COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

PROGRAM ANNOUNCEMENT/SOLICITATION NO./CLOSING DATE/if not in response to a program announcement/solicitation enter NSF 11-1

NSF 11-1

FOR NSF USE ONLY

NSF PROPOSAL NUMBER

1153164

FOR CONSIDERATION BY NSF ORGANIZATION UNIT(S) (Indicate the most specific unit known, i.e. program, division, etc.)

EAR - SURFACE EARTH PROCESS SECTION

DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	DUNS# (Data Universal Numbering System)	FILE LOCATION
08/01/2011	3	06030000 EAR	7570	007431505	08/03/2011 11:18am S

EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN): **846000555**

SHOW PREVIOUS AWARD NO. IF THIS IS
☐ A RENEWAL
☐ AN ACCOMPLISHMENT-BASED RENEWAL

IS THIS PROPOSAL BEING SUBMITTED TO ANOTHER FEDERAL AGENCY? YES ☐ NO ☒ IF YES, LIST ACRONYM(S)

NAME OF ORGANIZATION TO WHICH AWARD SHOULD BE MADE: **University of Colorado at Boulder**

ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE:
3100 Marine Street, Room 481
572 UCB
Boulder, CO 80309-0572

AWARDEE ORGANIZATION CODE (IF KNOWN): **0013706000**

NAME OF PRIMARY PLACE OF PERF: **University of Colorado at Boulder**

ADDRESS OF PRIMARY PLACE OF PERF, INCLUDING 9 DIGIT ZIP CODE:
University of Colorado at Boulder
3100 Marine Street, Room 481
Boulder ,CO ,803031058 ,US.

IS AWARDEE ORGANIZATION (Check All That Apply) (See GPG II.C For Definitions)
☐ SMALL BUSINESS ☐ MINORITY BUSINESS
☐ FOR-PROFIT ORGANIZATION ☐ WOMAN-OWNED BUSINESS

☐ IF THIS IS A PRELIMINARY PROPOSAL THEN CHECK HERE

TITLE OF PROPOSED PROJECT: **Integrated Data Management System for Critical Zone Observatories**

REQUESTED AMOUNT	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PRELIMINARY PROPOSAL NO. IF APPLICABLE
$ 1,503,590	24 months	10/01/11	

CHECK APPROPRIATE BOX(ES) IF THIS PROPOSAL INCLUDES ANY OF THE ITEMS LISTED BELOW
☐ BEGINNING INVESTIGATOR (GPG I.G.2)
☐ DISCLOSURE OF LOBBYING ACTIVITIES (GPG II.C.1.e)
☐ PROPRIETARY & PRIVILEGED INFORMATION (GPG I.D, II.C.1.d)
☐ HISTORIC PLACES (GPG II.C.2.j)
☐ EAGER* (GPG II.D.2) ☐ RAPID** (GPG II.D.1)
☐ VERTEBRATE ANIMALS (GPG II.D.6) IACUC App. Date ________
PHS Animal Welfare Assurance Number ________
☐ HUMAN SUBJECTS (GPG II.D.7) Human Subjects Assurance Number ________
Exemption Subsection ________ or IRB App. Date ________
☐ INTERNATIONAL COOPERATIVE ACTIVITIES: COUNTRY/COUNTRIES INVOLVED (GPG II.C.2.j)
____
☐ HIGH RESOLUTION GRAPHICS/OTHER GRAPHICS WHERE EXACT COLOR REPRESENTATION IS REQUIRED FOR PROPER INTERPRETATION (GPG I.G.1)

PI/PD DEPARTMENT: **Institute of Arctic and Alpine Research**

PI/PD POSTAL ADDRESS:
450 UCB
Boulder, CO 803090450
United States

PI/PD FAX NUMBER: **303-492-6388**

NAMES (TYPED)	High Degree	Yr of Degree	Telephone Number	Electronic Mail Address
PI/PD NAME Mark W Williams	PhD	1991	303-492-8830	markw@snobear.colorado.edu
CO-PI/PD Anthony K Aufdenkampe	PhD	2002	610-268-2153	aufdenkampe@stroudcenter.org
CO-PI/PD				
CO-PI/PD				
CO-PI/PD				

Electronic Signature

CERTIFICATION PAGE

Certification for Authorized Organizational Representative or Individual Applicant:

By signing and submitting this proposal, the Authorized Organizational Representative or Individual Applicant is: (1) certifying that statements made herein are true and complete to the best of his/her knowledge; and (2) agreeing to accept the obligation to comply with NSF award terms and conditions if an award is made as a result of this application. Further, the applicant is hereby providing certifications regarding debarment and suspension, drug-free workplace, lobbying activities (see below), responsible conduct of research, nondiscrimination, and flood hazard insurance (when applicable) as set forth in the NSF Proposal & Award Policies & Procedures Guide, Part I: the Grant Proposal Guide (GPG) (NSF 11-1). Willful provision of false information in this application and its supporting documents or in reports required under an ensuing award is a criminal offense (U. S. Code, Title 18, Section 1001).

Conflict of Interest Certification

In addition, if the applicant institution employs more than fifty persons, by electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative of the applicant institution is certifying that the institution has implemented a written and enforced conflict of interest policy that is consistent with the provisions of the NSF Proposal & Award Policies & Procedures Guide, Part II, Award & Administration Guide (AAG) Chapter IV.A; that to the best of his/her knowledge, all financial disclosures required by that conflict of interest policy have been made; and that all identified conflicts of interest will have been satisfactorily managed, reduced or eliminated prior to the institution's expenditure of any funds under the award, in accordance with the institution's conflict of interest policy. Conflicts which cannot be satisfactorily managed, reduced or eliminated must be disclosed to NSF.

Drug Free Work Place Certification

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Drug Free Work Place Certification contained in Exhibit II-3 of the Grant Proposal Guide.

Debarment and Suspension Certification

(If answer "yes", please provide explanation.)

Is the organization or its principals presently debarred, suspended, proposed for debarment, declared ineligible, or voluntarily excluded from covered transactions by any Federal department or agency? Yes ☐ No ☒

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Debarment and Suspension Certification contained in Exhibit II-4 of the Grant Proposal Guide.

Certification Regarding Lobbying

The following certification is required for an award of a Federal contract, grant, or cooperative agreement exceeding $100,000 and for an award of a Federal loan or a commitment providing for the United States to insure or guarantee a loan exceeding $150,000.

Certification for Contracts, Grants, Loans and Cooperative Agreements

The undersigned certifies, to the best of his or her knowledge and belief, that:
(1) No federal appropriated funds have been paid or will be paid, by or on behalf of the undersigned, to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with the awarding of any federal contract, the making of any Federal grant, the making of any Federal loan, the entering into of any cooperative agreement, and the extension, continuation, renewal, amendment, or modification of any Federal contract, grant, loan, or cooperative agreement.
(2) If any funds other than Federal appropriated funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with this Federal contract, grant, loan, or cooperative agreement, the undersigned shall complete and submit Standard Form-LLL, ''Disclosure of Lobbying Activities,'' in accordance with its instructions.
(3) The undersigned shall require that the language of this certification be included in the award documents for all subawards at all tiers including subcontracts, subgrants, and contracts under grants, loans, and cooperative agreements and that all subrecipients shall certify and disclose accordingly.

This certification is a material representation of fact upon which reliance was placed when this transaction was made or entered into. Submission of this certification is a prerequisite for making or entering into this transaction imposed by section 1352, Title 31, U.S. Code. Any person who fails to file the required certification shall be subject to a civil penalty of not less than $10,000 and not more than $100,000 for each such failure.

Certification Regarding Nondiscrimination

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative is providing the Certification Regarding Nondiscrimination contained in Exhibit II-6 of the Grant Proposal Guide.

Certification Regarding Flood Hazard Insurance

Two sections of the National Flood Insurance Act of 1968 (42 USC §4012a and §4106) bar Federal agencies from giving financial assistance for acquisition or construction purposes in any area identified by the Federal Emergency Management Agency (FEMA) as having special flood hazards unless the:
(1) community in which that area is located participates in the national flood insurance program; and
(2) building (and any related equipment) is covered by adequate flood insurance.

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant located in FEMA-designated special flood hazard areas is certifying that adequate flood insurance has been or will be obtained in the following situations:
(1) for NSF grants for the construction of a building or facility, regardless of the dollar amount of the grant; and
(2) for other NSF Grants when more than $25,000 has been budgeted in the proposal for repair, alteration or improvement (construction) of a building or facility.

Certification Regarding Responsible Conduct of Research (RCR)
(This certification is not applicable to proposals for conferences, symposia, and workshops.)

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative of the applicant institution is certifying that, in accordance with the NSF Proposal & Award Policies & Procedures Guide, Part II, Award & Administration Guide (AAG) Chapter IV.B., the institution has a plan in place to provide appropriate training and oversight in the responsible and ethical conduct of research to undergraduates, graduate students and postdoctoral researchers who will be supported by NSF to conduct research.
The undersigned shall require that the language of this certification be included in any award documents for all subawards at all tiers.

AUTHORIZED ORGANIZATIONAL REPRESENTATIVE	SIGNATURE	DATE
NAME **Joan H Eaton**	**Electronic Signature**	**Aug 1 2011 7:27PM**

TELEPHONE NUMBER	ELECTRONIC MAIL ADDRESS	FAX NUMBER
303-492-2691	**Joan.Eaton@colorado.edu**	**303-492-6421**

* EAGER - EArly-concept Grants for Exploratory Research
** RAPID - Grants for Rapid Response Research

Integrated Data Management System for Critical Zone Observatories
Mark Williams and Anthony Aufdenkampe

Project Summary

This proposal requests funding to develop an integrated data management system for the NSF-funded Critical Zone Observatories (CZOs). Our goal is to support, empower, and broaden the impact of CZO science and maximize the return on investment of the CZO program by ensuring the preservation and open sharing of data acquired by the CZOs. We will engage the CZO community to build the CZO integrated data management system (CZOData) upon several extant, NSF-funded cyberinfrastructure (CI) systems, to facilitate seamless and integrated access to and analysis of the wide range of multi-disciplinary data within and across CZOs. The CZOData design is based on the experience of the project team in building the current CZOData prototype. The design follows the general SOA "publish-find-bind" pattern. In particular, these requirements affect the "publish" component which is represented as two interlinked modules: publishing CZO data at individual web sites as human-readable ASCII files, and their harvesting and republishing as standard-compliant web services at the central CZO data portals. The CZO Central System for point-based time series and the CZchem System for geochemical data will store, index, analyze, and publish the data via web services using WaterML and EarthChemXML encodings, respectively. These web services allow the CZO portals to send data and metadata to CUAHSI HIS and EarthChem, allowing CZO users to take full advantage of those systems.

Intellectual Merit:
Here we have a unique opportunity to integrate environmental research with data management and CI from the ground up. A major contribution of this project is the integration of site data managers with critical zone researchers. As a result of this project, an interactive, distributed data publication and sharing system for diverse data will be developed and made operational for the six participating CZO sites. Once operational, it will enable researchers and students to quickly discover and access a variety of CZO data, including time series of water and atmosphere measurements, observations derived from geochemical samples, high density LiDAR and other vector and raster spatial data. CZOData will significantly ease data discovery and integration for CZO researchers and collaborators, thus a) supporting new tasks that were not possible before, such as cross-CZO comparison and modeling, and b) easing the data management burden on individual researchers and data managers and thus letting them focus on science issues.

Broader Impacts:
By incorporating system research and development with graduate education, we will move towards cultivating a new generation of researchers skilled in both different facets of environmental research and in advanced information management and computing. The collaboration between CUAHSI HIS, IEDA and CZO data management teams on this proposal, has, during proposal development, and will in the ongoing project, allow us to identify common elements in the data models used by the different disciplines involved and work towards a generalized data model that brings together and serves the broader interests of the multiple disciplines involved. Exposing CZO data as services compliant with OGC specifications will enable a higher level of interoperability with data and software developed by other groups and will lead to a sustainable operational model with linkages to IEDA, EarthChem, DataONE, OpenTopography, LTER, NEON, and other environmental observatories and systems.

TABLE OF CONTENTS

For font size and page formatting specifications, see GPG section II.B.2.

	Total No. of Pages	Page No.* (Optional)*
Cover Sheet for Proposal to the National Science Foundation		
Project Summary (not to exceed 1 page)	1	
Table of Contents	1	
Project Description (Including Results from Prior NSF Support) (not to exceed 15 pages) **(Exceed only if allowed by a specific program announcement/solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)**	15	
References Cited	3	
Biographical Sketches (Not to exceed 2 pages each)	12	
Budget (Plus up to 3 pages of budget justification)	28	
Current and Pending Support	14	
Facilities, Equipment and Other Resources	5	
Special Information/Supplementary Documents (Data Management Plan, Mentoring Plan and Other Supplementary Documents)	2	
Appendix (List below.) **(Include only if allowed by a specific program announcement/ solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)**		

Appendix Items:

*Proposers may select any numbering mechanism for the proposal. The entire proposal however, must be paginated. Complete both columns only if the proposal is numbered consecutively.

1. INTRODUCTION

This proposal requests funding to develop an integrated data management system for the NSF-funded Critical Zone Observatories (CZOs). Our goal is to support, empower, and broaden the impact of CZO science and maximize the return on investment of the CZO program by ensuring the preservation and open sharing of data acquired by the CZOs. We will engage the CZO community to build the CZO integrated data management system (CZOData) upon several extant, NSF-funded cyberinfrastructure (CI) systems, to facilitate seamless and integrated access to and analysis of the wide range of multi-disciplinary data within and across CZOs.

The CZOs are a group of NSF-funded, site-based observatories that are dedicated to investigate in a coordinated manner Earth processes in the critical zone, which is the region between bedrock and the atmospheric boundary layer [Anderson et al., 2008]. The six CZOs are comprised of: the Southern Sierra CZO (University of California, Merced), the Boulder Creek CZO (University of Colorado at Boulder), the Shale Hills CZO (Pennsylvania State University), the Christina River Basin CZO (University of Delaware), the Jemez River Basin and Santa Catalina Mountains CZO (University of Arizona), and the Luquillo CZO in Puerto Rico (University of Pennsylvania).

The vision of the CZO program as articulated by the CZO PIs to the New Research Opportunities in the Earth Sciences (NROES) committee is to develop, within the next ten years, a robust predictive ability for how critical zone attributes, processes, and outputs will respond to projected climate and land-use changes. This predictive ability must be founded on sufficiently broad knowledge of critical zone processes to describe how the varied climatic and geologic factors that distinguish different regions interact. Over the next decade, the CZO program will produce new, fundamental understanding of the critical zone based on four-dimensional datasets that will stimulate, inspire, and test the resulting predictive models. The CZOs will: 1) develop a unifying theoretical framework of critical zone evolution; 2) develop coupled systems models to explore how critical zone services respond to anthropogenic, climatic, and tectonic forcings; and 3) develop data sets that document differing critical zone geologic and climatic settings, inform the theoretical framework, constrain conceptual and coupled systems models, and test model-generated hypotheses. Achieving these goals requires an unprecedented integration of data from different disciplines on diverse materials and over enormous ranges in space and time (Fig. 1). Thus, an integrated data management system is essential to address the vision of the CZO program.

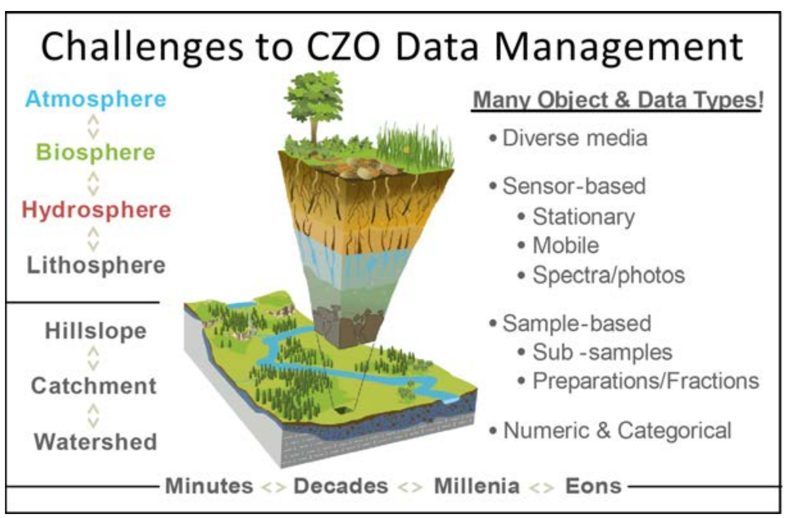

Figure 1. Critical Zone Observatories require integration of exceptionally diverse datasets

In order for the proposed CZOData to successfully fulfill its mission, its architecture and services have to be highly integrative, facilitating cross-disciplinary discovery, access, analysis, and modeling using the many diverse CZO data types. Typical research scenarios for data at CZO sites involve accessing geochemical data analyzed in the lab on samples collected at CZO sites along with hydrologic time series of water quality and quantity within experimental watersheds, relating the dynamics of hydroclimate parameters, modeling soil nutrients under different topographic, geologic, hydrologic and vegetation conditions, and analysis of fluxes across watershed boundaries. While closely connected research teams have been successful in such cross-discipline analysis and modeling using data with which they are familiar, accomplishing such integration by investigators who did not collect the data, potentially across CZO sites and at multiple spatio-temporal scales, faces a range of interoperability challenges. They stem, in particular, from differences in information models used in different disciplines and by different research groups to describe observations, differences in data representation and access, and discrepancies in metadata and their semantics. These interoperability challenges are common to emerging environmental science observatories, such as the large scale, cross-observatory systems that are being developed within the Long Term Ecological Research Network (LTER) and the National Ecological Observatory Network (NEON) and have spurred significant NSF investment in systems for data management [e.g. Tarboton et al. 2010; Horsburgh et al. 2011; Zaslavsky and Maidment 2011; DataONE 2011].

Our strategy is to support publication, discovery and integration of diverse CZO research data as comprehensively as possible, from easy-to-use tools that enable researchers and data managers to publish their quality controlled site data in a timely fashion, to tools that enable scientists within and outside the CZO program to identify, access and interpret the data of interest to them. We will achieve this using iterative and user need driven system development, and we will enhance the sustainability of the CZOData by engaging the CZO community throughout the development process. This approach will support several levels of data interoperability (Fig. 2) that will enable scientists, both within and beyond the CZO community, to perform integrated analysis and synthesis that otherwise would not be feasible, and create and sustain data legacies for the advancement of critical zone science.

