task 0 to explicitly outline and highlight the priority of community involvement efforts described in proposal section 4. We describe our vision of future oversight, maintenance and enhancement of Integrated Data Management for CZOs. We conclude with responses to specific reviewer comments.

CZOData Overview and Integration

The goal of the project is to create a comprehensive integrated data management system for CZOs that enables publication and sharing of CZO-collected data of different types. In particular, it focuses on integrating information from hydrologic observations, with geochemical and other data types that require extensive sample and method metadata, and with various forms of geospatial data. The project also focuses on providing a consistent data publication, cataloguing, discovery and access infrastructure. Specific tasks of the project are designed to develop components of the integrated CZOData system, leveraging CI efforts in related earth science fields (including CUAHSI HIS, IEDA, EarthChem, DataONE, OpenTopography, LTER, NEON). The focus of this effort is enabling CZO science in a way that: (a) supports new cross-CZO research tasks that were not possible before, and (b) eases the data management burden on CZO researchers and data managers.

The tasks derive from the proposed architecture of CZOData as presented in the proposal, and provide for:

a. extensive and iterative interaction and feedback from the community of CZO PIs, scientists and data managers (Section 4, Task 0)
b. uniform web portal appearance for the CZO sites and the national CZO program (Task 1);
c. development of a consistent metadata strategy for CZO data (Tasks 0 and 2), supported by a respective collection of data submission forms and tools (Task 3);
d. enhancing publication and data discovery workflows for geochemical (Tasks 4 and 6), hydrologic (Task 9) and spatial (Task 10) information;
e. creating a uniform data discovery portal (Task 8);
f. ensuring that the data descriptions follow consistent semantics (Task 5);
g. integrating with the EarthChem system (Task 11); and
h. developing a consistent online data visualization interface for CZO time series data (Task 12).

Taken together, these tasks will lead to a comprehensive end-to-end system that will be used by CZO researchers and data managers to reliably publish, discover, access and integrate CZO observations regardless of sites' physical locations or existing differences in metadata and access protocols. One specific focus of the proposed project is working with CZO site information managers and researchers to understand their data management challenges and provide infrastructure solutions that would be

accepted by the CZO research community (Task 7). While additional details of the architectural design and rationale are contained in the proposal , here we want to emphasize that the proposed project work will be comprehensive in addressing both technical and organizational issues of an integrated CZO data infrastructure, and that the list of tasks described below is designed to be undertaken in close collaboration and with guidance from CZO research teams.

Project Management

Our team comprises a high level of expertise, bringing together a group of investigators who have successfully developed and implemented effective cyberinfrastructure and data systems such as the CUAHSI HIS, IEDA, and EarthChem. We add to our team strong collaboration from CZO representatives who can ensure effective communication between our project team, CZO PIs, and CZO information managers.

Internal project management will involve:

1. Regular bi-weekly phone conferences of the team members
2. Code sharing infrastructure, as well as mailing lists and document sharing
3. Close interaction with CZO PIs and the Information Management Committee to steer project development (Task 0)
4. Continuous interaction with site information managers (Tasks 0 and 7)
5. Regular reporting on the progress of the project, to seek feedback from CZO PIs (Task 0)

Project Tasks

The following is a list of project tasks that were expanded from the project description with additional information within the statements of work from each of the project collaborators and the proposal's Data Management Plan. Within the list, project collaborators who will be responsible for each each major and sub-task are indicated. Following the list is a detailed matrix that summarizes this information and provides a timeline..

New Task 0. Management and community involvement, as described in proposal section 4 (BC-CZO, CR-CZO, with others)

   a. Instigate and Support an Information Management Committee (IMC) (BC-CZO, CR-CZO, Q1-Q8)
   b. Web-based information events and workshops (at least 2 per year); mailing lists (BC-CZO, CR-CZO, Q1-Q8)
   c. Synthesis working group (two workshops) (BC-CZO, CR-CZO, Q3 and Q7)
   d. Subdiscipline workshops (three workshops, in geochemistry, GIS data management; hydrologic data management) (BC-CZO, CR-CZO, GfG, Q2, Q5, Q8)

2. Task 1. CZO-wide Information Portal (BC-CZO, CR-CZO, SDSC).
   a. Develop the CZO online content management system (BC-CZO, Q1-Q3)
   b. Further iterate on the CZO CMS, add content (BC-CZO and CR-CZO, Q4-Q8)
   c. Host CZO online content management system, and, in collaboration with BC-CZO, integrate it with the CZO data discovery portal (SDSC: integrate with data discovery portal: Q4-Q8; host the CMS: Q1-Q8)
3. Task 2. Consistent metadata (All).

    a. Extend CZO display file format to geochemical data (GfG, UCHIC, SDSC; Q2-Q4)
    b. Extend CZO display file format to spatial data (BC-CZO, SDSC, UCHIC; Q3-Q5)
4. Task 3. Data and metadata publication tools and templates (UCHIC, SDSC, GfG).
    a. Develop metadata forms and tools to support automatic generation of display files from local CZO data management systems, in particular metadata files for time series (UCHIC, SDSC, Q2-Q4)
    b. Develop similar templates for geochemical data submission (GfG, UCHIC, Q3-Q5)
    c. Develop similar templates for spatial data submission (SDSC; Q4-Q6)
5. Task 4. Web-enable CZchemDB (GfG).
    a. Transfer CZchemDB to new relational database management system (RDBMS) (GfG, Q1-Q2)
    b. Develop specifications for CZchemDB User Interface (GfG, Q3-Q4)
    c. Develop web application for CZchemDB UI (GfG, Q5-Q8)
6. Task 5. Shared vocabulary system (UCHIC, SDSC, SWRC, CZO scientists).
    a. Extend the existing vocabulary system to variables and other terms used by CZOs (UCHIC, Q1-Q8)
    b. 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).
    c. Setup a SKOS system for CZO controlled vocabularies (SDSC, Q4-Q8)
7. Task 6. Implement IGSN Registration Agent for CZO samples (GfG).
    a. Develop specification for CZO object types and metadata profiles (GfG, Q1-Q2)
    b. Build the CZO IGSN Registration Agent (GfG, Q3-Q6)
8. Task 7. Support for site data managers (SDSC, UCHIC, BC-CZO).
    a. 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, BC-CZO, Q1-Q2)
    b. 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, BC-CZO, Q3-Q8)
    c. Bi-weekly VTC with data managers (SDSC, Q1-Q8)
    d. Maintain a document and file sharing portal for data managers (password protected). (SDSC, Q1-Q8)
9. Task 8. CZO Central metadata catalog, data discovery portal and harvester (SDSC, UCHIC).
    a. 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, GfG, Q5-Q7).
    b. Assist the GfG team in development of standards-based web services for geochemical data and registering them to the CZO data portal (SDSC, Q5-Q8)
    c. 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)
    d. 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).
    e. Enhance the CZO online system that assists data managers in registering and managing their observation networks (password protected) (SDSC, Q1-Q6)
    f. Add DataONE Member Node interfaces to CZO Central (SDSC, UCHIC, Q6-Q8)
10. Task 9. Central data repository of time series and point data (SDSC).

a. 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)
b. 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)
c. Develop reliable hardware and software environment for the central data repository (redundant mirrored databases and web servers) (SDSC, Q3-Q5)
d. 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).

11. Task 10. Publication and sharing of spatial data and LiDAR (SDSC)
    a. Continue working with the OpenTopography project to utilize the LiDAR project resources for archiving and processing of CZO LiDAR data (SDSC, Q5-Q8).
        i. 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/)
    b. Register available CZO LiDAR data managed by OpenTopography, via the CZO data portal (SDSC, Q5-Q8)
    c. Replicate and manage CZO spatial data at the CZO central data repository at SDSC (SDSC, Q3-Q8)
12. Task 11. Integrate CZchemDB data into the EarthChem system (GfG)
    a. Develop scripts and procedures to convert the CZchemDB data into XML documents (GfG, Q1-Q3)
    b. Develop web services that will harvest the XML encoded data from CZchemDB for inclusion in the EarthChem Portal database (GfG, Q4-Q6)
    c. Update EarthChem vocabularies to include terminology used by CZchemDB, and integrate with CZO shared vocabulary system (GfG, Q5-Q8)
13. Task 12. Visualization tools (APL).
    a. Enhance NVS framework with generic time-handling components (APL, Q1-Q4)
    b. 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)
    c. 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)
    d. Implement search, access and visualization capabilities that integrate across CZO sites; enhance data selection and download tools (APL, Q5-Q8)
    e. 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)

Table 1. Summary matrix of project tasks and collaborator responsibilities.