The overarching goal of the proposed project is to create the essential data management information system for the CZO program to provide fundamental services for the preservation, discovery, access, analysis and synthesis of CZO-generated data across different sites and research domains, and make this infrastructure operational through close collaboration among CZO researchers and students, CZO data managers, and information system developers.

2. VISION FOR A CZO INTEGRATED DATA MANAGEMENT SYSTEM

The distributed nature of CZO research, the heterogeneity and complexity of CZO teams and computing environments, and the unique role of the CZO program as an evolving, cross-disciplinary, multi-site effort, create a range of specific requirements for CZOData. Our interaction with CZO researchers and data managers, and experience in the design of large scale environmental CI systems in neighboring fields, point to the following key components needed for a successful CZOData:

- Enabling CZO research designs that rely on finding and integrating relevant data collected by cross-CZO researchers at a scale that was not possible before. This new capability has the potential to transform how interdisciplinary research is done in the earth sciences, and in critical zone research in particular [NSF 2007; NSF 2011].
- Building on the experience of environmental observatory CI projects, which suggest multiple common requirements and infrastructure issues [e.g., CEON 2011; Dozier and Gail 2009; Lehning et al. 2009; Horsburgh et al. 2011], and leveraging several closely-related NSF-supported CI development efforts.
- Reliance on standards for data exchange adopted in research communities comprising the CZO program and on state-of-the art data cataloguing, discovery, integration, and visualization tools.
- Minimizing requirements on the individual CZO sites to reduce the impact of the proposed system on their operations, while creating centralized components that achieve cross-CZO integration and that are scalable and extensible.

- Uniform data modeling, data description and formatting practices to ensure that the CZO data and metadata published on site web sites can be unambiguously interpreted and their provenance can be traced.
- Ensuring that CZO data are available both in a human-readable form at individual CZO websites and via interoperable web services from the centralized components of CZOData, to enable programmatic access to data from multiple client applications.

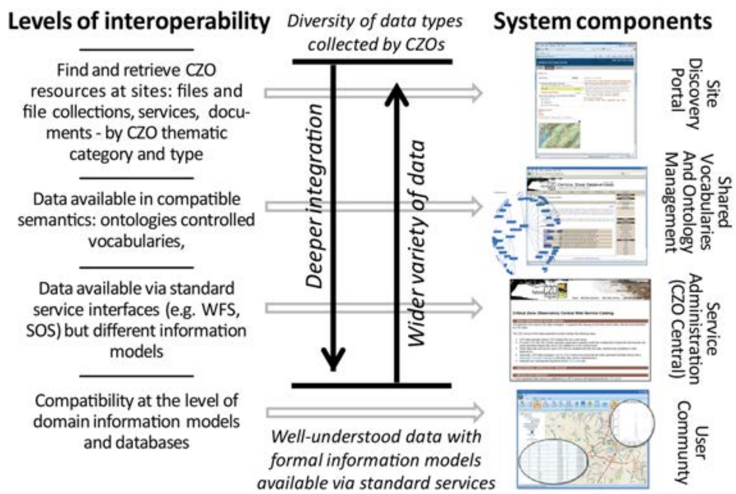

Figure 2. Levels of data interoperability, and corresponding components of CZOData.

CZOData is being designed as an evolving operational system, capable of supporting cross-site and cross-domain data integration at different interoperability levels (Fig. 2.) At the first level, different types of CZO data resources will be registered at a CZO data portal so that they can be discovered and retrieved regardless of their type or domain system that manages them. At the next level, the resources will have compatible semantics, implemented as shared vocabularies for variable names, methods, units, etc., to ensure that resource metadata can be unambiguously interpreted in a cross-disciplinary setting. Further, resources of certain types will become available via standard web service interfaces, such as those developed by the Open Geospatial Consortium (OGC), so that they can be accessed programmatically from standards-aware client applications. Finally, at the fourth level, the data that are available via standard web service interfaces will use standard encodings that reflect domain information models, to enable a much wider range of operations across different compliant sources.

As different sub-domains (geochemistry, hydroclimatology, geophysics) of CZO evolve towards emerging community agreements and standards in information modeling, service interfaces and vocabularies, CZOData will adapt to these changes by supporting all levels of interoperability described in Fig. 2. This extensible and scalable design for cross-domain and cross-site interoperability within the CZO data system is detailed in the next section.

3. OVERALL SYSTEM DESIGN

Our overall design goal is to facilitate seamless and integrated access to and analysis of the wide range of multi-disciplinary data within and across CZOs without reinventing solutions developed earlier. In this proposed phase of CZOData development, we will focus on CZO geochemical and timeseries data (e.g., climate, hydrological), and provide linkages to OpenTopography (LiDAR repository), the System for Earth Sample Registration (SESAR), Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) Hydrologic Information System (HIS), EarthChem, and Data Observation Network for Earth (DataONE) projects. The foundation for integrating these components will be the existing CZO Data Management System Prototype developed as a pre-cursor to this project [Zaslavsky et al. 2011].

3.1 Existing Systems

The CZO Data Management System Prototype. Developed over the past two years, the prototype implements the key components shown in Fig. 2. While it focuses primarily on publishing hydrologic observations collected at CZO sites and leverages service oriented architecture approaches and some software components developed by the CUAHSI HIS project [Tarboton et al. 2009], it was developed to be extensible to other types of data. The general CZO data publication workflow envisioned in the prototype is shown in Fig. 3.

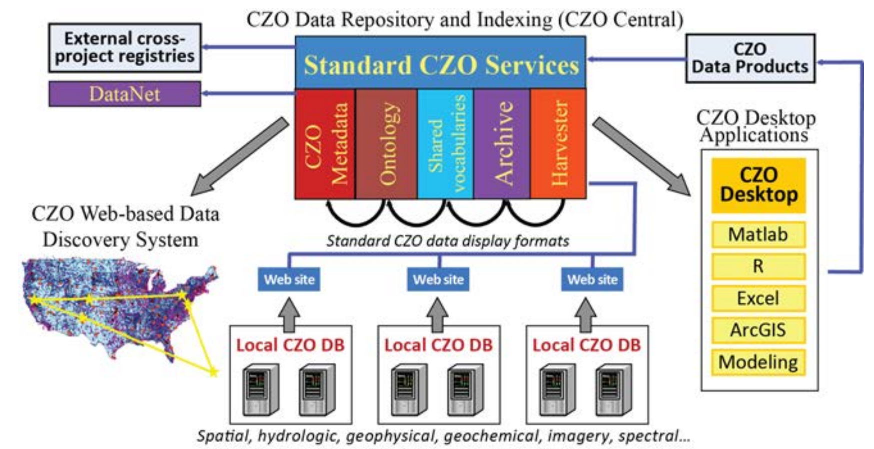

Figure 3. Main components of the CZO data publication workflow

The current prototype uses a group of ASCII files that follow a Display File format created to provide a simple way for CZO site data managers to publish their data for public access. The display files are designed to be both human-readable and serve as a computer-readable transfer format. Once published at an individual CZO web site, these files are automatically harvested into the CZO Central Data Repository housed at the San Diego Supercomputer Center (SDSC). The harvested data are then validated against shared vocabularies and a variable ontology, and archived in a set of databases based on the CUAHSI HIS Observations Data Model (ODM) [Horsburgh et al. 2008] established for each CZO. Upon harvest, validation, and entry into the data portal at SDSC, standard CZO data services are automatically updated to include the new data. For hydrologic observations, these are CUAHSI WaterOneFlow web services, which transmit data according to the WaterML 1.x specification [Zaslavsky et al. 2007], and Web Feature Services (WFS) specified by the OGC [Botts et al. 2008], which are used to exchange time series metadata. The services for each CZO are registered and indexed in the CZO Central's service registry, and can be discovered via a CZO Data Portal, which is compliant with OGC's Catalog Services for the Web (CSW) standard [Nebert et al. 2007]. Data managers use the CZO Central's administration interface to manage web services for their individual CZOs, define service metadata, trigger data harvesting and verify that the harvesting has been successful. The standard services can be consumed by various client applications and registered in cross-project domain registries including CUAHSI HIS Central (for hydrologic time series). As the Display Files are harvested from CZO sites, their metadata sections are also automatically indexed in the data portal, to enable metadata search and discovery of the published display files. While still at an early development stage, the CZO Central discovery system currently indexes about 90 resources, including services and display files, while brokering web service access to over 15 million hydrologic observations collected at CZOs.

For hydrologic observations, the current display file format is based on the information model adopted from CUAHSI ODM. The display files include a configuration file (specifying which files shall be regularly harvested from a CZO web site), sites file, methods file, series metadata file, and a data file. In a typical scenario, each configuration file housed by a CZO will point to single sites and methods files, and to one or more series metadata files, each of which would reference one or more data files. The shared vocabularies of the prototype are adapted from the CUAHSI HIS ODM controlled vocabulary submission system [Horsburgh et al. 2009] but extensible to other types of data collected by CZO sites. These vocabularies, available via a web interface and via web services, are used to establish semantic conventions within the CZO system, ensuring that terms describing key metadata elements are well defined, unique and unambiguous, which, in turn, supports cross-CZO attribute-based data discovery. The web interface for the shared vocabulary system allows data managers to browse the vocabulary content, and propose additions and edits, while the web service API is used by the CZO Central's harvesting application to validate submitted metadata for compliance with the vocabularies.

In addition to the centralized data harvesting, archiving and discovery system prototype, from our initial prototype funding, each CZO has an information portal that presents site information (research design, personnel, research highlights, site characteristics, etc.) and data that is accessible to the user community. The individual web sites at each CZO present a similar look, feel, and navigation system to provide a consistent program presence. Each individual site can be accessed through the first-generation national site at http://criticalzone.org/. The CZO websites reflect the program's focus on interdisciplinary science, integration, linkages, and feedbacks. They highlight the combined field and theoretical approaches and emphasize easy data access. The websites are thus valuable resources for education and outreach, and provide a venue for broader project impacts.

For an operational cross-site and cross-domain integrated data management system, the current prototype must be extended beyond its current limited focus and incorporate data management components for multiple types of CZO data developed in related projects. The projects and software systems to be integrated in the current phase of CZOData development are briefly described below.

CZchemDB. Geochemcial data pose a unique challenge for CZOData. CZchemDB is a relational database for CZO rock and regolith geochemical data that was developed by CZO PI S. Brantley and her group at Pennsylvania State University in collaboration with K. Lehnert from EarthChem [Niu et al. 2011]. CZchemDB's structure is based on the geochemical database model developed for igneous rock geochemical databases [Lehnert et al. 2000] and that have been in use for the PetDB, GEOROC, NAVDAT, and other geochemical databases for over a decade. By adopting this database structure, CZchemDB data and metadata can be seamlessly encoded in EarthChemXML to be integrated into EarthChem's existing, web-based, global portal for geochemical data, which will broaden the accessibility and usability of CZO geochemical data. CZchemDB contains information about the analyzed samples, the sampled sites, the observed values, and extensive metadata that describe the analytical methods, data quality, data source and author [Niu et al. 2011], as well as controlled vocabularies for variables and units.

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

OpenTopography. OpenTopography supports the processing, management, visualization and downloading of high-resolution topographic data. The NSF-funded OpenTopography facility is based at SDSC at the University of California, San Diego and is operated in collaboration with colleagues in the School of Earth and Space Exploration at Arizona State University. Several methods for interacting with the data are used, including Google Earth to visualize LiDAR-derived elevation grids. The data can be downloaded in a range of "pre-processed" standard digital elevation model (DEM) file formats (e.g., ESRI Arc Binary). Also available is the option to generate custom derivative products, such as DEMs produced with user-defined resolution and algorithm parameters, and downloaded in a number of different file formats. The CZO LiDAR data will be managed by the OpenTopography facility and will also be discoverable through the Central CZO data discovery system. OpenTopography is currently working with the CZO program to establish a memorandum of understanding on hosting of CZO LiDAR data and is ingesting a sample dataset from the Christina River Basin CZO as a pilot.

SESAR/ISGN. CZO investigators have agreed to implement the registration of CZO samples and appropriate other objects such as sites with the International Geo Sample Number (IGSN) so that all samples collected as part of the CZO program can be uniquely identified. Unique identifiers for samples are essential for proper referencing of data to the samples on which they were acquired, to link and integrate multiple datasets generated in different labs over time on the same sample, and to track relationships between related samples (multi-tiered hierarchies of samples and sub-samples (e.g., sub-samples taken from a core, or mineral concentrates separated from a soil). The IGSN was developed by the GfG group, which operates the System for Earth Sample Registration (SESAR) as the registry of the IGSN and metadata catalog for Earth samples [Lehnert et al. 2005]. An international SESAR workshop in February 2011 that was attended by CZOData investigators Zaslavsky and Whitenack recommended a new architecture for the IGSN to become a more modular and scalable approach that could advance broad disciplinary and international implementation of the IGSN. In the new architecture the IGSN, registry is separated from a central Sample Metadata Clearinghouse (SESAR) and 'Local Registration Agents' are introduced that provide registration services to specific user communities with tools for metadata submission, management, and archiving. The new IGSN architecture will be established over the next year with funding from NSF's Office of International Science and Engineering. This new architecture will provide the ability to incorporate and "brand" samples from CZO sites and other earth observatories.

CUAHSI HIS. With NSF support over the last 8 years, CUAHSI has developed its HIS for sharing hydrologic data. While initially developed for hydrologic data, the system design is sufficiently general to support most point-based time series. It follows Service Oriented Architecture (SOA) principles and consists of three major types of components, linked together by web services. The first of the components is HydroServer, a server for storing and publishing data and its descriptive metadata. The second is HydroCatalog, an index of all registered HydroServers and a subset of their metadata to facilitate system-wide data searching. The third component is HydroDesktop, a desktop application for searching, accessing, and analyzing HIS data. The system currently provides web service access to over 70 government and academic water observation networks, presenting over 5.2 billion data points for 1.96 million measurement sites in the US and over 18,000 parameters of water quality and quantity, and brokering over 7,000 service requests for data download per day. CZOData will build upon the web service interfaces, hydrologic time series encoding (e.g., WaterML), and metadata cataloging and search capabilities developed by CUAHSI HIS.

EarthChem. EarthChem is a community-driven effort to facilitate the discovery, access, and preservation of geochemical data. Since its inception in 2005, EarthChem has established itself as a leading community resource that offers to its users a range of data services comprising: the EarthChem Portal as a central access point to geochemical data in federated databases with tools for data discovery, data access, and data analysis; the Geochemical Resource Library that serves as a repository for user-submitted data and features an online submission tool and a searchable metadata catalog; topical data collections such as the Deep Lithosphere Dataset that are developed and maintained by EarthChem; and GEOCHRON, a system for easy population and access of geochronology and thermochronology data [Lehnert and Vinayagamoorthy 2007].

One of EarthChem's main achievements has been the community-driven development of metadata requirements for geochemical data through workshops, the Editors Roundtable, and in collaboration with other geochemical data systems [Lehnert et al. 2009]. EarthChem developed templates to guide investigators with formatting their data and providing the required metadata,for submission to EarthChem's geochemical data collections. EarthChem will generate modified templates that fit the needs of CZO geochemical data and that will ensure CZO geochemical data to be adequately documented. From these data templates the data can be seamlessly imported into the EarthChem portal for discovery by a wide community of scientists.

NANOOS Visualization System. The NANOOS Visualization System [NVS 2010] was developed by several partners of the Northwest Association of Networked Ocean Observing Systems (NANOOS), led by the Applied Physics Laboratory (APL) of the University of Washington. The system has received considerable attention within the U.S. Integrated Ocean Observing System (IOOS) community for its user-friendly yet rich interface, its rapid development pace, and its consistent data presentation that is independent of the original data source [Mayorga et al. 2010]. NVS can currently access hydrologic time series data via CUAHSI water data services used in the CZO prototype, as well as time series data from other services. This tool will be adapted to CZO needs by providing user-friendly data access to CZO time series data and visualizations for time series and relevant, core geospatial data from all CZO sites, as provided by CZOData services hosted at the CZO Central Data Management System. It will be accessible through the CZO main web portal. The APL team is currently finishing the development of an NVS adaptation to the Christina River Basin CZO focused on near-real-time time series data, together with background GIS layers and relevant Citizen Science information [NVS 2011].

LTER. The LTER Network is a collaborative effort involving more than 1800 scientists and students investigating ecological processes over long temporal and broad spatial scales. The Network promotes synthesis and comparative research across sites and ecosystems and among other related national and international research programs. The National Science Foundation established the LTER program in 1980 to support research on long-term ecological phenomena in the United States. The 26 LTER sites represent diverse ecosystems and research emphases. The LTER program has long been the gold-standard for data accessibility and integration generated by environmental observatories [e.g., Ingersoll et al. 1997]. We will build on the metadata standards and web presentation for site data developed by the LTER program, and incorporate LTER EML standards with our ecological data. Additionally, we will adopt and build on LTER software tools for data managers, as shown in Task 3 below.

DataONE. The Data Observation Network for Earth (DataONE) is part of NSF's ongoing DataNet (Sustainable Digital Data Preservation and Access Network Partners) program. DataONE is poised to be the foundation of new innovative environmental science through a distributed framework and sustainable cyberinfrastructure that meets the needs of science and society for open, persistent, robust, and secure access to well-described and easily discovered Earth observational data. DataONE will ensure the preservation and access to multi-scale, multi-discipline, and multi-national science data and is developing a SOA that consists of a distributed set of "Member Nodes" on which scientific data are published and Coordinating Nodes that provide data discovery and other services. Member Nodes can join DataONE by implementing the necessary web service interfaces and sharing data and metadata in compatible formats. DataONE then indexes the available data and provides discovery and other services through its Coordinating Nodes. Integration with DataONE will bring to CZOData a robust mechanism for replicating and archiving CZO data, a library of tools for data integration, analysis, and synthesis, as well as increasing the exposure of CZO data to broader scientific communities.

3.2 CZOData Design

The CZOData design is based on the experience of the project team in building the CZO data management prototype and in a range of related CI projects. The design follows the SOA model, and is composed of components that link a set of web services and web sites that pull together numerous NSF-funded systems described above and shown in Figure 4. Local CZO web sites host CZO Display Files and tools that support generating such files from CZOs local databases and other records. These Display Files will then be harvested by the CZO Central Data Management System which, as in the prototype, will review the files and then process them according to the workflows specific for each type of data. Unlike the prototype, which trained on hydrologic time series, many more data types will be supported. The workflows will check Display Files against terms in the shared vocabulary and then add metadata from the Display File to the CZO Metadata Catalog that is associated with the CZO Data Discovery Portal, before forwarding the harvested data to the sub-system matching its type.