Project Tasks	BC-CZO	GfG	SDSC	CR-CZO	UCHIC	APL
	M. Williams (1 semester/year teaching relief), Lubinski/ Parrish (8 mo total), 6 mo in yr 2 for staff liason to central CZO office	K. Lehnert (0.5 cal mon/year), C. Chan (1 cal mon/year), IT Developer TBH (10 cal mon/year)	I.Zaslavsky (1.2 cal mon/year), T. Whitenack (10.9 cal mon/year)	A. Aufdenkampe (8k/year stipend ≈ 0.4 mo/year)	J. Horsburgh (0.5 cal mon/year), D. Tarboton (0.25 cal mon/year), K. Schreuders (1.5 cal mon/year)	E. Mayorga (1.05 cal mon/year), T. Tanner (1.15 cal mon/year)
0. Community	a:Q1-Q8 b:Q1-Q8 c:Q3,Q7 d:Q2,Q5,Q8	d: Q2		a:Q1-Q8 b:Q1-Q8 c:Q3,Q7 d:Q2,Q5,Q8		
1	a:Q1-Q3 b:Q4-Q8		c:Q1-Q8	b:Q4-Q8		
2	a:Q2-Q4 b:Q3-Q5	a:Q2-Q4	a:Q2-Q4 b:Q3-Q5	a:Q2-Q4	a:Q2-Q4 b:Q3-Q5	
3		b:Q3-Q5	a:Q2-Q4 c:Q4-Q6		a:Q2-Q4 b:Q3-Q5	
4		a:Q1-Q2 b: Q3-Q4 c: Q5-Q8				
5			b:Q1-Q8 c:Q4-Q8		a:Q1-Q8	
6		a:Q1-Q2 b:Q3-Q6				
7	a:Q1-Q2 b:Q3-Q8		a:Q1-Q2 b:Q3-Q8 c:Q1-Q8 d:Q1-Q8		a:Q1-Q2 b:Q3-Q8	
8			a:Q5-Q7 b:Q5-Q8 c:Q3-Q6 d:Q5-Q8 e:Q1-Q6 f:Q6-Q8		g:Q6-Q8	
9			a:Q1-Q8 b:Q2-Q4			

			c:Q3-Q5 d:Q5-Q7			
10			a:Q5-Q8 b:Q5-Q8 c:Q3-Q8			
11		a:Q1-Q3 b:Q4-Q6 c:Q5-Q8				
12			c:Q2-Q7			a:Q1-Q4 b:Q1-Q6 c:Q2-Q7 d:Q5-Q8 e:Q4-Q8

Future of Integrated Data Management for CZOs

The proposal here is to develop an integrated data management system for the CZOs over the next two years, but the projected lifespan of the CZOs will go well beyond that timeframe. We are taking those disparate timescales into consideration throughout this project in the following vision:

1. **Future data system oversight and maintenance.** All data systems need to continually be maintained and evolve to meet future needs. Our care to build upon existing and widely used standards, technologies and systems maximizes the probability of continuing independent maintenance and enhancement of most of the foundational components of the CZO system. In addition, all code that is custom developed for the CZO data system will be well documented and managed via public online version control system following norms for open source software projects (i.e. such as the codeplex system used by CUAHSI, http://hydroserver.codeplex.com/). This will not only facilitate our own project collaboration but enable others to participate in code development in collaboration with us during this project and after, or even take over where we leave off if funding for our team does not get renewed. Our understanding is that a national CZO program central office is likely to be formed before this 2 year project is completed, and will likely take on high-level oversight of the integrated CZO data system, among many other activities.
    a. **Transitioning the data management system to CZO central.** We envision working closely with that office before the completion of this project's funding in 2014 to ensure longevity and evolution of the system that we will develop. To that end, we have budgeted 6 months of funding in year 2 for a staff member (Boulder) to serve as a technical liason to the future CZO central office. However, it is the experience of team members over many projects that handing a system from developers to a production/support team doesn't usually happen quickly and may not be complete by 2014, depending on many factors.
    b. **Future costs** for system oversight and maintenance after 2014 are likely to total at least **$250k-$500k/year**, and include:

i. 1-2 months/year of salary support is a minimum for the future CZOData PI to maintain high-level system oversight. The future CZOData PI should be linked with CZO National Program Central Office, and perhaps be the Director, a staff member or an affiliate faculty, which could include one or more members of the current CZOData team. Regardless, the future CZOData PI will need very strong knowledge of the universe of efforts in hydro/geo/eco-informatics in order to effectively oversee future maintenance, because CZOData is built upon a wide-ranging foundation of cyber-infrastructures serving different domains. Thus an integrated team of subcontracted specialists will most likely perform much of the actual work, but that team needs to be well coordinated by a PI who is current and active in the evolving world of cyber-infrastructures for earth and environmental sciences.

ii. CI specialists will be required to not only oversee the data server and database infrastructure, but also to maintain the functionality of web services, data ingestion from CZOs, shared vocabulary/ontology standards, catalogs, web-portals, visualization systems, archiving systems, code management, wikis and tutorials, etc. Although we are designing the present CZOData system in a way that could be taken over by a new team if necessary, significant efficiencies would be gained by maintaining the present CZOData development team as sub-contractors for maintenance functions. Estimated costs might be a minimum of $200-400k/year, and distributed between 2-4 subcontractors. Additional funding for "support" would be required if the current CZOData team were not subcontracted to continue with system maintenance.

2. **Future data system enhancement.** Although we expect to deliver a solid data system by the end of the proposed 2-year project, to fulfill the envisioned potential of a central CZO data system we foresee a need to develop several additional enhancements that are not budgeted within the scope of this project. These could either be pursued in parallel with current efforts but with separate funding, or become part of a phase 3 plan starting in late 2014 and overseen by the CZO central office. These envisioned enhancements include:
    a. **A shared information model,** which we are calling the Observations Data Model 2.0 (ODM2), which would serve as a foundation for complete interoperability between CUAHSI-HIS, EarthChem, CZOData and potentially many other programs such as IOOS, DataONE, LTERs, NEON and OGC. In short, we would like rewrite the core logical structure of all three data systems so that they would be identical, and then work through the OGC process to develop it as an international standard, similar to what CUAHSI and EarthChem have each done with WaterML and GeochemML. The benefits of this can not be understated. Our team has just submitted a separate proposal to the NSF Geoinformatics program to develop the shared information model with engagement from the broad community, and to develop prototype implentations for testing and demonstration. This effort should be pursued until completed, and will likely cost a **total of $2.5M to $3.5M** to fully implement, as detailed below.

i. Our team just submitted a proposal to GeoInformatics (EAR-1224638) for $551,176 to develop and prototype a functioning system (i.e. distributed databases with web services, central cataloging and central archiving). This is step one, and will hopefully be well underway by the end of the CZOData Phase II project (2012-2014).

ii. Full implementation of a complete functioning system for all CZO data, including adopting/developing shared vocabulary/ontologies for new data types not previously addresses (i.e. phylogentic data, molecular biology data), extending the information system to handle these new data types, rebuilding all the currently used portals and tools, and testing. This could be done by the current CZOData team and begin by overlapping with the second year of step one, above, and would likely take 2-4 additional years at a total cost of approximately $2M-$4M.

b. **A shared modeling platform**. The Community Surface Dynamics Modeling System (CSDMS) and CUAHSI's Community Hydrologic Modeling Platform (CHyMP) have both separately developed shared modeling platforms for Earth surface science communities. These existing efforts have made substantial progress but have not yet achieved the full vision that that we see for a seamless, highly efficient use of CZO data by models in high-performance cloud environments. We envision very large benefits from partnering with these two efforts to create such a seamless system.

i. A very rough approximation of development costs might be **$1M-$4M total**, depending on how much synergism is possible with existing efforts such as CSDMS and CUAHSI-CHyMP. The team to do this would need to evolve out of a mashup between many different groups in the greater community, and perhaps under the umbrella of a Scientific Software Innovation Institute (NSF solicitation 11-589) dedicate to environmental observatories.

c. **Other**. It is safe to say that there will be future needs that will become apparent over time. Many may become very clear in just the next 6-12 months, due to the substantial community engagement activities that are planned by the pending CZOData Phase II project. Staying relevant to the evolving needs of advanced CZ science is critical to the future success of a central CZOData system.