This design will have two central portals: the CZO Central System for point-based time series and the CZchem System for geochemical data. These two portals will store, index, analyze, and publish the data via web services using WaterML and EarthChemXML encodings, respectively. These web services allow the CZO Central portals to send data and metadata to CUAHSI HIS and EarthChem, allowing CZO users to take full advantage of those systems. LiDAR data will be separately sent to OpenTopography, and their metadata will be also added to the CZO Central Metadata Catalog directly from the OpenTopography registry. Other spatial datasets maintained at local CZO sites (e.g., shapefiles of local hydrography, land use, vegetation) will have defined metadata, will be registered at the CZO Central Metadata Catalog, and will be exposed via the CZO Central Data Discovery Portal, which, as a custom application of ESRI's GeoPortal, provides sufficient support for such data. CZO Central will also implement a DataONE interface, allowing CZOData to function as a DataONE Member Node and linking it into the DataONE network. Several of the components of CZO Central also produce OGC compliant WFS and CSW web services, allowing standards-based clients of those services access to the CZO data.

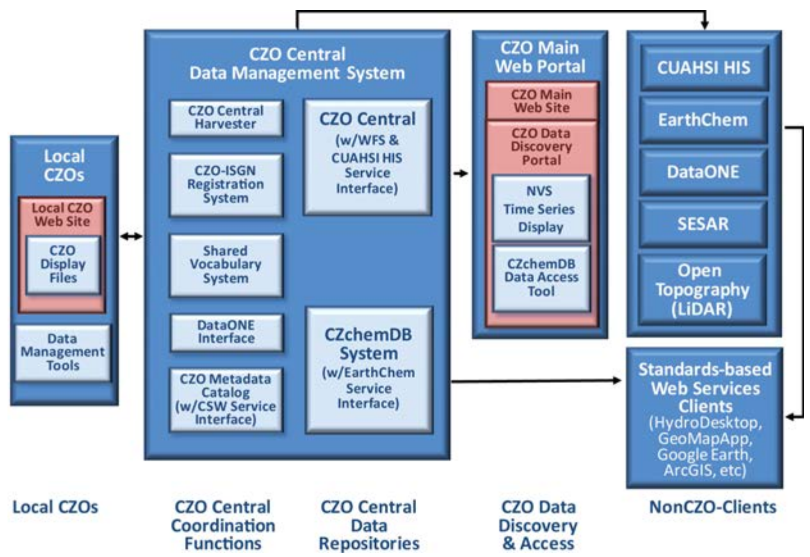

Figure 4. CZOData design architecture diagram.

CZOData will take advantage of the unique sample ID system offered by SESAR. A CZO-ISGN Registration System will be set up, allowing the Central CZO system to request batches of unique IDs, and then issue IDs to investigators at local CZOs, as needed. ISGN assignment tools at the local CZOs and at CZO Central will then collect the metadata for the samples and forward it back to SESAR.

CZOData will rely on standards-based formats and services. Data and metadata harvested from individual CZO web sites into the central CZO system will be republished as standards-compliant web services. This will enable programmatic access to CZO data from multiple client applications, including discovery portals, visualization systems, and data analysis and modeling applications. Compliance with international standards for data exchange is beneficial because it will open CZO data for access by a wide range of third party standards compliant applications, and will make it easier to integrate CZO data with data collected in related earth science projects. As mentioned above, we will rely on the WaterML 1.x specification to publish hydrologic time series information and will use encodings and service interfaces for spatial data standardized by OGC. WaterML 1.x has been a successful specification, adopted by USGS, NCDC, and multiple state-wide, local and academic projects. WaterML 1.x became one of the foundations for a new specification, WaterML 2.0, which is being developed by the OGC/WMO Hydrology Domain Working Group, with the expected adoption by OGC in late 2011. Once the specification is adopted as an international standard, the CZO Central system will additionally publish web services following this new encoding and using OGC Sensor Observation Service (SOS) as the standard service interface. We will communicate with the OGC Hydrology and Sensor Web Enablement Working Groups to provide feedback on the implementation of the standards in the context of the CZO project and request specification changes as needed.

3.3. Project Team

The proposed project will be led by an executive management team consisting of M. Williams from the Boulder Creek CZO who will chair the team (BC-CZO), A. Aufdenkampe from the Christina River CZO who will be the vice-chair (CR-CZO), K. Lehnert from the Geoinformatics for Geochemistry (GfG) program at the Lamont-Doherty Earth Observatory (LDEO) of Columbia University, I. Zaslavsky of the Spatial Information Systems Lab at SDSC, E. Mayorga of the Applied Physics Laboratory (APL) at the University of Washington, and J. Horsburgh of The Utah Center for Hydrologic Information and Computing (UCHIC) at Utah State University. More detailed information on additional team members and responsibilities can be found in the budget justifications and scope of work. Site data managers and scientists will be an integral part of this team, as documented below.

4. COMMUNITY INVOLVEMENT

The key measure of success for CZOData is the acceptance and use of its services and systems by the CZO investigators and the broader CZ community. A critical component of the proposed project will therefore be to involve this community in the development process. 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. We plan to achieve this via the following activities:

Instigate and Support an Information Management Committee (IMC). One of the keys to accomplishing community involvement is development of an IMC made up of the executive management team, the SIMs, and a science representative from each CZO site appointed by the site PI. The IMC will hold meetings and workshops to develop CZO program standard procedures and protocols, to exchange approaches, best practices and technical information, and to ensure regular communication among its participants. The IMC will have a monthly telecom and will meet in person once a year. 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. Participation in this committee will be time-consuming, so a stipend has been budgeted for each of the participating scientists.

Cyber-security and Data Management Web-based Information Events. Security and management of individual CZO websites and systems will remain the responsibility of the individual sites. However, the IMC, 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.

Synthesis Working Groups. At the direction of the CZO PIs, a synthesis working group will meet annually to discuss a topic of interest that is cross-CZO. The goal of the workshop is a manucript 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. This feedback each year will provide important mid-course corrections for CZOData. It will also marry data management to CZO science.

Subdiscipline Workshops. IMC participants will convene additional workshops on subdisciplines to enhance skill sharing and communication among data managers and scientists. For example, funding in this proposal to LDEO will support a workshop for CZO geochemists to refine procedures and tools for the management and access of geochemical data, including metadata requirements, templates for data submission, IGSN registration procedures, and user interface of CZchemDB. Similarly, there is funding for a skill-sharing workshop on GIS and visualization tools for SIMs. An example is a visualization tool for geophysical data under development by the BC-CZO (http://czo.colorado.edu/html/data-befuss3_3d.shtml).

5. PROJECT PLAN

The research proposed here is a unique integration of CI, community processes, and software engineering to create a system that will support multidisciplinary and cross CZO research. Responsible organization for each of the tasks listed below is given in parenthesis.

5.1 Project Tasks

Task 1. CZO-wide Information Portal (BC-CZO, CR-CZO, SDSC). The CZO web sites (individual sites and national site) are the major communication platform between the CZO program and the user community, including non-CZO scientists and students, the general public, and K-12 students. The individual web sites are the essential backbone for getting data from sites to central portals. Through existing funding, we are developing a new, fully-integrated web system for promoting and displaying all CZO content, under the direction of David Lubinski of the BC-CZO. This system will be deployed at SDSC and is scheduled to roll out in September 2011. Wireframes of mockups for the revised system can be accessed at: http://instaar.colorado.edu/temp/czo_mockups_v3/National_Home.html. Through this proposal we seek support to keep the information portal running and also to support updates to the information portal.

Maintenance and Support. We want to ensure an easy-to use experience for everyone involved, including both the people visiting the web sites as well as the CZO personnel updating site content. Successfully reaching this goal requires a number of basic tasks such as regular software updating for both the core and third party aspects of the content management system. Such updates are necessary to ensure security, speed, and reliability. They also keep more options open for the future. Additional tasks include bug fixes and assistance with content entry.

Design and Development. 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.

Task 2. Consistent metadata (All). 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. While our prototype was focused on hydrologic time series, the Display File format was designed to be extensible to other types of data. To illustrate, 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.

Task 3. Data and metadata publication tools and templates (All). Our interactions with SIMs show that 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. Similar templates and software programs are being developed for other tools, such as geophysical data.

Task 4. Develop a web-based user interface to CZchemDB (GfG). Currently, CZchemDB is not web-enabled and can only be accessed locally at PSU. Building on Task 3 above, 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 GfG databases such as EarthChem, PetDB, SedDB, and VentDB 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.

Task 5. Shared vocabulary system (UCHIC, SDSC). 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.

Task 6. Unique sample identifier for CZO data (GfG). 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. For example, the CZO geochemistry group is discussing making IGSN registration a mandatory component of their system.

Task 7. Support for site data managers (SDSC, UCHIC, BC-CZO). Identifying, analyzing and prioritizing information management issues experienced by CZO sites, developing recommendations for shared technological approaches, and getting constant feedback from SIMS and users about system design and operation are important conditions of system quality and relevance and its eventual acceptance by our community. The IMC is an essential component of addressing these needs. Additionally, we will support SIMS as follows: 1) biweekly teleconferences with SIMs under the supervision of SDSC; 2) a document sharing web site, initially developed with the prototype by SDSC, will continue to be maintained and managed as a communication tool for the SIMs and CZOData team; 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 by the team of M. Williams of Boulder, T. Whitenack of SDSC, and Kim Schreuders of USU. 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.

Task 8. CZO Central metadata catalog, data discovery portal and harvester (SDSC). 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, invoking the viewer directly from the portal (described in Tasks 4 and 12); and 3) implementation of 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. For example, when Display Files describing geochemical data are published by SIMs, the central harvester will send a request to CZChemDB to initiate either automatic or manual data ingestion into the database.

Task 9. Central data repository of time series and point data (SDSC). 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. In our experience, data published via CZO web sites may change. Variables and measurement points may be renamed, and metadata or data values updated after additional calibration. To ensure that the central time series catalog of hydrologic measurements remains consistent with the data files and that integrity and accessibility of the data are maintained, the Display Files will be periodically re-harvested, and, if differences in previously recorded data and metadata are detected, new versions of the datasets will be made available via web services. At the same time, the previous versions will continue to be retained, so that a researcher could reconstruct the state of the data archive for any date if interest. This is useful in situations where analysis needs to be replicated and validated against information that was available in the CZO system for a specific date.

Another key component of this task is working with SIMs to publish additional (and eventually, all) time series information collected by CZOs or important for CZO research, such as ecological data. We will ensure that the configured display files are successfully harvested into the central system, and troubleshoot harvesting problems when required. Maintaining a smooth, intuitive and efficient data publication workflow for time series is critical for the system to be accepted by the CZO community.

Task 10. Publication and sharing of spatial data and LiDAR (SDSC). Besides geochemical and hydrologic data, this phase of the CZOData project will provide integrated access to CZO spatial data and LiDAR products. As described in Section 3.1, the OpenTopography project has built a scalable system for processing and analysis of large volumes of high-resolution (meter to sub-meter) topography data acquired with LiDAR and other technologies, accessible via web services. 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 incorporating elements of the OpenTopography portal (visualization, in particular) in the CZO portal. This work will be done collaboratively with the OpenTopography team at SDSC.

Centralized publishing and archiving of spatial information, in particular vector data collected or organized by CZOs, as well as remote sensing data products, are common data management needs expressed by SIMs and PIs. We will work with SIMs to replicate and archive such data at SDSC, and register them through the CZO Central data portal, taking advantage of the portal's ability to manage common spatial data formats and spatial metadata profiles.

Task 11. Integrate CZchemDB data into the EarthChem system (GfG). 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. Similar web services have already been developed for the PetDB and SedDB data collections, but will need to be modified as EarthChem is going to rebuild the EarthChem Portal site using a SOA to provide a web service aspect to all search and output functions, responding to the increasing demands for interoperability. The new web service based architecture will offer a superior range of options and possibilities for interaction and data exchange with other developers and systems. The conversion of CZchemDB data into EarthChemXML documents requires that vocabularies used in CZchemDB are aligned with those in EarthChemXML, for example lithologies, geographical names, journal names, etc. We will update EarthChem vocabularies to include terminology used by CZchemDB, and ensure that CZchemDB uses existing vocabulary in EarthChem if applicable. We will also ensure coordination with the CZO shared vocabulary system that we are developing.

Task 12. Visualization tools (APL). 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, working in close collaboration with the SDSC and GfG teams. 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.

5.2. Metrics Of Success And End-User Involvement

CZOData will use a range of metrics to measure its success, evaluating both the internal process used to create it, as well as the adoption and successful use by the community it serves. Selection and validation of metrics will be an iterative process, subject to community input and prioritization: 1) *Growth of data holdings*, e.g., number of data sets posted at individual sites, including data values, sites, variables, and web service data; types and amounts of data stored within the data portals; 2) *Use statistics*, e.g., the number of times data sets at individual sites are accessed; number of unique users per time unit; average log-on duration and total use time; geographic diversity of system users (based on IP address), etc; 3) *Scientific impact*, e.g., number of citations using CZOData; number of papers submitted to and published in peer-reviewed forums about this project and its implementation and their citations; posters, invited talks, panel sessions; 4) *User satisfaction*, e.g., user surveys to gather feedback on data content, data delivery, data quality, functionality of applications for search, integration, visualization, and analysis of data; adoption of standard specifications for data format, metadata content, quality assurance/quality control (QA/QC), and exchange protocols across all sites and data portals; 5) *Impact on Geoscience cyberinfrastructure*, e.g., number of client systems using CZOData web services; adoption of CZOData developments (e.g., metadata schemas, architecture) by other systems and projects.

6. EDUCATION AND OUTREACH

The CZO's existing development, outreach, and training activities raise the profile of the CZO program; benefit the program by creating collaborative research, cyberinfrastructure development, and education opportunities; and transfer requisite technological skills and knowledge to CZO scientists and information managers. Outreach supports information flows within CZO and between CZO and the broader community of scientists, educators and the public. Proposed development, outreach, and training activities aim to: 1) enhance understanding of CZO—its structure (e.g., sites), capabilities, achievements and contributions to science and society; 2) promote CZO as a center of excellence offering that support understanding long-term critical zone research and processes; 3) facilitate and contribute to partnerships between CZO and other networks, mission agencies, and organizations that lead to new mutually beneficial research, cyberinfrastructure, and educational opportunities; and 4) provide CZO scientists, SIMs, and students with training in new technologies, standard methods and approaches that facilitate data acquisition, integration and synthesis, as well as communication of knowledge to the broader community of scientists, students, and the public.

Proposed development and outreach will focus on activities that most successfully communicate CZO achievements, demonstrate broader scientific impacts, create significant opportunities for collaboration and building partnerships, and provide training that is critical for CZO in meeting its objectives. Education goals of the project include: 1) serve CZO graduate and undergraduate students and involve them in data management and metadata description; 2) cultivate a data sharing and collaborative culture; 3) facilitate access to and interpretion of CZO data from neighboring earth science domains; and 4) create a new generation of earth science researchers that combine understanding of physical processes in the critical zone with the ability to contribute to and use advance data management and data integration systems.

7. INTELLECTUAL MERIT

Here we have a unique opportunity to integrate environmental research with data management and CI from the ground up. A major contribution of this project is the integration of site data managers with critical zone researchers. As a result of this project, an interactive, distributed data publication and sharing system for diverse data will be developed and made operational for the six participating CZO sites. Once operational, it will enable researchers and students to quickly discover and access a variety of CZO data, including time series of water and atmosphere measurements, observations derived from geochemical samples, high density LiDAR and other vector and raster spatial data. To accomplish this, the project will develop data publication guidelines and protocols and implement them via data and metadata submission forms, standard data and metadata encodings and service interfaces, and a uniform, cross-CZO mechanism for generating Display Files at individual sites, harvesting them, validating and archiving in the central system, and thus making diverse data accessible via web services. Such consistent standards-based publication of data services from different subdomains of CZO, and making them available through an integrated discovery and access system, will create a comprehensive portrait of the study regions and experimental watersheds that will enhance research and education in the earth sciences. CZOData will focus on specific needs expressed by the CZO research community, in particular, by creating and maintaining a new data publication model that simplifies data management requirements for each site while creating a comprehensive operational system for publishing, discovering, accessing and integrating research data across CZOs and scientific domains.

8. BROADER IMPACTS

CZOData will significantly ease data discovery and integration for CZO researchers and collaborators, thus a) supporting new tasks that were not possible before, such as cross-CZO comparison and modeling, and b) easing the data management burden on individual researchers and data managers and thus letting them focus on science issues. The data system will be used by graduate and undergraduate students across CZO and will make it easier for students and researchers to access data collected in other subdomains of CZO, thus enabling more comprehensive research designs. By incorporating system research and development with graduate education, we will move towards cultivating a new generation of researchers skilled in both different facets of environmental research and in advanced information management and computing. The collaboration between CUAHSI HIS, IEDA and CZO data management teams on this proposal, has, during proposal development, and will in the ongoing project, allow us to identify common elements in the data models used by the different disciplines involved and work towards a generalized data model that brings together and serves the broader interests of the multiple disciplines involved. Exposing CZO data as services compliant with OGC specifications will enable a higher level of interoperability with data and software developed by other groups and will lead to a sustainable operational model with linkages to IEDA, EarthChem, DataONE, OpenTopography, LTER, NEON, and other environmental observatories and systems.

9. RESULTS FROM PRIOR NSF SUPPORT

S. Anderson: **EAR-0724960, 2009 – 2011 Supplement to the Boulder Creek Critical Zone Observatory: Weathered Profile Development in a Rock Environment and Its Influence on Watershed Hydrology and Biogeochemistry ($330,000)**. ($85K to SS-CZO and $85K to SH-CZO directly). A $500,000 supplement was received by CZO in October 2009 to develop a prototype for CZOData under the direction of M. Williams, with subcontracts to SDSC to develop a central data harvesting and services system for hydrologic data, and USU to develop web-based tools and a shared vocabulary. To date, this supplement has resulted in 1) an information portal at each CZO that has a common look, feel, and navigation scheme to provide a program identity and brand (http://criticalzone.org/); 2) a prototype of a data catalog and display that are consistent across the CZO program; 3) single portal access for a subset of CZO (water quantity and quality, climate data) that harvests these data from each site and supports geographic based data inquiries, sophisticated data displays (http://maxim.ucsd.edu/czoportal/), and incorporation of CZO data directly into other data systems (e.g., CUASHI HIS); 4) development of CZchemDB; 5) acquisition of LiDAR data for all six CZO sites and an MOU with OpenTopography to display the data; and 6) a prototype shared vocabulary (http://sv.criticalzone.org/). There have been numerous presentations about CZOData at national and international meetings (e.g., Whitenack et al., 2010), invited talks (e.g., Williams, 2011), and one peer-reviewed paper (Zaslavsky et al., 2011).