3. **CZO central data repositories and archives**. CZO data is currently being served by individual CZOs and collected into a central repository at SDSC. However, none of this data is being archived in a way that ensures its availability beyond the currently funded project periods (see section below for details). As part of this project we will work with the CZO community over the next 2 years to develop recommendations for how to address this issue after 2014.

a. **Data archiving**. There are several parts to archiving, some of which we commit to in this proposal, and others that need to be pursued with additional funds:

i. Creating a versioned database of time series, so that we keep old versions that might have been overwritten/modified at CZO sites - yes;

ii. Add DataOne interfaces -yes;

iii. Store both the database and display files - retrieved from web sites and validated - yes;
iv. Providing access to data as encoded in standard formats (which is what archivable formats typically are) - yes;
v. Providing redundant storage, backups, etc. - yes;
vi. Committing to managing the data beyond the 2-year project period - no, at the moment.

b. **Future Costs** of long-term data archiving are likely to be quite modest (i.e. $5k-$15k per year total). SDSC recently rolled out a cloud storage facility. It is a very cost effective way to provide access to data long term (because no data movement fees, and also you can partly charge it to equipment, so less IDC.) Also, provides a Dropbox-like interface. For example, in one recent grant I added additional 5-year archiving support to the last year of the project, which cost, for 8.5 Tb of dual-copy storage, $5K in equipment and $1260/year in maintenance."
    i. Do we want to consider pre-paying for future data archiving, for a period of 20-50 years?

Response to Specific Reviewer Concerns

Reviewer #1 described the need to "address the cultural issues of [shared/controlled] vocabulary maintenance (the potentially contentious process relating to extension and deprecation of terms) and governance."

- Response: We understand through first-hand experience the critical importance of the human dimension in managing any integrated data management system. Our proposal reflects this need with the strong emphasis on engaging the community of CZO scientists and data managers (section 4 and now outlined under Task 0), with $175,945 budgeted to participant support and supplies for workshops and web-conferences

Reviewer #2 expressed no concerns.

Reviewer #3 described four concerns:

1. "a more detailed [technical?] explanation of how this [cyber-infrastructure] integration will be achieved."
    - Response: The team has presented several technical descriptions of how cyberinfrastructure integration will be accomplished at the technical level, including a paper presented at EIM'2011 (Zaslavsky, I., T. Whitenack, M. Williams, D. G. Tarboton, K. Schreuders, and A. Aufdenkampe, (2011) "The Initial Design of Data Sharing Infrastructure for the Critical Zone Observatory", in *Proceedings of the Environmental Information Management Conference*, Santa Barbara, CA, 28-29 September, EIM'2011, pp. 145-150), an EarthCube white paper (I. Zaslavsky, M. Williams, A. Aufdenkampe, K. Lehnert, E. Mayorga, J. Horsburgh (2011) Data Infrastructure for the Critical Zone Observatories (CZOData): an EarthCube Design Prototype

(http://earthcube.ning.com/group/earthcube-design-approaches/forum/topics/white-paper-data-infrastructure-for-the-critical-zone), and several AGU abstracts. Without repeating the content of these texts, here we'll list the key data integration principles we follow in the technical design of CZOData. These principles derive from requirements expressed by CZO PIs and researchers, and follow state-of-the-art technical architecture developed in other large scale successful projects, leveraging the best practices and accomplishments of these projects. They include:

i. reliance of international information encoding and service standards, e.g. developed by OGC (through close involvement with the OGC community standardization process; co-PI Zaslavsky is a co-chair of the OGC/WMO Hydrology Domain Working Group)
ii. supporting interoperability at several levels: catalogs and discovery services (using the OGC Catalog Services for the Web (CSW) standard interface); shared vocabularies (using Simple Knowledge Organization System (SKOS) interface; standard data access services (using OGC Sensor Observation Service (SOS) for exchanging observations data and other OGC service standards (WMS, WFS) for exchanging spatial data); following standard information encodings, based on OGC Observations & Measurements specification to present observational data (a profile of which, WaterML 2.0, has been just submitted to OGC for approval as international standard, at the end of 2011), and GML for spatial data
iii. developing tools and services to bring legacy infrastructure components in line with the standards listed above,
iv. developing a common CZO data discovery portal, to support online discovery of resources of different types
v. reliance on standard service interfaces and information models will result in easier use of third-party analysis and visualization tools