References

Anderson, S.A., R. C. Bales, and C. J. Duffy. Critical Zone Observatories: Building a network to advance interdisciplinary study of Earth surface processes. Mineralogical Magazine, 72(1), pp 7-10.CZO Community Report to New Research Opportunities in the Earth Sciences, 2010.

Botts, M. G. Percivall, C. Reed, and J. Davidson, OGC® Sensor Web Enablement: Overview and High Level Architecture, GeoSensor Networks, Lecture Notes in Computer Science, 2008, Volume 4540/2008, 175-190, DOI: 10.1007/978-3-540-79996-2_10

Critical Zone Exploration Network, http://www.czen.org/ (last accessed June 1, 2011).

Cyberinfrastructure for Environmental Observation Networks (CEON) Workshop Report, held February 25 & 26, 2008 at The National Science Foundation, Arlington, VA. http://feon.wdfiles.com/local--files/start/Feb2008WorkshopFinalReport.pdf (last accessed June 1, 2011).

Dozier, J. and W. B. Gail, (2009), "The Emerging Science of Environmental Applications," in "The Fourth Paradigm", Ed. by T. Hey, S. Tansley and K. Tolle, Microsoft Research, Redmond, WA, pp. 13-29.

DataONE Project Team, (2011), DataONE structure and potential partnership as a Member Node, http://www.dataone.org/content/dataone-structure-and-potential-partnership-member-node, accessed 3/13/11.

The EarthChem Portal, www.earthchem.org (last accessed June 1, 2011)

European Petroleum Survey Group, Geodetic Parameter Dataset, http://www.epsg-registry.org/ (last accessed June 1, 2011)

ESRI Geoportal Server, http://geoportal.sourceforge.net/ (last accessed June 1, 2011)

Ingersoll, R.C.; Seastedt, T.R.; Hartman, M.A. 1997. A model information management system for ecological research. BioScience 47(5):310-316.

Horsburgh, J. S., D. G. Tarboton, D. R. Maidment and I. Zaslavsky, (2008), "A Relational Model for Environmental and Water Resources Data," Water Resour. Res., 44: W05406.

Horsburgh, J. S., D. G. Tarboton, M. Piasecki, D. R. Maidment, I. Zaslavsky, D. Valentine and T.Whitenack, (2009), "An integrated system for publishing environmental observations data," Environmental Modelling & Software, 24(8): 879-888.

Horsburgh, J. S., D. G. Tarboton, D. R. Maidment and I. Zaslavsky, (2011), "Components of an environmental observatory information system," Computers & Geosciences, 37(2): 207-218.

Lehnert, K., Su, Y, Langmuir, C.H., Sarbas, B., Nohl, U., 2000. A global geochemical database structure for rocks. Geochem. Geophys. Geosyst. vol. 1. Paper Number 1999GC000026.

Lehnert, K. A.; Goldstein, S. L.; Lenhardt, W. C.; Vinayagamoorthy, S. "Unique Sample Identification in the Solid Earth Sciences: SESAR (Solid Earth Sample Registry) and the International Geo Sample Number IGSN", Geophysical Research Abstracts, Vol. 7, 05498, 2005

Lehnert, K., and S. Vinayagamoorthy, (2007) "Geoinformatics for Geochemistry (GFG): Integrated Digital Data Collections for the Earth and Ocean Sciences", Proceedings of Geoinformatics 2007 - Data to Knowledge. USGS Scientific Investigation Report 2007-5199. U.S. Geological Survey, Reston, VA, pp. 32-34.

Lehnert, K. A.; Walker, D.; Block, K. A.; Ash, J. M.; Chan, C., "EarthChem: Next Developments to Meet New Demands", American Geophysical Union, Fall Meeting 2009, abstract #V12C-01

Lehning, M., N. Dawes, M. Bavay, M. Parlange, S. Nath, and F. Zhao, (2009), “Instrumenting the Earth: Next-Generation Sensor Networks and Environmental Science”, in “The Fourth Paradigm”, Ed. by T. Hey, S. Tansley and K. Tolle, Microsoft Research, Redmond, WA, pp. 45-51.

The US Long Term Ecological Research Network, www.lternet.edu (last accessed June 1, 2011).

Mayorga, E., T. Tanner, R. Blair, A.V. Jaramillo, N. Lederer, C.M. Risien and C. Seaton. 2010. The NANOOS Visualization System (NVS): Lessons learned in data aggregation, management and reuse, for a user application. Proc. MTS/IEEE Oceans'10.

Niu, X., K. A. Lehnert, J. Williams, and S. L. Brantley S. (2011) “CZChemDB and EarthChem: Advancing management and access of critical zone geochemical data”, Applied Geochemistry 26, S108–S11

National Critical Zone Observatory program (CZO), www.criticalzone.org (last accessed June 1, 2011).

National Science Foundation, 2007. Cyberinfrastructure Vision for 21st Century Discovery (http://www.nsf.gov/pubs/2007/nsf0728/index.jsp?org=NSF)

National Science Foundation, 2011. Dear Colleague Letter: The "Earth Cube" - Towards a National Data Infrastructure for Earth System Science (http://www.nsf.gov/pubs/2011/nsf11065/nsf11065.jsp?org=NSF)

Nebert, D., A. Whiteside, P. Vretanos, eds. (2007), OpenGIS Catalog Services Specification, OGC 07-006r1, Version 2.0.2, http://portal.opengeospatial.org/files/?artifact_id=20555.

NVS 2010. NANOOS Visualization System. Viewed 2011 August 1. http://nvs.nanoos.org

NVS 2011. Christina River Basins Visualization System. Viewed 2011 August 1. http://www.nanoos.org/nvs_crb/nvs.php [Interim URL, available through September 2011].

Open Geospatial Consortium – Hydrology Domain Working Group, http://external.opengis.org/twiki_public/bin/view/HydrologyDWG/ (last accessed June 1, 2011).

Observations and Measurements – XML Implementation. OGC Document 10-025r1, http://portal.opengeospatial.org/files/41510 (last accessed June 1, 2011).

Tarboton, D. G., J. S. Horsburgh, D. R. Maidment, T. Whiteaker, I. Zaslavsky, M. Piasecki, J. Goodall, D. Valentine and T. Whitenack, (2009), "Development of a Community Hydrologic Information System," 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Ed. R. S. Anderssen, R. D. Braddock and L. T. H. Newham, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, July 2009, p.988-994.

Tarboton, D. G., D. R. Maidment, I. Zaslavsky, D. P. Ames, J. Goodall, and J. S. Horsburgh (2010), CUAHSI Hydrologic Information System 2010 Status Report, Consortium of Universities for the Advancement of Hydrologic Science, Inc, 34p. http://his.cuahsi.org/documents/CUAHSIHIS2010StatusReport.pdf (last accessed June 1, 2011).

Whitenack, T., M. W. Williams, D. G. Tarboton, I. Zaslavsky, M. Durcik, R. G. Lucas, C. Dow, X. Meng, B. Bills, M. Leon, C. Yang, M. Arnold, A. K. Aufdenkampe, K. Schreuders, and O. Alvarez (2010): Development of an integrated information system for Critical Zone Observatory data, Eos Trans. AGU, Fall Meet. Suppl., Abstract IN31B-1289.

Williams M,, Developing an integrated data management system for the Critical Zone Observatory program, CUAHSI Conference on Hydrologic Data and Information Systems, Utah State Univesity, June 22□24, 2011.

Zaslavsky, I., D. Valentine and T. Whiteaker, (2007), "CUAHSI WaterML," OGC 07-041r1, Open Geospatial Consortium Discussion Paper, http://portal.opengeospatial.org/files/?artifact_id=21743 (last accessed June 1, 2011).

Zaslavsky, I., and D. R. Maidment (2011), Service orientation in the design of a community hydrologic information system, in Geoinformatics: Cyberinfrastructure for the Solid Earth Sciences, edited by G. R. Keller and C. Baru, pp. 193-209, Cambridge University Press.

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", Conference on Environmental Information Management, EIM'2011 (in press) .

MARK W. WILLIAMS

Institute of Arctic and Alpine Research and Department of Geography, 303.492.8830 (tel)
University of Colorado at Boulder, CB 450 303.492.6388 (fax)
Boulder, CO 80309 markw@snobear.colorado.edu

EDUCATION
All degrees in Biological Sciences, with emphasis on Biogeochemistry and Hydrology
Ph.D. 1991 University of California, Santa Barbara
B.A. 1985 University of California, Santa Barbara

EMPLOYMENT
Present-2005: Professor of Geography and Fellow at the Institute of Arctic and Alpine Research, University of Colorado, Boulder
2005-1998: Associate Professor and Research Associate III, CU-Boulder.
1991-1998: Assistant Professor of Geography and Research Associate, CU-Boulder.

AWARDS
2008: Denali Award for Recent Outstanding Recent Achievement in Mountain Geography, Mountain Geography Specialty Group (MGSG) of the American Association of Geographers
2006: Residence Life Academic Teaching Award, CU-Boulder.
2006: National Park Service, Intermountain West Annual Award for Research to Support Park Service Resource Management by a non-Federal Scientist.
2004: ARCUS visiting speaker, Fairbanks and Juneau, Alaska.
2000-1999 academic year: Faculty Fellowship, CU-Boulder.
1999: Fulbright Research Fellowship, Ecuador.
1998: US EPA Region VIII Outstanding Environmental Achievement Award.
1991-1987: NSF pre-doctoral fellowship for minority students

FIVE RELEVANT PUBLICATIONS
Zaslavsky I, Whitenack T, Williams M, Tarboton D, Schreuders KAT and Aufdenkampe A, Initial Design of Data Sharing Infrastructure for the Critical Zone Observatory, Environmental Information Management, in press.
Leopold M, Williams MW, Caine N, Volkel J, Dethier D; Internal Structure of the Green Lake 5 Rock Glacier, Colorado Front Range, USA; Permafrost and Periglacial Processes, 2011, Published online in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/ppp.706
Williams MW, TA Erickson, and JL Petrzelka, Visualizing meltwater flow through snow at the centimetre-to-metre scale using a snow guillotine, Hydrological Processes, DOI: 10.1002/hyp.7630, 2010, in press
Williams MW, C Seibold, and K Chowanski, Storage and Release of Solutes from a Subalpine Seasonal Snowpack: Soil and Stream Water Response, Niwot Ridge, Colorado, 95(1), pp 77-94, DOI: 10.1007/s10533-009-9288-x, Biogeochemistry, 2009
Nanus, L., M. W. Williams, D. H. Campbell, E. M. Elliott, and C. Kendall, Evaluating Regional Patterns in Nitrate Sources to Watersheds in National Parks of the Rocky Mountains using Nitrate Isotopes, Environmental Science and Technology, 2008, 42, 6487-6493

FIVE ADDITIONAL PUBLICATIONS
Miller MP, BE Simone, DM McKnight, RM Cory, MW Williams and EW Boyer, New light on a dark subject: comment, Aquatic Sciences, DOI: 10.1007/s00027-010-0130-2, 72:269-275, 2010.
Fassnacht SR, MW Williams, MV Corrao, Changes in the surface roughness of snow from millimetre to metre scales, Ecological Complexity 6 (2009) 221-229

Marshall JD, John M. Blair, Deb Peters, Greg Okin, Albert Rango, Mark Williams, Predicting and Understanding Ecosystem Responses to Climate Change at Continental Scales, Frontiers in Ecology and the Environment, 6(5): 273-280, doi:10.1890/070165, 2009.

Litaor MI, M. W. Williams, and T. Seastedt, Topographic controls on snow distribution, soil moisture, and species diversity of herbaceous alpine vegetation, Niwot Ridge, Colorado, JGR: Biogeosciences, 113, G02008, doi:10.1029/2007JG000419, 2008.

Molotch, N. P., P. D. Blanken, M. W. Williams, A. Turnipseed, R. Monson, and S. Margulis, Estimating sublimation of intercepted and sub-canopy snow using eddy covariance systems, Hydrologic Processes, DOI:10.1002/hyp.6719, V21, N12, pp 1597-1607, 2007.

SYNERGISTIC ACTIVITIES
NSF (PI), NWT RIDGE LTER program, 2004-2016
NEON (PI): NWT RIDGE core site for domain 13
NSF (senior personnel); Boulder Creek Critical Zones Observatory
Lead: Integrated data management system for Critical Zones Observatory

COLLABORATORS
(last 48 months) Anthony Aufdenkampe (UD), Roger Bales (UA), Paul Brooks (UA), Nel Caine (CU-B), Don Campbell (USGS), Dave Clow (USGS), Jeff Dozier (UCSB), Tissa Illangasekare (CU-B), John Melack (UCSB), Tom Meixner (UCR), Tad Pfeffer (CU-B), Robert Sanford (DU), Dave Tarboten (USU), Ilay Zaslavsky (UCSD)

Ph.D. Mentor:
John Melack (UC Santa Barbara)

Current graduate students:
Adina Racoviteanu, Rory Cowie (PhD), Tony Krupicka, Daniel Cordalis, Morgan Zeliff, Adrianne Kroespch, Katya Hafich, Alana Wilson

Past graduate students:
Jennifer Petrzelka, Jordan Parman, Rory Cowie (MA), Leora Nanus, Fengjing Liu, Eran Hood, Tyler Erickson, Meredith Knauf, Daniel Cordalis, Craig Anderson, Kim Raby, Paul Brooks, Dave Manthorn, Jen Hazen, Tom Davinroy, Andreas Torrizo, Mindia Brown, George Ingersoll

BIOGRAPHICAL SKETCH
ANTHONY K. AUFDENKAMPE

Address: Stroud Water Research Center, 970 Spencer Road, Avondale, PA 19311
Tel: (610) 268-2153 (x263) FAX: (610) 268-0490, aufdenkampe@stroudcenter.org

(i) Professional Preparation

Dartmouth College	Chemistry	B.A., 1991
University of Washington	Chemical Oceanography	M.S., 1998
University of Washington	Chemical Oceanography	Ph.D., 2002

(ii) Appointments

Associate Research Scientist, 2011-Present. Stroud Water Research Center, PA, USA.

Assistant Research Scientist, 2003-2011. Stroud Water Research Center, PA, USA.

Postdoctoral Fellow, 2002. NSF International Research Fellow. Univ. of South Carolina & Centro de Energia Nuclear na Agricultura (CENA), Brazil.

Research Assistant, 1994-2002. School of Oceanography, University of Washington.

(iii) Publications—Five Most Closely Related to this Project

Aufdenkampe, A. K., E. Mayorga, P. A. Raymond, J. M. Melack, S. C. Doney, and S. R. Alin. 2011. Riverine coupling of biogeochemical cycles between land, oceans and atmosphere. *Frontiers in Ecology and the Environment* 9(1): 53-60, doi:10.1890/100014.

Yoo, K., J. Ji, **A. Aufdenkampe**, and J. Klaminder. 2011. Rates of soil mixing and associated carbon fluxes in a forest versus tilled agricultural field: Implications for modeling the soil carbon cycle. Journal of Geophysical Research - Biogeosciences **116**:G01014, doi:10.1029/2010jg001304.

Aufdenkampe, A. K., E. Mayorga, J. I. Hedges, P. D. Quay, J. E. Richey, C. Llerena and A. V. Krusche. 2007. Organic matter in the Peruvian headwaters of the Amazon: Compositional evolution from the Andes to the lowland Amazon mainstem. *Organic Geochemistry*. 38(3): 337-364.

Mayorga, E., **A. K. Aufdenkampe**, C. A. Masiello, A. V. Krusche, J. I. Hedges, P. D. Quay, J. E. Richey and T. A. Brown. 2005. Young organic matter as a source of carbon dioxide outgassing from Amazonian rivers. *Nature* 436(7050): 538-541.

Aufdenkampe, A. K., J. I. Hedges, A. V. Krusche, C. Llerena and J. E. Richey. 2001. Sorptive fractionation of dissolved organic nitrogen and amino acids onto sediments within the Amazon Basin. *Limnology and Oceanography*, 46(8), 1921-1935.

Five Other Significant Publications

Battin, T. J., S. Luyssaert, L. A. Kaplan, **A. K. Aufdenkampe**, A. Richter, and L. J. Tranvik. 2009. The boundless carbon cycle. *Nature Geoscience* **2**, 598 – 600, doi:10.1038/ngeo618.

Richardson, D. C., L. A. Kaplan, J. D. Newbold and **A. K. Aufdenkampe**. 2009. Temporal dynamics of seston: A recurring nighttime peak and seasonal shifts in composition in a stream ecosystem. *Limnology and Oceanography* **54**(1): 344–354.

Hernes, P. J., A. C. Robinson and **A. K. Aufdenkampe**. 2007. Fractionation of lignin during leaching and sorption and implications for organic matter “freshness”. *Geophysical Research Letters* 34: L17401, doi:17410.11029/12007GL031017.

Aufdenkampe, A. K., D. B. Arscott, C. L. Dow, and L. J. Standley. 2006. Molecular tracers of soot and sewage contamination in streams supplying New York City drinking water. *Journal of the North American Benthological Society* 25(4): 928-953.

Richey, J. E., J. M. Melack, **A. K. Aufdenkampe**, V. M. Ballester and L. L. Hess. 2002. Outgassing from Amazonian rivers and wetlands as a large tropical source of atmospheric CO2. *Nature*, 416, 617-620.

(iv) Synergistic Activities

Advisor/Supervisor. 4 Post-Docs, 1 MS Student, 4 NSF REU Interns, 1 NASA OUR Earth Intern and >30 other project-supported undergraduate interns.

Reviewer. *Aquatic Geochemistry; Analytical Chemistry; Biogeochemistry; Belgium FWO proposals; Global Biogeochemical Cycles; Geochimica et Cosmochimica Acta; Estuarine, Coastal and Shelf Science; European Journal of Soil Science; Hydrological Processes; Journal of Environmental Quality; Journal of the North American Benthological Society; Limnology and Oceanography; Marine Chemistry; Nature; NSF proposals; NSF Hydrology Review Panel; NOAA proposals; Organic Geochemistry; United Kingdom NERC proposals; Water Research.*

Guest Lecturer and Field Instructor. Taught lectures and week-long mini-units in:
Introduction to Freshwater Ecology with T. Bott, University of Pennsylvania, 2004-2011.
Isotope Geochemistry with W. Ullman, University of Delaware, 2006.
Introduction to Field Ecology with J. Thorne, University of Pennsylvania, 2003.
Various Oceanography and Biogeochemistry courses at the Univ. of Washington and Seattle University, 1997 to 2001.

Member. American Soc. of Limnology & Oceanography (ASLO), Am. Geophysical Union (AGU).

Sustainable Agriculture Educator, U.S. Peace Corps, 1992-1994. Developed and implemented soil conservation programs and trained local extension agents. Central African Republic.

(v) Collaborators

(a) Collaborators in last 48 months (see publications for additional co-authors)

Rolf Aalto	Univ. of Washington, Seattle, WA.
Katarina Billups	Univ. of Delaware, Newark, DE
Paul Buckavekas	Virginia Commonwealth University, Richmond, VA
Peter Hernes	Univ. of California, Davis, CA
George Hornberger	Univ. of Virginia, Charlottesville, VA.
Louis Kaplan	Stroud Water Research Center, Avondale, PA.
Alex E. Krusche	Centro de Energia Nuclear na Agricultura (CENA), Brazil
Carlos Llerena	Universidad Nacional Agraria la Molina (UNALM), Peru
Laurence Maurice	Institut de recherche pour le Développement (IRD), France
Emilio Mayorga	Univ. of Washington, Seattle, WA
Denis Newbold	Stroud Water Research Center, Avondale, PA
Paul Quay	University of Washington, Seattle, WA
Jeffrey E. Richey	University of Washington, Seattle, WA
Donald L. Sparks	Univ. of Delaware, Newark, DE
William Ullman	Univ. of Delaware, Newark, DE
Kyungsoo Yoo	University of Minnesota. Minneapolis, MN

(b) Graduate and Postdoctoral Advisors:

John I. Hedges	M.S. & Ph.D. Advisor	University of Washington, Seattle
Jeffrey E. Richey	Ph.D. Co-Advisor	University of Washington, Seattle
James W. Murray	M.S. Co-Advisor	University of Washington, Seattle
Ronald Benner	Postdoctoral Supervisor	University of South Carolina, Columbia

(c) Theses Advisor and Postgraduate-Scholar Sponsor:

Diana Karwan	Post-Doc	SWRC. 2010-present.
Carl Rosier	Post-Doc	Univ. of Delaware. 2010-present.
Olesya Lazareva	Post-Doc	Univ. of Delaware. Co- w/ D. Sparks. 2010-present.
Rebecca Hays	Post-Doc	Univ. of Delaware. Co- w/ K. Billiups. 2010-present.
Linda Carter	Post-Doc	Stroud Water Research Center. 2005-2006.
Michelle C. Cogo	M.S.	Universidade de São Paulo, Brazil. Co- w A. Krusche

Jeffery S. Horsburgh, PhD
Research Assistant Professor, Utah Water Research Laboratory
Utah State University, Logan, UT 84322-8200
jeff.horsburgh@usu.edu

Professional Preparation

Utah State University	Environmental Engineering	BS, 1999
Utah State University	Civil and Environmental Engineering	MS, 2001
Utah State University	Civil and Environmental Engineering	PhD, 2009

Professional Appointments

2009-present	Research Assistant Professor, Utah Water Research Laboratory and Department of Civil and Environmental Engineering	Utah State University
2001-2009:	Research Engineer, Utah Water Research Laboratory	Utah State University
1999-2001:	Research Technician, Utah Water Research Laboratory	Utah State University
Summers 1997, 1998:	Summer Fellow	Idaho National Laboratory

Five Publications Relevant to Proposed Project

Horsburgh, J. S., D. G. Tarboton, D. R. Maidment, and I. Zaslavsky (2010), Components of an integrated environmental observatory information system, *Computers & Geosciences*, 37(2), 207-218, doi:10.1016/j.cageo.2010.07.003.

Horsburgh, J. S., A. Spackman Jones, D. G. Tarboton, D. K. Stevens, N. O. Mesner (2010), A sensor network for high frequency estimation of water quality constituent fluxes using surrogates, *Environmental Modelling & Software*, 25, 1031-1044, doi:10.1016/j.envsoft.2009.10.012.

Horsburgh, J. S., D. G. Tarboton, M. Piasecki, D. R. Maidment, I. Zaslavsky, D. Valentine, and T. Whitenack (2009), An integrated system for publishing environmental observations data, *Environmental Modeling and Software*, 24, 879–888, doi:10.1016/j.envsoft.2009.01.002.

Horsburgh, J. S., D. G. Tarboton, D. R. Maidment and I. Zaslavsky (2008), A relational model for environmental and water resources data, *Water Resources Research*, 44, W05406, doi:10.1029/2007WR006392.

Goodall, J. L., J. S. Horsburgh, T. L. Whiteaker, D. R. Maidment, and I. Zaslavsky (2007), A first approach to web services for the National Water Information System, *Environmental Modeling and Software,* 23(4), 404-411, doi:10.1016/jenvsoft.2007.01.005.

Five Other Publications

Spackman Jones, A., D. K. Stevens, J. S. Horsburgh, N. O. Mesner (2010), Surrogate measures for providing high frequency estimates of total suspended solids and total phosphorus concentrations, *Journal of the American Water Resources Association*, 47(2), 239-253, doi:10.1111/j.1752-1688.2010.00505.x.

Jackson-Smith, D. B., M. Halling, E. de la Hoz, J. P. McEvoy, and J. S. Horsburgh (2010), Measuring conservation program BMP implementation and maintenance at the watershed scale, *Journal of Soil and Water Conservation*, 65(6), 413-423, doi:10.2489/jswc65.6.413.

Horsburgh, J. S., D. G. Tarboton, K. A. T. Schreuders, D. R. Maidment, I. Zaslavsky, and D. Valentine (2010), HydroServer: A platform for publishing space-time hydrologic datasets, in Proceedings of the AWRA Spring Specialty Conference on GIS and Water Resources, Orlando, FL, March 29 – 31.

Tarboton, D. G., J. S. Horsburgh, D. R. Maidment, T. Whiteaker, I. Zaslavsky, M. Piasecki, J. Goodall, D. Valentine, and T. Whitenack (2009), Development of a community Hydrologic Information System, in Anderssen, R. S., R. D. Braddock, and L.T.H. Newham (eds) 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Modelling and

Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, July 2009, pp. 988-994, ISBN: 978-0-9758400-7-8.

Ames, D. P., J. S. Horsburgh, J. Goodall, D. G. Tarboton, T. Whiteaker, and D. R. Maidment (2009), Introducing the open source CUAHSI Hydrologic Information System desktop application (HIS Desktop), in Anderssen, R. S., R. D. Braddock, and L.T.H. Newham (eds) 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, July 2009, pp. 4353-4359, ISBN: 978-0-9758400-7-8.

Synergistic Activities

- *I lead a team of computer programmers and technicians* in the development of cyberinfrastructure and hydroinformatics databases, software and systems.
- *I am leading efforts to establish Internet-based watershed information systems* for the Bear River Watershed (UT, ID, and WY) and the Great Salt Lake watershed that serve as common data repositories for spatial and temporal data and provide users with data analysis and visualization tools over the Internet.
- *I serve on the Core Cyberinfrastructure Team (CCIT)* and as co-lead of the data Semantics and Integration Working Group for the NSF sponsored DataNet DataONE project designing cyberinfrastructure for long term data sharing and preservation.
- *I work on the CUAHSI Hydrologic Information System (HIS) team* participating in the design of the Observations Data Model, in developing software applications that use technology created by the CUAHSI HIS, and in deploying the HIS software.
- *I operate and maintain the hydrologic monitoring infrastructure* for the Little Bear River WATERS network test bed/experimental watershed.

Collaborators and Co-Editors

Daniel P. Ames (Idaho State University); Matthew E. Baker (University of Maryland, Baltimore County); Michelle Baker (USU); Christina Banderagoda (Silver Tip Solutions); Steven Corbato (University of Utah); Ernesto de la Hoz (USU); Jeff Dozier (University of California, Santa Barabara); Susan Gill (Stroud Water Research Center); Jonathan L. Goodall (University of South Carolina); Michael Halling (USU); Charles P. Hawkins (USU); Richard Hooper (Consortium of Universities for the Advancement of Hydrologic Science, Inc.); Douglas Jackson-Smith (USU); Norm Jones (Brigham Young University); David R. Maidment (University of Texas at Austin); Jamie P. McEvoy (USU); Nancy O. Mesner (USU); William Michener (University of New Mexico); Bethany T. Neilson (USU); Fred L. Ogden (University of Wyoming); Michael Piasecki (The City College of New York); David Rosenberg (USU); Ronald J. Ryel (USU); Kimberly A. T. Schreuders (USU); Charles Sims (USU); Amber Spackman-Jones (USU); David K. Stevens (USU); David G. Tarboton (USU); David Valentine (San Diego Supercomputer Center); James P. Verdin (United States Geological Survey); Timothy L. Whiteaker (University of Texas at Austin); Mark Williams (University of Colorado Boulder); Thomas Whitenack (San Diego Supercomputer Center); Ilya Zaslavsky (San Diego Supercomputer Center)

Graduate Advisor and PostDoctoral Sponsor:

David K. Stevens, Utah State University, Logan, UT (MS advisor)
David G. Tarboton, Utah State University, Logan, UT (PhD advisor)

Thesis Advisor and Postgraduate-Scholar Sponsor

Brant Whiting, Utah State University, Logan, UT (current MS student)
Total Graduate Students supervised: 1
Total Postdoctoral Fellows supervised: 0

KERSTIN A. LEHNERT

LAMONT-DOHERTY EARTH OBSERVATORY OF COLUMBIA UNIVERSITY
PALISADES, NY 10964
Phone: 845-365-8506   Fax: 845-365-8162
Email: lehnert@ldeo.columbia.edu

Professional Preparation:

1979	Diplom, Technische Universität Braunschweig, Germany
1989	Ph.D., Albert-Ludwigs-Universität Freiburg i. Breisgau, Germany Thesis title: "Petrology of Carbonatite Dikes in the Kaiserstuhl"

Appointments:

Since 2009	Senior Research Scientist (Project), Director, Integrated Earth Data Applications Research Group
2006-2009	Director, Geoinformatics for Geochemistry Program
Since 2005	Administrative Director for Research, Lamont-Doherty Earth Observatory
2003-2005	Coordinator for Research Administration, Office of the Director, Lamont-Doherty Earth Observatory
1996-2003	Sr. Staff Associate, Lamont-Doherty Earth Observatory of Columbia University project manager RIDGE PetDB; lab manager (DCP, EMP, ICP-MS), research associate
1985-1996	Staff associate, Max-Planck-Institut für Chemie, Department of Geochemistry, Mainz, Germany Lab manager (XRF, HPLC); system administrator; executive assistant of the Director of the Geochemistry department; safety engineer.
1980-1982	Research Assistant, Institut für Mineralogie und Petrologie, Universität Freiburg i. Breisgau, Germany

RELEVANT PUBLICATIONS TO THE PROPOSED WORK

Niu, X., Lehnert, K. A., Williams, J., Brantley, S. L., "CZChemDB and EarthChem: Advancing management and access of critical zone geochemical data", Applied Geochemistry, Vol. 26, Suppl. 1, S108-S111, 2011

Lehnert, K. A., Goldstein, S. L., Johansson, A., Murray, R. W., Pisias, N. G., Vinayagamoorthy, S., Djapic, B., "SedDB – A New Information System to Facilitate Use of Marine Sediment Geochemistry in Science and Education", MARGINS Newsletter, Vol. 18: 9-11, 2007.

*Cervato, C., Goldstein, S. L., Grossman, E. L., Lehnert, K., McArthur, J. M., "*Joint Discussion of Sedimentary Geochemistry Data Management Systems That Cross the Waterline", EOS Trans. AGU, 85, 44, 2004

Lehnert, K.A., Su, Y., Langmuir, C.H., Sarbas, B., Nohl, U., "A global geochemical database structure for rocks", G^3, Volume 1, 2000.

Staudigel, H., Helly, J., Koppers, A.A.P., Shaw, H., McDonough, W., Hofmann, A., Langmuir, C.H., Lehnert, K., Sarbas, B., Derry, L., Zindler, A., "Electronic Data Publication in Geochemistry". G^3, Volume 4, 2003.

FIVE OTHER PUBLICATIONS

Goldstein, S. L., Soffer, G., Langmuir, C.H., Lehnert, K. A., Graham, D. W., Michael, P. J., “Origin of a ‘Southern Hemisphere’ geochemical signature in the Arctic upper mantle”, Nature, 453, 89-93, 2008.

Carbotte, S.M., Arko, R. Chayes, D., Haxby, W., Lehnert, K., O’Hara, S., Ryan, W.B.F., Weissel, R.A., Shipley, T., Gahagan, L., Johnson, K., Shanks, T., “New Integrated Data Management System for Ridge2000 and MARGINS Research.” EOS Trans AGU, 85, 51, 553, 2004

Michael, P.J., Langmuir, C.H., , Dick, H. J. B. , Snow, J. E., Goldstein, S. L., Graham, D. W., Lehnert, K., Kurras, G., Jokat, W., Mühe, R., Edmonds, H. N., “Magmatic and amagmatic seafloor generation at the ultraslow-spreading Gakkel ridge, Arctic Ocean”, Nature, 423, 956-961, 2003

Hubberten, H.-W., Katz-Lehnert, K.A., Keller, J., "Carbon and oxygen isotope investigations in carbonatites and related rocks from the Kaiserstuhl, Germany", Chem. Geol., 70, 257-274, 1988.

Arndt, N.T., Lehnert, K.A., Vasil'ev, Y., "Meimechites: highly magnesian alkaline magmas from the subcontinental lithosphere?", Lithos 34, 41-59, 1995.

SYNERGISTIC ACTIVITIES

- Co-founder of the EarthChem consortium (http://www.earthchem.org)
- Member of the Policy Committee of UCAR’s Unidata facility
- Member of various Geoinformatics committees, e.g. Vice-chair of the Geoinformatics Division of the Geological Society of America, member of the Executive Committee of the Earth & Space Science Informatics Focus Group of the American Geophysical Union
- Organization of numerous short courses & workshops
  - *“Using Online Igneous Geochemical Databases for Research and Teaching”.* Washington State University, March 2008.
  - *“Data Resources for the Geosciences”* (Joint short course of GfG and MGDS at GSA 2007); Denver CO, October 2007.
  - *“Data Resources for the Geosciences”* (Joint short course of GfG and MGDS at AGU 2007, sponsored by MARGINS); San Francisco, December 2007.
- Mentoring of Postdoctoral Research Scientist Dr. Leslie Hsu

COLLABORATORS AND OTHER AFFILIATIONS

Lee Allison (Arizona State Geological Survey), Susan Brantley (Pennsylvania State University), Carla Moore (National Geophysical Data Center, NOAA), Michael Mottl (University of Hawaii), Douglas J. Walker (Kansas University)

EMILIO MAYORGA – BIOGRAPHICAL SKETCH

Address: Applied Physics Laboratory, University of Washington
1013 NE 40th St., Seattle, WA 98105-6698
Tel: (206)-543-6431, mayorga@apl.washington.edu

(a) Professional Preparation

Massachusetts Institute of Technology	Environmental Engineering Science	B.S., 1992
University of Washington	Chemical Oceanography	M.S., 1997
University of Washington	Chemical Oceanography	Ph.D., 2004
Rutgers University	Global River Nutrient Exports	2007-2008

(b) Appointments

Oceanographer 4, Jan. 2009 – present, Applied Physics Laboratory, University of Washington, Seattle, WA, USA.

Research Associate, Jan. 2007 – Dec. 2008, Institute of Marine & Coastal Sciences, Rutgers University, New Brunswick, NJ, USA.

Principal GIS Analyst, Sept. 2001 – Dec. 2006, Surface Water Management, Snohomish County, Everett, WA, USA.

Research Assistant, 1993-2001, School of Oceanography, University of Washington.

(c) Publications – Five most closely related to this project

Aufdenkampe, A.K., **E. Mayorga**, P.A. Raymond, J. Melack, S.C. Doney, S.R. Alin, R.E. Aalto and K. Yoo. 2011. Riverine coupling of biogeochemical cycles between land, oceans and atmosphere. *Frontiers in Ecology & Environment* 9(1): 53-60, doi:10.1890/100014

Mayorga, E., T. Tanner, R. Blair, A.V. Jaramillo, N. Lederer, C.M. Risien and C. Seaton. 2010. The NANOOS Visualization System (NVS): Lessons learned in data aggregation, management and reuse, for a user application. *Proc. MTS/IEEE Oceans'10*

Mayorga, E., S.P. Seitzinger, J.A. Harrison, E. Dumont, A.H.W. Beusen, A.F. Bouwman, B. Fekete, C. Kroeze and G. Van Drecht. 2010. Global Nutrient Export from WaterSheds 2 (NEWS 2): Model development and implementation. *Environmental Modeling & Software* 25: 837-853, doi:10.1016/j.envsoft.2010.01.007

Seitzinger, S., **E. Mayorga**, A.F. Bouwman, C. Kroeze, A.H.W. Beusen, G. Billen, G. Van Drecht, E. Dumont, B.M. Fekete, J. Garnier and J.A. Harrison. 2010. Global nutrient river export: A scenario analysis of past and future trends. *Global Biogeochemical Cycles* 24: GB0A08, doi:10.1029/2009GB003587

Risien, C.M., J.C. Allan, R. Blair, A.V Jaramillo, D. Jones, P.M. Kosro, D. Martin, **E. Mayorga**, J.A. Newton, T. Tanner and S.A. Uczekaj. 2009. The NANOOS Visualization System: Aggregating, displaying and serving data. *Proc. MTS/IEEE Oceans'09*

Five Other Significant Publications

Fekete, B.M., D. Wisser, C. Kroeze, **E. Mayorga**, L. Bouwman, W.M. Wollheim and C. Vörösmarty. 2010. Millennium Ecosystem Assessment scenario drivers (1970-2050): Climate and hydrological alterations. *Global Biogeochemical Cycles* 24: GB0A12, doi:10.1029/2009GB003593

Mayorga, E. 2008. Carbon cycle – Harvest of the century. *Nature* 451: 405-406

Aufdenkampe, A.K., **E. Mayorga**, J.I. Hedges, C.A. Llerena, P.D. Quay, J. Gudeman, A.V. Krusche and J.E. Richey. 2007. Organic matter in the Peruvian headwaters of the Amazon:

Compositional evolution from the Andes to the lowland Amazon mainstem. *Organic Geochemistry* 38: 337-364

Mayorga, E., M.G. Logsdon, M.V.R. Ballester and J.E. Richey. 2005. Extracting cell-to-cell land surface drainage paths from digital channel networks, with an application to the Amazon basin. *Journal of Hydrology* 315: 167-182

Mayorga, E., A.K. Aufdenkampe, C.A. Masiello, A.V. Krusche, J.I. Hedges, P.D. Quay, J.E. Richey and T.A. Brown. 2005. Young organic matter as a source of carbon dioxide outgassing from Amazonian rivers. *Nature* 436: 538-541

(d) Synergistic Activities

Reviewer. *Journals:* Biogeochemistry, Biogeosciences, Ecosystems, Global Biogeochemical Cycles, Hydrological Processes, Limnology and Oceanography, Nature. *Proposals and Research Programs:* NASA, NSF, USGS.