2. "proposal's aims are perhaps more grandiose than its requested funding and staffing profile allow."
    - Response: Indeed the need and vision for an integrated CZO data management system is great, and certain components of that vision needed to be spun off for separate funding (i.e. the effort to develop a second generations Observations Data Model (ODM 2.0) is being submitted to NSF Geoinformatics this week). However, the tasks proposed here substantially leverage prior and existing efforts by the project team (see section 2 of proposal). Although we will need to prioritize efforts judiciously we believe that this projects goals are attainable with the budgeted time and funding.
3. "there is little discussion of contingencies"
    - Response: As described above, we believe our goals are achievable because of the substantial synergies with complementary efforts by project team members. For example, the migrating CZChemDB from MS Access to a web-accessible PostgreSQL database management system (Task 4) is a relatively straightforward task for the GfC team because they co-developed CZChemDB with S. Brantley's group with that future in mind and because such a system would closely mimic GfC's existing PetDB, SedDB and

VentDB database infrastructure. Likewise, to develop a DataOne Member Node interface to CZOcentral (Task 8) is relatively straightforward. Project member J. Horsburgh is a member of the Core Cyberinfrastructure Team and Co-Leading the DataONE Data Integration and Semantics Working Group. Horsburgh has confirmed that specifications are in place for us to develop CZO data interfaces to a DataOne Member Node.

4. "not clear whether the success of this project will contribute to easing cross-organizational programmatic data access issues." In particular, because "CZO site PIs to control their own data... [this] disallows true data integration, because uncertainty estimation cannot be performed reliably without access to datum-level data and metadata transparency."
    - Response: While CZO PIs have ultimate control of their data, the proposal describes harvesting the data (with sufficient metadata as defined in the CZO display file specification) into a central database, cataloguing and indexing them, and making them available via standard services. The prototype implementation of CZOdata to date has already demonstrated that hydrologic time series information collected by all CZOs can be accessed seamlessly via a standard set of web services (following the same service specification as used in CUAHSI HIS and adopted by several federal agencies including USGS). In a sense, this service-based approach already allows cross-organizational programmatic data access. In this proposal, we are planning to extend this approach to managing other types of data collected across different CZO sites, in particular geochemical data.

Reviewer #4 felt "that the implementation plan leaves a lot to be desired."

- Response: This document provides a detailed description of tasks, time line and responsibilities of project partners, clarifying possible omissions of the proposal. In particular, we demonstrate that the 12 tasks (with the added management task) represent a coherent and tightly integrated effort by a distributed group of researchers and developers, which is tightly integrated with the needs of the national CZO program and provides specific steps enabling cross-CZO data access and integration

Reviewer #5 described a need for close feedback from CZO scientists to "avoid this infrastructure becoming "yet-another" legacy system that cannot be maintained by its user base," including "looking at long-term operational costs (how will this group streamline development and reduce the amount of CZO custom (and task-custom) code to reduce the ongoing costs of maintaining these capabilities?)."

- Response: efficient on-going maintenance and involving CZO the (and larger research) community in operating the system is a serious concern of the project team, as described above in "Future of Integrated Data Management for CZOs." There are several specific steps that the project describes, to make the proposed infrastructure a success:
    - reliance on standards and standards-based software, to minimize custom code development
    - reliance on best practices of software development and code management, compilation and testing, to minimize long-term code management burden

- close integration and synergies with existing operational infrastructures and NSF-supported projects (CUAHSI HIS, EarthChem, OpenTopography, DataONE)
- significant community engagement and outreach effort described in Tasks 0 and 7, aimed at engaging broad groups of scientists and students in research projects that use the developed infrastructure, and making this use as easy and straightforward as possible via proven service interfaces and analytical applications

Reviewer #6 had two concerns:

1. "the applicability of hydraulic data sets of fluids to the far-less mixed solid phases of the CZ, and eventually to the gaseous fractions of theses systems"
    - Response: Although the currently functional prototype CZO data system indeed focuses on hydrologica data sets, this proposal will specifically develop a system of web services for CZChemDB -- which focuses on soil and other subsurface sample fraction geochemistry data -- to be interoperable with the current system (see tasks 3, 4, 5, 6).
2. "who does what is not at all clear" and "relatively little of [the proposal text describes] what the parts will do to complement each other."
    - Response: We have substantially clarified the integration of project tasks and responsibilities in the text above.