Member. American Geophysical Union (AGU), American Society of Limnology and Oceanography (ASLO), Ecological Society of America (ESA)

Working Group Participation.

Carbon in Tropical Rivers, Organization for Tropical Studies (OTS) NSF Research Coordination Network, 2010 – present.

REgional Carbon Cycle Assessment and Processes (RECCAP), 2010 – present.

Global Nutrient Export from Watersheds (Global NEWS – UNESCO's Intergovernmental Oceanographic Commission), 2007 – present.

Merging terrestrial and aquatic perspectives of biogeochemistry, National Center for Ecological Analysis and Synthesis (NCEAS), 1999 – 2002.

Environmental Informatics Involvement. *EPA Network Exchange (Pacific NW Water Quality Data Exchange), Geochemical Earth Reference Model (GERM), Ocean Biogeographic Information System (OBIS)*, *US Integrated Ocean Observing System (IOOS), Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI).*

(e) Collaborators & Other Affiliations

Collaborators and Co-Editors in the last 48 months

Anthony Aufdenkampe	Stroud Water Research Center, Avondale, PA
Lex Bouwman	Environmental Assessment Agency (MNP), The Netherlands
Balazs Fekete	The City College of New York (CCNY/CUNY), New York, NY
John Harrison	Washington State University, Vancouver, WA
Jens Hartmann	University of Hamburg, Germany
David Jones	Applied Physics Lab., University of Washington, Seattle, WA
Albert Kettner	University of Colorado, Boulder, CO
Carolien Kroeze	University of Wageningen, The Netherlands
Kon-Kee Liu	National Central University, Taiwan
Jan Newton	Applied Physics Lab., University of Washington, Seattle, WA
Peter Raymond	Yale University, New Haven, CT
Weijin Yan	Chinese Academy of Sciences, China
Ilya Zaslavsky	San Diego Supercomputer Center, Univesity of California, CA

Graduate and Postdoctoral Advisors

Jeffrey E. Richey	M.S. & Ph.D. Advisor	University of Washington, Seattle, WA
Allan Devol	M.S. Co-advisor	University of Washington, Seattle, WA
John I. Hedges	Ph.D. Co-advisor	University of Washington, Seattle, WA
Paul Quay	Ph.D. Co-advisor	University of Washington, Seattle, WA
Sybil Seitzinger	Postdoctoral Supervisor	Rutgers University, New Brunswick, NJ

Ilya Zaslavsky

San Diego Supercomputer Center	Telephone: 858 534-8342
University of California, San Diego	Fax: 858 534-5113
9500 Gilman Drive, MC 0505	E-mail: zaslavsk@sdsc.edu
La Jolla, CA 92093-0505	

Professional Preparation

Moscow State University, Moscow, Russia	M.A. Geography	1985
USSR Academy of Sciences Moscow, Russia	Ph.D. Equivalent	1990
University of Washington	Ph.D. Geography	1995

Appointments

2000	Director, Spatial Information Systems Laboratory, San Diego Supercomputer Center
1997	GIS Staff Scientist, Education Center on Computational Science and Engineering, Adjunct Professor, Mathematical and Computer Sciences, San Diego State Univ.
1995	Assistant Professor, Geography, Western Michigan University
1993	GIS Prog./Systems Analyst, Schlosser Geographic Systems, Inc., Seattle, WA
1987	Research Scientist, Institute of Geography, USSR Academy of Sciences
1987	Consultant, Software Engineer, Research Center “ROST”, Small Enterprise “Context”, Moscow, Russia

Related Publications

[1] Zaslavsky, I., and Maidment, D. R., (2011), “Service orientation in the design of a community hydrologic information system”, In G. R. Keller and C. Baru (eds.) *Geoinformatics. Cyberinfrastructure for the Solid Earth Sciences*, Cambridge Univ Press. pp. 193-209.

[2] Zaslavsky, I., C. Baru, K. Bhatia, A. Memon, P. Velikhov, V. Veytser, “Grid-enabled mediation services for geospatial information”. In *Next Generation Geospatial Information. From Digital Image Analysis to Spatiotemporal Databases*, ed. by P. Agouris & A. Croitoru), Taylor & Francis, 2005, pp. 15-24.

[3] CBEO Project Team (W. Ball, D. Brady, M. Brooks, R. Burns, B. Cuker, D. DiToro, T. Gross, W.M. Kemp, L. Murray, R. Murphy, E. Perlman; M. Piasecki, J. Testa, I. Zaslavsky), "A Prototype System for Multi-Disciplinary Shared Cyberinfrastructure – Chesapeake Bay Environmental Observatory (CBEO)," *ASCE J. Hydrologic Engineering*, 13(10): 960-970 (2008).

[4] Horsburgh, J. S., D. G. Tarboton, D. R. Maidment and I. Zaslavsky, “A Relational Model for Environmental and Water Resources Data,” *Water Resour. Res*., 44, 2008: W05406, doi:10.1029/2007WR006392.

[5] Maidment, D. R., I. Zaslavsky, J. Horsburgh, 2006. Hydrologic Data Access Using Web Services. *Southwest Hydrology* 5(3): 16-17.

Other Significant Publications

[1] Ioffe, G., T. Nefedova, I. Zaslavsky, "From spatial continuity to fragmentation: the case of Russian farming", *Annals of the Association of American Geographers*, **94** (4): 913-943.

[2] I. Zaslavsky, 2003 "Online Cartography with XML", in "*Maps and the Internet*", Elsevier, pp. 171-196.

[3] Ioffe, G., T. Nefedova, I. Zaslavsky, 2006, *The End of Peasantry? The Disintegration of Rural Russia.* University of Pittsburgh Press, 258 pp. ISBN: 082294295X.

[4] I. Zaslavsky, H. He, J. Tran, M.E. Martone, A. Gupta, 2004. Integrating Brain Data Spatially: Spatial Data Infrastructure and Atlas Environment for Online Federation and Analysis of Brain Images, Proceedings of the 15th International Workshop on Database and Expert Systems Applications (DEXA 2004), Biological Data Management (BIDM'04), pp. 389-393.

[5] Gupta, A., R. Marciano, I. Zaslavsky, C. Baru, "Integrating GIS and Imagery through XML-Based Information Mediation." In P. Agouris and A. Stefanidis (Eds.) *Integrated Spatial Databases: Digital Images and GIS*, Lecture Notes in Computer Science, Vol. 1737, 1999.

Synergistic Activities

- Co-Chair, Hydrology Domain Working Group, Open Geospatial Consortium and WMO (http://external.opengis.org/twiki_public/bin/view/HydrologyDWG/)
- Lead, INCF Digital Atlasing Infrastructure Task Force (http://www.incf.org/about/programs/atlasing)
- Editor, CUAHSI WaterML specification (http://www.opengeospatial.org/standards/dp)
- Designer and lead developer of "Sociology Workbench", a collection of on-line tools for survey data analysis on the Web (1998 – 2001) (http://www.edcenter.sdsu.edu/swb/)
- Technical coordinator for the "Mapping People with ALS" project (http://geo.sdsc.edu/ilya/als/index.html)

.

Collaborators

Chaitanya Baru (SDSC), Amarnath Gupta (SDSC), Grigory Ioffe (Ratford University), David Maidment (UT Austin), Maryann Martone (UCSD), Keith Pezzoli (UCSD), Michael Piasecki (Drexel University), David Tarboton (USU)

Graduate Students and Postdoctoral Researchers

Debra Alfonso, Western Michigan University
Fahad Alhassan, Western Michigan University
Lotta Jarnefelt-Burns, Western Michigan University
John Burt, Western Michigan University

Graduate and Postgraduate Advisors

Nicholas Chrisman, University of Washington
Georgy Lappo, Russian Academy of Sciences

Budget Justification
University of Colorado-Boulder

Salaries and Wages
Investigators. No salary is requested for PI Williams. Salary for teaching relief is requested. Williams teaching schedule is 2 and 1 classes per semester per year. The teaching relief will free him from teaching responsibilities one semester each for the next two years. The teaching relief will provide Williams with time to concentrate on administering this proposal, communicate with CZO PIs, host and/or attend workshops, and conduct site visits to each CZO. No salary is requested for Co-I Aufdenkampe. However, he will receive a stipend as the lead on the synthesis workshops.

Staff salaries are included for maintenance of the web site (Lubinski) and developing web GIS/visualization tools for sites (Parrish). We also request support for a 50% FTE in year two to help with the transition of CZOData (person to be determined) to an operational system that will come online with the CZO renewal in two years.

Student hourly support includes $2,000 to support numerous activities, including data entry, web updates, and GIS visualization help. In most cases we hire work-study students, which provides a 4:1 match to our CZO dollar.

Worshops.
Skill sharing and knowledge transfer is a large component of this proposal. Implementation of the IMC is integral to the success of CZOData. Therefore, we have allocated a stipend of $8,000 dollars each year that goes to each CZO science representative to the IMC. Stipends and travel to the annual IMC meeting results in a commitment of approximately $54,000 per year (before overhead) to the IMC.

The annual synthesis workshop provides an opportunity for mid-course correction for CZOData. The science topic will be decided by the PIs. Aufdenkampe will chair the synthesis workshops and a stipend of $8,000 per year is allocated for his participation.
Aufdenkampe will serve as a liasion between CZOData and the synthesis scientists, though some site information managers and subcontractors will attend the annual synthesis workshop.

A GIS/visualization workshop will provide the opportunities for site information managers to exchange information on software and implementation of tools to improve the look, feel, and accessibility of site web pages. For example, one goal of the workshop is to develop “fly-bys” at each site that provide an appealing overview of the spatial domain of each site.

Additional workshops are included in either the subcontractors budget (LDEO will lead a workshop on geochemistry data in collaboration with Penn State) or are included in carry-over funds from existing CZO supplements. These include a geophysics workshop to be hosted by BC-CZO and a spatial workshop to be hosted by SS-CZO.

Travel.
The travel budget is $14,000 per year to cover expenses of PI Williams to attend data manager meetings, visit each CZO site, etc. The subcontractor budgets include similar travel costs. Travel is budgeted for participants for each of the designated workshops.

Subcontractors. Each of the subcontractors has a major role in this proposal and each is integral to the success of this proposal. Below is a synopsis of the role of each subcontractor. More detailed information is in their individual budget justifications. We have also included their individual Scope of Work in the “Other Supplementary Documents” section of Fastlane to emphasize their individual contributions to this proposal.

University of California San Diego (UCSD). UCSD will conduct research and development activities focused on the development of central CZO data harvesting, archiving and dissemination system, and supporting infrastructure and protocols. Specific tasks include, but are not limited to (see their Scope of Work in supplementary documentation):

- 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. 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. Ensure that the data are periodically re-harvested, and the information provided via services is appropriately updated.
- In coordination with CZO PIs and project leadership team, 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.
- In coordination with the USU team, 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.
- In collaboration with the other project teams, develop solutions for registering CZO-collected data of other data types (not time series). This includes, primarily, geochemical samples and spatial data layers, and can be extended based on project priorities.
- Ensure that the harvested data become available via standards-based service interfaces, and update the interfaces when new standards are adopted.
- Host CZO online content management system, and, in collaboration with the UC-Boulder team, integrate it with the CZO data discovery portal.

Columbia University. The Geoinformatics for Geochemistry (GfG) Program at Columbia University will contribute to the development of CZOData by developing data systems and services that support the management of geochemical and sample data. The GfG Program is part of the NSF-funded Integrated Earth Data Applications (IEDA) facility (http://www.iedadata.org). With funding provided through this subcontract, they will complete a suite of developments as described below that are aimed to (a) ensure proper formatting and documentation of geochemical data for preservation, display, and inclusion in the CZO Portal, aligned with community-based data reporting standards, (b) enhance access to CZO geochemical data by developing a web-based search interface for CZchemDB and incorporating CZchemDB data in the EarthChem portal, (c) ensure registration and unique identification of CZO samples by developing the CZO Registration Agent in the International Geo Sample Number (IGSN) system.

Utah State University (USU) will assist in the design and implementation of CZOData. More specifically, USU will complete the following tasks:

1. Assist in the overall design of the architecture for and implementation of the CZO Integrated Data Management System
2. Provide coordination between the proposed project and the existing CUAHSI Hydrologic Information System project

3. Host, populate, and perform maintenance and upgrades to the CZO Shared Vocabulary System
4. Assist in the identification and documentation of metadata requirements for CZO datasets
5. Assist in design of templates to be used by CZO Data Managers for creation of metadata describing CZO datasets
6. Assist in the design and development of tools to support the creation and publication of display files by CZO Data Managers
7. Assist in the development of training materials for CZO Data Managers and attend and assist with CZO Data Manager trainings and workshops
8. Assist in developing an information management working group for CZO Data Managers
9. Attend at least one annual project meeting and annual data manager's meetings

University of Washington. The Applied Physics Laboratory (APL) of the University of Washington will develop an online data access and visualization tool for CZOData that is interactive, feature-rich, user-friendly and responsive to CZO needs. APL will deploy this capability by adapting and extending the regional data integration and visualization system used by the Northwest Association of Networked Ocean Observing Systems (NANOOS). APL will develop the "CZO VS" in close collaboration with SDSC, LDEO and the CZO network. The CZO VS will focus on ingestion of time series data generated by CZO sites and distributed via the SDSC system using CUAHSI HIS-based web services and Open Geospatial Consortium (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. The capability to search across CZO sites and explore the resulting network-side data will be developed jointly with SDSC and its CZO GeoPortal catalog, in parallel with the site-focused approach of the CZO VS. 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 ontologies implemented by SDSC (e.g., measurement variables, platform types, asset types). 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. This functionality will represent a substantial component of the APL contribution to this project that will involve both user-interface development by APL and data-discovery optimizations developed and implemented jointly with SDSC.

The Institute of Arctic and Alpine Research (INSTAAR) is a unique unit of the University of Colorado. It is solely dedicated to research and completely separate, though complementary, to the traditional academic department. INSTAAR is an interdisciplinary institute, and as such, houses large and complex research programs. Each individual investigator has access to a depth and broad range of scientific expertise that complements his/her research. Support is requested for technical assistance that is
necessary for executing, administering, and accounting for the proposed project.
Communication and duplication includes such items as toll calls, postage, fax charges, telephone equipment and photocopies that specifically support this project. The University of Colorado maintains telecommunications services that include the capability of tracking expenses to specific projects. In addition, INSTAAR maintains photocopy services that include capability of tracking expenses to specific projects.

Budget Justification

The work proposed of this subcontract will be carried out by staff members of the Geoinformatics for Geochemistry (GfG) Program at the Lamont-Doherty Earth Observatory. We request salary support for the staff members who will execute the work and for travel to our collaborators at Pennsylvania State University (PSU) at State College, PA, and to annual meetings with the CZOData investigators at University of Colorado in Boulder, CO.

1. Salaries

0.5 months/year salary support for GfG Director Kerstin Lehnert will be required to contribute to her oversight and management of the GfG staff who will complete the development tasks of this subcontract. K. Lehnert has been involved with CZO data management activities for several years and will also provide scientific guidance for the design of the CZOData components to be developed by GfG, coordinate development activities with CZOData PI Mark Williams and other collaborators of the CZOData project, and participate in CZOData meetings.

Sze-Ling Chan, Senior System Architect for the Geoinformatics for Geochemistry Program, will oversee and guide the technical aspects of the developments for CZODataby the GfG Program. We request 1 month of salary support per year.

We request 2.5 months/year of salary support for a TBH Data Manager to help design the geochemistry data templates for posting geochemical data at the site web sites, to help ensure posting of geochemical data at the site web sites, and to add new data to the CZchemDB database.

We request 10 months/year of salary support for a TBH IT Developer to build the sample-based data components of the CZOData, which will include development of the web-accessible CZchemDB, scripts to export CZchemDB data to EarthChemXML, and to set up the IGSN Registration Agent for the CZO community.

2. Fringe Benefits & Indirect Costs

Fringe benefits are assessed at Columbia University's government negotiated rate of 33.3%, which became effective July 1, 2011 and is estimated beyond that.

3. Travel

We have budgeted funds for two people per year to visit the CZchemDB group at Pennsylvania State University in State College, PA. In year 1, these visits are required to work with the PSU group on the development of the scripts to export CZchemDB data to EarthChemXML, on templates for geochemical data submission and on preparing specifications for web enabling CZchemDB. In year 2, these visits will be used to test and review the web interface to CZchemDB.R/T airline tickets to Denver, ground transportation, and accommodation for two trips per year have been budgeted to participate in the CZOData working group meetings in Boulder.

4. Participant Support

In year 1, we will organize a workshop together with our collaborators at Pennsylvania State University for CZO investigators, who work with sample-based geochemical data, to refine reporting requirements for geochemical data, design the data templates, and gather their input on the specifications of the web interface for CZchemDB. We expect to invite 10 participants. We have budgeted $1,000 per participant to support participants' travel expenses and other logistical expenses for the workshop. The location of the workshop has not been determined.

5. Other Direct Costs

Amounts requested for project materials computer supplies, and communication, are the minimum required for this proposal's operation.

Material and Supplies

We are requesting $200 per year to cover expenses for project related materials and computer supplies.

Communication/Shipping:

We are requesting $300per year to cover expenses for communication with the team members at the collaborating institutions to offset postage and costs for conference calls.

Indirect Costs:

Are assessed at Lamont-Doherty's government-negotiated rate of 53%.

Budget Justification (UCSD)

This request is to cover costs for the project PI, Ilya Zaslavsky, and Programmer-Analyst III, Thomas Whitenack.

We are requesting 1.2 calendar month's effort for Dr. Zaslavsky in both years of the project. He will focus on technical issues of the overall cross-CZO data management system design and direct the development of the CZO Central data harvesting, archiving and dissemination system. In addition, he will manage SDSC's contractual, budgetary and reporting obligations and collaborate with other project PIs and with CZO PIs.

We are requesting 11.1 and 10.7 calendar month's effort for Thomas Whitenack, in the two years of the project respectively. He is the primary developer of the CZO Central system. He will work with the project PIs and with CZO data managers to further develop the CZO central site and publish CZO data as web services, following specific tasks outlined in the statement of work.

Senior personnel have staff appointments and are supported 100% on soft money.

Fringe Benefits
Fringe benefits have been listed at the projected rate for each salaried employee in accordance with University policy.

Travel
Ten domestic trips are scheduled for the project members in the two years of the project: to visit each of the CZO sites, and to participate in CZO-wide data management workshops. The site visits are critical for understanding specific needs and data management workflows adopted by each CZO site, and developing strategies for integrating locally collected data into the central system.

Equipment
No equipment is requested. The SDSC Spatial Information Systems Lab has purchased and installed a CZO Central server during the previous pilot stage of CZO data management project, and will continue to manage it.

Other Direct Costs
Supplies include materials for talk preparation, hardware and software upgrades, and other computer related supplies to be used in conjunction with this research. Computer-related costs include yearly rack fees and other maintenance and update charges associated with supporting the CZO Central server and archive. Charges for telephone lines and tolls, research journal subscriptions and other project specific costs when directly related to this work have also been included. Communications/Computing costs have been included for telephone and associated voice and data communications charges directly related to the individuals working on the project. Other direct costs include technical support, equipment maintenance, and costs for project-specific expenses directly related to this work. In each instance, such support is crucial to the accomplishment of this work.

Indirect costs

The University of California, San Diego/San Diego Supercomputer Center incurs indirect costs at a rate 54.5% on modified total direct costs (MTDC). MTDC excludes tuition remission, equipment and the excess of $25,000 on subcontracts. IDC increases to 55% 7/1/12 in accordance with DHHS agreement.

BUDGET JUSTIFICATION
APPLIED PHYSICS LAB, UNIVERSITY OF WASHINGTON

SALARIES: The salaries will be paid in accordance with University policy and at the rates in effect at the time. This budget is based on actual 15 July 2011 salaries, plus cost of living allowance and anticipated merit increases.

Emilio Mayorga, Principal Investigator will oversee all project activities at the University of Washington and coordinate with external project partners. He will also lead the development of data harvesting and integration schemes for the adaptation of the NANOOS Visualization System (NVS) to be developed for this project. Mayorga and Tanner will jointly translate project needs into a consistent architecture and prioritized system enhancements.
(Yr1: 1.20 months; Yr2: 0.90 months)

Troy Tanner, Engineer IV, is the primary NVS developer. He will be responsible for the adaptation of the NVS user interface to the project's needs. Tanner and Mayorga will jointly translate project needs into a consistent architecture and prioritized system enhancements.
(Yr1: 1.00 months; Yr2: 1.30 months)

BENEFITS AND LEAVE ALLOWANCE: The fringe benefits rates used are in accordance with the University of Washington Grants Information Memorandum Number 3, http://www.washington.edu/research/osp/gim/gim3.html. While they are the rates that the University began using effective 7/16/2011, they are not yet approved by the Department of Health and Human Services (Cognizant Federal Agency). The University has proposed these fringe benefit rates in its F&A rate negotiations with DHHS, and expects them to be approved as reasonable and allowable. APL-UW benefit rate for professional staff salary is 30.3%. The rate for APL leave allowance for professional staff is 25.7% of salaries. Therefore, the total fringe benefit rate for professional staff is 56% of salaries.

EQUIPMENT: An Iris 1251 server will be purchased in year 2 to support data management activities. Estimated cost is $5,222 with WA state sales tax and shipping and handling fee included.

TRAVEL: Travel costs are budgeted in accordance with University and State of Washington regulations. Costs for each trip are detailed below based on contract airfare, current per diem rates, and appropriate fees. The detailed budget includes an inflation factor of 4% in years 1 and 2.

Travel in year 1 for 1 investigator from Seattle, WA – Oracle, Arizona to attend a CZO network meeting.

Roundtrip airfare Seattle, WA – Oracle, AZ	$768
Per diem 4 day @ $128/day	$512
Misc. Expenses/ground transportation	$248
Total	$1,528

Travel in year 2 for 1 investigator from Seattle, WA – San Francisco, CA to attend the Fall American Geophysical Union (AGU) conference and meet with CZO partners during the conference to discuss project priorities, status and plans.

Roundtrip airfare Seattle, WA – San Francisco	$337
Per diem 4 day @ $265/day	$1,060
Misc. Expenses/ground transportation	$260
Total	$1,657

OTHER DIRECT COSTS:

1. MATERIALS AND SUPPLIES: Project specific supplies are included at $200 in year 1 and $251 in year 2.

2. APL PRORATED DIRECT COSTS: The University Facilities and Administration (F&A) rate applied to APL-UW is lower than the rate elsewhere on campus (17% vs. 54%) and does not recover the Laboratory's central costs. These are recovered by applying a Prorated Direct Cost to total salaries. Prorated Direct Costs include such expenses as salaries and employee benefits for central service employees, administrative data processing, communications, and some facilities costs. APL-UW's Prorated Direct Costs have been reviewed and accepted by the Navy's Resident Administrative Contracting Officer, Sandra Thomson, Office of Naval Research, Seattle Regional Office, per letter dated 06 October 2010.

FACILITIES AND ADMINISTRATION RATE: APL-UW's negotiated rate is 17% of Modified Total Direct Costs (MTDC). MTDC includes all direct costs less equipment, graduate operating fees, and the amount of subawards above the initial $25,000. The current F&A Rate Agreement with DHHS is dated 30 November 2010.

The APL-UW may rebudget within the total estimated cost.

FACILITIES, EQUIPMENT AND OTHER RESOURCES

LABORATORY: The Applied Physics Laboratory (APL) of the University of Washington is a self-contained, fully equipped research facility. APL is equipped with machine and electronics shops, a library, and drafting and publications facilities. Computing needs are addressed by a combination of special purpose machines, central facilities, and networked personal computers and workstations. Special laboratories are available for electronic system development and physical acoustics, and ice studies.

COMPUTER: APL computers include Linux clusters for computationally intensive tasks, Apple Macintosh and Windows workstations, black and white and color printers and a myriad of software applications.

OFFICE: APL employees have offices and standard office capabilities such as fax, electronic mail, mail services, phones, voicemail, etc. No additional office or office capabilities are necessary to conduct the proposed research.

OTHER: APL maintains several vessels for use in fresh and saltwater activities.

MAJOR EQUIPMENT: APL maintains an equipment pool of oceanographic instrumentation. This instrumentation is quality controlled both by in-house expertise and manufacturer's calibration and repair. Current meters, temperature-conductivity sensors, acoustic releases, deck sets and other equipment necessary may be taken from this pool, at cost only of expendables (e.g. batteries, mooring line, hardware), calibration and refurbishment. Projects can draw from the APL logistics equipment pool for additional equipment or field support items necessary to perform field work.

Budget Justification
Utah State University

Senior Personnel

The budget for Utah State University (USU) requests 0.5 months per year of salary for Jeffery S. Horsburgh, 0.25 months per year for David G. Tarboton, and 1.5 months per year for Kimberly A. T. Schreuders. Horsburgh will act as the USU Principal Investigator and will coordinate the work done at USU. He will assist in the design and maintenance of the CZO Shared Vocabulary System and will participate in the design and implementation of the CZO Data Management System Architecture. Tarboton will provide coordination with ongoing work on the CUAHSI Hydrologic Information System and will also participate in the high level design and implementation of the CZO Data Management System Architecture. Schreuders will assist in performing maintenance and upgrades to the CZO Shared Vocabulary System, will attend and assist with the CZO Data Manager trainings and workshops, will assist in the development of a CZO information management working group, and will assist in the design and development of templates used by CZO Data Managers for creating metadata describing CZO datasets. Senior personnel salary rates in year 2 include a 3% inflation rate as authorized by USU.

If this proposal is funded, total NSF support from all NSF projects on which Jeffery S. Horsburgh is working could exceed 2 months per year. Dr. Horsburgh is a Research Assistant Professor at Utah State University and is on a 12-month soft money funded research appointment (all 12-months of support from research projects). If awarded, the requested salary will need to be specifically approved in the award.

Students

A total of 1.5 months in year one and one month in year two has been allocated at $2,080 per month ($12 per hour with a 3% inflation rate in year 2 as authorized by USU) for a student programmer who will assist USU senior personnel in maintaining the CZO Shared Vocabulary System and in assisting Schreuders with the development of CZO metadata templates. The student programmer will by jointly supervised by Horsburgh and Schreuders.

Fringe Benefits

Fringe benefits for faculty and professional staff at USU are calculated at a rate of 43% for year 1 and 43.5% for year two. Fringe benefits for the student programmer are calculated at a rate 8.3 % (as per Utah State University standard benefit rates).

Travel
We have budgeted $9000 in years one and two for domestic travel for all project personnel at USU. This amount will cover 2 USU personnel (Schreuders and Horsburgh) to attend one project meeting in each year. The remaining balance of the budgeted travel will support Schrueders in attending CZO Data Manager trainings and CZO site visits to assist and train CZO data managers.

Other Direct Costs

We have budgeted $1000 in years one and two to cover the costs of hosting the CZO shared vocabulary system on computer servers maintained at USU. These funds will be used to maintain necessary software licenses and necessary computer maintenance. Additionally, we have budgeted $850 in year one and $1000 in year two to cover office materials and supplies, telephone charges, other supplies, and to cover the costs of publishing our work.

Indirect Costs

Utah State University recovers indirect costs at a rate of 41% of modified total direct costs, which is the Federally Negotiated Rate for USU. Indirect costs are not charged for capital equipment in excess of $5,000.

FACILITIES, EQUIPMENT, AND OTHER RESOURCES

1. University of Colorado Boulder

The Boulder Creek CZO is housed at INSTAAR, a research unit within the University of Colorado-Boulder. INSTAAR has established a major computational facility called the Environmental Computation and Imaging (ECI) Facility. This facility houses a Sun Enterprise 6500 supercomputer (named deeppurple) with twenty-four 450 MHz processors supported with 24GB, which is available for use by BC-CZO. Additionally, BC-CZO has available three Unix Servers, located at INSTAAR, along with numerous Unix clients. We also have numerous desktop and laptop PCs and Macs, along with digitizing tablets, etc.

For this proposal, the major facilitites and equipment needed are computational ability and storage. These duties are performed by our subcontractors. Below we highlight the facilitites, equipment, and other resources of our contractors.

2. Utah State University

Utah State University (USU)'s team is located at the Utah Water Research Laboratory (UWRL), which is a multi-building facility located on the Logan River, in Logan, Utah. The UWRL operates within the academic environment of the Department of Civil and Environmental Engineering and collaborates with government and private sectors to address technical and societal aspects of water-related issues, including quality, quantity, distribution, and conjunctive use. This is accomplished through providing more than 100,000 square feet of state-of-the-art laboratory, computer, and office space. The UWRL facility contains offices and laboratories that house engineers, scientists, laboratory technicians, and students from a variety of engineering and water science disciplines. Designed to meet the needs of these disciplines, the UWRL has well equipped environmental quality and hydraulics laboratories; computer facilities; conference rooms; and administrative, faculty, staff, and student offices (see http://uwrl.usu.edu). The 11,000-square foot Environmental Quality Laboratory (EQL) at the UWRL is equipped for analyses of organic and inorganic constituents in air, water, and soil. All of these resources, but most notably the office space, computer facilities, conference rooms, and administrative staff and student offices will be available for use by project researchers for the proposed project.

Computers: Investigators at USU have been working for the past several years on the NSF-supported CUAHSI Hydrologic Information Systems project. The project team already has available computer servers (web servers, database servers, map servers, and virtual server hosts) that are housed at the Utah Water Research Laboratory and are capable of serving as the research platform and development environment for some of the software and middleware cyberinfrastructure systems that will be deployed to support this proposed research.

Function	Model
File Server	Poweredge 1900
Web Servers	Poweredge 700
	Dell Poweredge 1800
	Dell Poweredge 1900
GIS Server	Dell Poweredge 1900
Database Server	Dell Poweredge 2900
Virtual Server Hosts	Dell Poweredge 2800
	Dell Poweredge 2950
	Dell Poweredge 2950
CUAHSI HIS Server Appliance	Dell Poweredge 840

These servers are currently deployed within a machine room at the UWRL that has emergency backup power and redundant chilling systems. This existing infrastructure represents a significant investment in computer hardware and software available to the project.

In addition, the project team has access to existing computer resources at the Utah Water Research Laboratory and Utah State University, including personal computers or laptops for all project personnel. The project team already has access to all of the software and development environments that will be used in the completion of this project. USU IT Services provides both wired and wireless network services to all of USU's campus. The full time IT professionals on staff ensure that the physical infrastructure of USU's network is maintained and that it meets the requirements and applications in support of teaching, research, and the outreach mission of USU. USU IT takes a leadership role for university-wide voice, video, and data systems supporting the delivery of data, online course offerings, and video. USU IT services currently provides 1 GB network connections to offices of each of USU's project personnel and their graduate student offices. USU IT Services maintains dual 10 GB network lines to the rest of the Utah Education Network (UEN) ensuring the USU project personnel are reliably connected to the collaborating project partners.

3. Lamont-Doherty Earth Observatory, Columbia University

The Geoinformatics for Geochemistry team at located at the Lamont-Doherty Earth Observatory (LDEO) of Columbia University on the Lamont campus in Palisades, NY. LDEO is a leading research institution for the Earth Sciences, and a key unit of Columbia's Earth Institute. LDEO provides office space for all GfG team members, meeting space, space for computer hardware, administrative support, grants management and financial services, library and other academic services under the negotiated ICR.

The computer infrastructure of the Geoinformatics for Geochemistry Program is operated and maintained as part of the Integrated Earth Data Applications (IEDA) facility, which is funded by the NSF through a Cooperative Agreement to provide data services for the Earth, Ocean, and Polar Sciences. The IEDA data facility operates with a well-established organizational and technical infrastructure to fulfill these requirements of the Cooperative Agreement, including a highly skilled team with diverse expertise that comprises science experts, data specialists, database developers, and application programmers, and a solid project execution approach that ensures adequate monitoring of progress, assessment of achievements, and risk management. IEDA has a formal advisory structure, consisting of a Policy Committee and a User Committee that provide guidance to the Project Management Team on activities, priorities, and directions that best serve the broad solid-earth and polar geoscience communities.

Servers**:** The IEDA primary systems are currently hosted on a suite of eight rackmount Dell PowerEdge servers (4 x 1950, 2 x 2950, 2 x R310) distributed in 3 buildings in the Marine Geology and Geophysics (MG&G) division on the Lamont campus. The racks are located in server rooms with climate control, voltage-regulated battery-backed power, environmental monitoring, and diesel generators. The servers are protected by SonicWALL 3500 network security appliances, which in turn are connected to the campus fiber backbone and a Juniper NetScreen-208 enterprise firewall, with a redundant connection to the commodity Internet and Internet-2 via the Morningside campus.

Selected systems with specialized needs are hosted with off-site commercial providers including SoftLayer, Linode, and Amazon (CloudFront storage).

This server infrastructure simultaneously supports several other NSF programs at Lamont including Rolling Deck to Repository (R2R), Borehole Research Group (BRG), and the Office of Marine Operations (OMO).

Data Storage*:* The IEDA data and metadata collections are stored in a cluster of relational databases hosted on the four PowerEdge 1950 servers, combined with six rackmount ACNC JetStor disk arrays (2 x 412S, 4 x 516iS) with a total installed capacity of 128TB. A backup copy of selected data is stored on the Lamont Mass Store, an enterprise-class system with an ADIC Scalar i2000 tape library. A backup copy of very large objects (MCS data) is stored at the National Geophysical Data Center (NGDC) in Boulder.

Media Translation*:* A wide range of drives is available, attached to the server cluster in the Oceanography building, including DLT (DLT-4000/7000, SuperDLT 320), IBM 3490, 8mm tape

(Exabyte 82/85/87xx), 4mm tape (DDS-2/3/4/5), removable magnetic disk (DynaMO 2300, Zip/Jaz) and optical disc (CD/DVD+/-RW).

***Server Software**:* The IEDA primary servers run Red Hat Enterprise Linux (RHEL) and kernel-based virtual machines (KVM), in combination with the Nagios enterprise monitoring system and Bacula enterprise backup system. Data and metadata collections are stored in a PostgreSQL relational database management system (RDBMS) with PostGIS geospatial extensions. Primary search and download services are built on the Apache Web server with PHP scripting extensions and the Drupal content management system. The MapServer GIS platform and Apache Tomcat JSP servlet engine are deployed to support advanced Web services. We use the Atlassian enterprise suite (Confluence, JIRA, Crowd) for project management and the Request Tracker (RT) enterprise system to manage external email traffic.

A subset of IEDA systems are currently stored in a Oracle 11g database server with a WebLogic Express JSP servlet engine, during a transition phase into PostgreSQL and PHP.

Data processing software: A wide range of software programs is available at Lamont to process and display data including:

- Landmark ProMAX, Paradigm FOCUS, Seismic Unix (SU), and SioSeis for seismic processing
- Landmark SeisWorks, Paradigm VoxelGeo and SMT Kingdom Suite for seismic interpretation
- ESRI ArcGIS, ERMapper, IVS Fledermaus, MATLAB, IDL/ENVI, GMT, and MB-System for multi-purpose processing and visualization.

4. San Diego Supercomputing Center

The San Diego Supercomputing Center (SDSC) and its Spatial Information Systems Lab maintain facilities and equipment sufficient to support our role in this project. All project personnel have office space and computer workstations, as well as access to office and teleconferencing equipment and networks.

The research group for the proposal is housed in the SDSC building on the UCSD campus. The building co-locates the project team with a wide variety of scientists, developers, and support staff who are pursuing similar research agendas. The physical plant provides access to all essential facilities for a project of this type: conference rooms, teleconferencing facilities including the Access Grid, an international resource for visual telecommunication, and the SDSC/Calit2 Synthesis Center, a national resource for visually based collaboration on research projects. Facilities used for the participants are accessible to participants with disabilities.

The Spatial Information System Lab maintains a series of development and production servers and data storage systems that support the CUAHSI Hydrologic Information System and other projects. The production servers hosting the Hydrologic metadata catalog and services, the hydrologic ontology, portals and databases for CBEO, CZO and other projects, are located in the main SDSC machine room and managed 24/7 by the SDSC production support staff, with regular backup and updates. In addition, critical project servers are mirrored in the machine room of the California Institute for Telecommunication and Information Technology (CalIT2), to ensure high availability of project databases and services maintained by the lab. Several development servers are also maintained. SDSC has been providing collaboration services used in the lab's projects (mailing lists, Wiki, Sharepoint for document sharing, etc.).

IT Infrastructure: The center provides connectivity to the Internet via an internal fiber optic network, with multiple vBNS network connections that can carry approximately 2 Gbps in the aggregate, as well as connectivity to Grid resources through the TeraGrid backbone (max throughput at 30 Gbps). The project will have access to the SDSC storage resource broker (SRB), and the high performance storage system (HPSS) which makes available to the user a 18 Petabytes of tape storage capacity. This resource is available essentially for the cost of tapes. The participants will also have the ability to apply for computational cycles and online storage space through SDSC and TeraGrid allocation services.

The IT support infrastructure at SDSC provides 24/7 production level support, and working day (8/5) help desk/user services support to assist in resolving technical issues. This project will have access to the full expertise and insitutional experience to support state of the art computer equipment, including major servers, database storage devices, personal desktop machines, and notebooks.

Expertise: SDSC has a staff of more than 400 scientists, software developers, and support personnel. SDSC also sponsors several major data projects in biology, and several others in the geo- and environmental sciences. Each of these projects is developing a workbench area for its users. As a result, there is an "intellectual capital" of nearly 50 persons who are currently engaged in the development of cuttting edge resources for web based portals and data distribution. This includes a body of 30 software engineers who are creating production level tools to serve their respective communities. This makes SDSC an ideal environment for the construction and maintenance of proposed infrastructure resources.

High Perfomance Computers: SDSC is a well-established production facility with many computational resources. Most recently, the SDSC has put into production the Triton Resource, a High Performance cluster with 256 nodes, each containing 2 quad-core Intel 2.4 GHz processors and 24 GB memory; the machine offers 20 TeraFlops peak performance. Small allocations of time are made on this machine at no charge to SDSC researchers, and large time allocations are made to SDSC researchers on a recharge basis (0.03 per cpu hour). Other computational resources are available to the project via the NSF national allocation process. Computational time is awarded via a competitive application process, and development accounts are provided on request on a recharge basis. Access is also available to non-production servers where jobs can be deployed on experimental architectures committing them to production.

The Triton Resource at the San Diego Supercomputer Center is a high performance computing system supporting researchers at the University of California. The Triton Resource consists of two subsystems: The Triton Compute Cluster (TCC) and the Petascale Data Analysis Facility (PDAF). Both are supported by a high performance parallel file system and have connectivity to high bandwidth research networks.

The TCC is a 256-node Appro HyperGreen cluster. Each node contains dual quad-core Intel Xeon 5500 (formerly code-named Nehalem) processors running at 2.4 gigaHertz, and 24 gigabytes of main memory. The nodes are interconnected with a 10 gigabit-per-second Myrinet fabric. The TCC has a peak theoretical performance of 20 teraflops and six terabytes of memory.

The PDAF is a unique data-intensive computing resource with one of the largest main memory capacities of any publicly available system. The PDAF comprises 28 Sun x4600M2 symmetric multiprocessing (SMP) nodes. Each node contains eight quad-core AMD 8380 processors running at 2.5 gigaHertz. Twenty of the nodes have 256 gigabytes of main memory per node and eight of the nodes have 512 gigabytes of main memory. The PDAF nodes are interconnected with a 10 gigabit-per-second Myrinet fabric. The PDAF has a peak theoretical performance of nine teraflops, and nine terabytes of memory.

5. Applied Physics Laboratory, University of Washington

The Applied Physics Laboratory (APL) is a major research unit reporting directly to the Provost of the University of Washington and is located on the campus of the University of Washington close to Seattle’s Lake Union waterfront. Our buildings are equipped with extensive machine and electronics shops, CAD/CAM facilities, a library, publication facilities, and special purpose laboratories for microwave remote sensing, image processing, physical acoustics, transducer testing, operational marine data management and visualization facilities, and others. APL operates two vessels of its own. Many of APL’s 300 staff members teach university courses and supervise graduate students. The laboratory’s general computer center provides in-house computing and networking services, as well as access to nationwide networks. A variety of personal computers and servers are available for data acquisition and processing.

Environmental and Information Systems Department (EIS) and computing resources: EIS is a unit with diverse skills and expertise that may be grouped into three core research areas: acoustics and signal processing, environmental sensing and modeling, and data, information and control systems. EIS has substantial experience in the management, analysis and visualization of multi-dimensional environmental data from both observations and models in oceanographic, atmospheric and riverine systems. Data, information and visualization systems are supported through dedicated, high-capacity Linux Ubuntu and Fedora servers backed by redundant RAID arrays with automatic backup of data from files and relational databases. While the server-side environment is Linux based, EIS expertise encompasses Linux/UNIX, Mac OS and Windows platforms. Proprietary software installation and expertise includes Matlab, ESRI ArcGIS, PhotoShop, and many others. Our operational data storage, harvesting, management, and visualization systems rely on open-source software stacks that include: Apache and Tomcat web servers; MySQL and PostgreSQL relational database management systems (the latter including PostGIS spatial extension); PHP, Pyhon, Java, C, Perl, FORTRAN and other languages; support for Open Geospatial Consortium (OGC) standard web services via GeoServer, other community software, and custom software; support for SOAP and REST-based web services (server and client-based); custom, high-load map image tile server for Google Maps and similar systems.

Data Management Plan

Types of Data and Other Materials: Although no primary data collection will be conducted by this proposed project, the purpose of the CZO Integrated Data Management System (CZOData) is to promote collaboration among the CZO sites and sharing of data across CZO sites and with broader scientific communities. As such, the CZOData will support the full data life cycle – providing tools that assist CZO Data Managers in creating and publishing quality controlled data and tools that assist data consumers in discovering, visualizing, and downloading data in which they are interested. CZO data consist of field and laboratory measurements of many types of data: climate, hydrologic, geochemical, spatial, geophysical. Hydrologic and climate data include time series of observations from *in situ* sensors at fixed monitoring locations (e.g., streamflow gages, weather stations, etc.) as well as observations derived from water quality samples. Geochemical data include observations made from the collection and analysis of geochemical samples. CZO sites will also be sharing geospatial datasets describing the CZO sites (e.g., LiDAR, land use, digital elevation models, etc.). Finally, CZO investigators will be sharing derived datasets resulting from specific analyses or modeling exercises using combinations of the above data types. Development of the required repositories, protocols, services, and methods for enabling shared access to the full range of anticipated CZO data types is part of the intellectual contribution of this project.

Training materials, which will be aimed at teaching CZO Data Managers how to organize their data and metadata for publication in the CZO Integrated Data Management System, will be developed and may be of interest to the broader earth science observatory community. These materials will be made available via a publicly accessible CZO website.

Data and Metadata Standards: The CZOData will make full use of existing and emerging standards for sharing environmental data. We will continue development of the CZO Display File format as a cross-CZO standard for publishing data and metadata on individual CZO websites, extending this format to support different data types such as observations derived from geochemical samples. Time series of point observations harvested into the CZO Central System from individual CZO Display Files will be stored using the CUAHSI HIS Observations Data Model and published in Water Markup Language (WaterML) format using CUAHSI HIS WaterOneFlow web services. As a new, international standard for hydrologic data exchange, WaterML 2.0 is being developed through the Hydrology Domain Working Group of the Open Geospatial Consortium (OGC) and the World Meteorological Organization. The web service interfaces will be updated to support WaterML 2.0 and OGC Sensor Observation Service. The WaterML 2.0 specification is expected to be adopted as an international standard by the end of 2011. For geospatial datasets, we will use existing OGC standard interfaces such as Web Map Services, Web Feature Services, and Web Coverage Services. Metadata cataloging services for hydrologic data will use existing CUAHSI HIS Central discovery services. Geochemical data for all CZOs will be harvested from individual CZO Display Files and stored centrally in CZchemDB, which will be regularly harvested into the EarthChem Portal and will be made available via web services in EarthChemXML format. In addition, datasets of different types, as well as metadata display files, will be registered and indexed in the CZO Central portal, to enable cross-site and cross-domain data discovery by common metadata fields, spatial location, and full-text metadata search. This portal will rely on a cross-CZO metadata catalog exposed as OGC Cataloging Services for the Web. By using these accepted standards for interfaces and data encodings, we will ensure that CZO data are interoperable with existing data repositories such as the CUAHSI HIS and EarthChem systems, as well as other standards-compliant data systems in the earth sciences.

Policies for Data and Research Products: Data collection and Quality Assurance/Quality Control are the responsibility of CZO sites and will be conducted in accordance with adopted protocols developed by CZO investigators and Data Managers. The proposed CZOData will ensure that the data are validated against CZO Shared Vocabularies, checked for outliers in the process of harvesting and database ingestion, and that they are accompanied with metadata as defined in data type-specific metadata standards. The primary data dissemination method for the proposed system is web-based. Finalized data sets will be made publicly available and searchable via the data pages at individual CZO sites and via our central portals. For some data sets, the metadata will note the existence of finer temporal resolution data that will not be available online, but may be obtained from the CZO site Data Manager. Datasets are updated as often as possible by CZO Data Managers; continuously collected datasets have an update lag time of 1 month to a maximum of 2 years. All data products published in the CZOData will include appropriate attribute and citation information, and accessing published CZO data products may include agreement to an access/use agreement specified by the author and that is in conformance with established practices and policies already in place among the collaborating Universities and CZO investigators. This will ensure that intellectual rights of CZO investigators/data publishers are protected while granting redistribution rights to the CZOData and its archiving collaborators for purposes of long-term data sharing.

Source code developed by this project will generally be created using an open development model and distributed under the New BSD (BSD 2) open source license. Where feasible and as resources are available, our development team will use open-source code repositories for our common software development, enabling us to coordinate our development activities across multiple Universities, add transparency to our software development, and engage developers and contributors from outside of the immediate project team who may wish to contribute.

Security: Data at individual CZO sites are protected by the internal permissions and protocols used at each institution. Whether they are Linux or Windows systems, each CZO site implements internal network security with a firewall protecting their data files from unauthorized access or manipulation. Source code versioning protocols are used to retrieve earlier versions when necessary. Data files that are to be published will be placed on a webserver for read-only access at each CZO site by a CZO Data Manager. When data is copied to the CZO Central server, it will also be protected by internal network security and firewalls. Incremental backups of both systems will be performed nightly and full backups will be performed every 60 days. The centralized CZO servers will be monitored 24/7 to maintain uptime and reliability, and all patches and software updates will be applied as necessary.

Plans for Archiving Data: CZO investigators and Data Managers will curate and share data and derived data products in the CZO Display File format via the individual CZO websites. Data posted to the individual CZO websites will then be automatically harvested into the CZO Central system. The data harvesters will create an archive copy of the original Display Files, will validate the harvested Display Files against the CZO Shared Vocabularies to ensure that the harvested files are compliant with CZO semantics, and will then parse the data into the appropriate centralized database systems. This ensures that the data are consistently and permanently archived as well as highly available via either the individual CZO websites or via the CZO Central System. We anticipate that curated data products published in the CZOData will be citable for use in peer-reviewed journal articles and conference presentations and proceedings.

Integration with other Cyberinfrastructure: The CZOData will contain a body of knowledge that will be valuable to other research communities. We will implement appropriate interfaces within the Central CZO System that enable the curated CZO data to be indexed by the CI being built by the NSF-funded DataNet projects. For example, we will explore establishing the necessary web service interfaces for the CZOData to become a DataONE Member Node. By creating a DataONE Member Node, the CZOs can participate in the robust, distributed DataONE network and take advantage of the indexing, archival, and discovery services that DataONE provides, broadening the impact of published CZO data resources and encouraging their reuse by a broader scientific community.

National CZO Program

funded by NSF

CRITICAL ZONE OBSERVATORIES

studying the zone where rock meets life

July 20, 2011

Dear Dr. Williams,

After extensive discussions with you and CZO data managers, the board of the CZO PIs is pleased to support your proposal to help further develop the CZO information system. This proposal is a logical and important step in developing an integrated CZO information system.

Therefore individual CZO sites and PI's are more than willing to help this effort in any way possible. We also acknowledge that the success of this effort depends, in part, on CZO scientists and sites working together on shared vocabulary and meta-data standards.

Sincerely,

F.N. Scatena
CZO PI Steering Committee Chair
PI, Luquillo CZO

Jon Chorover
PI, Jemez Santa Catalina CZO

Suzanne Anderson
PI, Boulder Creek CZO

Christopher Duffy
PI, Susquehanna-Shale Hills CZO

Roger Bales
PI, Southern Sierra Nevada CZO

Donald Sparks
PI, Christina River Basin CZO

DRAFT

MEMORANDUM OF UNDERSTANDING
between
NSF OpenTopography Facility
and
NSF Critical Zone Observatories
for
AGREEMENT ON HOSTING AND DISTRIBUTION OF CRITICAL ZONE OBSERVATORIES LIDAR TOPOGRAPHY DATA VIA THE OPENTOPOGRAPHY PORTAL

I. Purpose: The purpose of this Memorandum of Understanding (MOU) is to document and solidify the agreement between the National Science Foundation (NSF)-funded OpenTopography Facility and the NSF-Critical Zone Observatories (CZO) projects, in the hosting, distribution, and Web-based processing of lidar topography data collected by the National Center for Airborne Laser Mapping (NCALM). This document identifies roles and responsibilities for the parties in support of their common objective to provide the greatest impact for these important data through open access to various user communities and the general public.

II. Background: The Critical Zone Observatories, with funding from the U.S. National Science Foundation has partnered with NCALM to collect lidar data covering the six CZO study areas. These high-resolution (~15 points/m^2) data were collected between March 2010 and April 2011. The data collection is approximately 2 terabytes (TB), and consist of classified lidar point cloud data in LAS format plus derived raster data products including bare earth and highest hit surfaces.

The OpenTopography Facility has been funded by NSF to facilitate community access to high-resolution, Earth science-oriented, topography data, and related tools and resources, by leveraging the range of cyberinfrastructure resources available at the San Diego Supercomputer Center (SDSC). OpenTopography has developed a highly scalable cyberinfrastructure-based system that delivers community LiDAR datasets using a tiered approach. At the core of this system is the capacity to host and provide Web-based access to full LiDAR point cloud datasets, user-defined queries based on an area of interest, on-the-fly generation of custom DEMs and derived products (e.g. slope maps), and visualization products. OpenTopography also provides download of pre-computed, ("standard"), DEM data, and access to network-linked LiDAR derived imagery within Google Earth. OpenTopography currently hosts approximately 78 billion lidar returns covering 16,000+ square kilometers and serves several thousand users.

III. Responsibilities: With this memorandum of understanding, OpenTopography and the Critical Zone Observatories agree to the following:

1. OpenTopography will host and distribute final versions of NCALM data deliverables received by the CZOs, including the project metadata, classified point cloud in LAS format, and gridded surfaces (DEMs).

CZO	Data Products	Release?	PI Signature
Boulder Creek	Classified Pt Cloud (LAS)	Y	
	Bare Earth DEM	Y	
	Highest Hit DEM	Y	
Christina River	Classified Pt Cloud (LAS)	Y	
	Bare Earth DEM	Y	

	Highest Hit DEM	Y	
Jemez-Catalina	Classified Pt Cloud (LAS)	Y	
	Bare Earth DEM	Y	
	Highest Hit DEM	Y	
Luquillo	Classified Pt Cloud (LAS)	Y	
	Bare Earth DEM	Y	
	Highest Hit DEM	Y	
Southern Sierra	Classified Pt Cloud (LAS)	Y	
	Bare Earth DEM	Y	
	Highest Hit DEM	Y	
Shale Hills	Classified Pt Cloud (LAS)	Y	
	Bare Earth DEM	Y	
	Highest Hit DEM	Y	

2. OpenTopography will work with the CZO PIs to ensure that data products are posted and approved for final release within one month of the arrival at SDSC.
3. OpenTopography provides no additional data QA/QC or additional processing of data delivered by NCALM, and assumes the data products delivered to be final. The CZOs or the data provider will address problems with the data set should any arise.
4. OpenTopography will provide access to CZO lidar usage statistics as collected by the OpenTopography system. CZO PIs and data managers will be granted login-based access to an OpenTopography statistics "dashboard", enabling them view information on users of the CZO data.
5. OpenTopography will ensure proper citation of the CZOs and the their funders on portal web pages and in metadata associated with these data sets. OpenTopography will also provide access to original data processing and acceptance reports delivered with the CZO data.
6. OpenTopography's multi-year, proposal- based funding prevents us from guaranteeing long-term digital data archiving. However, OT maintains a robust data backup policy to ensure preservation of data over the shorter term. OT is also in conversation with long-term digital data archive activities at SDSC to ensure that OT-hosted data will be preserved should the NSF no longer fund the facility. We recommend that the CZOs maintain an independent long-term data preservation strategy.
7. Large manual data transfers (e.g., via hard drive) of CZO data will be handled by the CZOs. OpenTopography's sole responsibility is for Web-based data downloads via the OpenTopography portal, and every effort is made to provide interfaces to access the data in bulk when required.
8. Through this agreement, the CZOs acknowledge that they are benefiting from significant NSF investment in OpenTopography's cyberinfrastructure, as well as resources at the San Diego Supercomputer Center at University of California, San Diego. These include:
    > Disk space and server resources
    > Double tape backup
    > On-going disk and server costs
    > Unlimited Internet2 SDSC Bandwidth
    > 24 x 7 live system monitoring
    > OT Help Desk support & Tutorials

> OT Discussion forums and community
> Portal workspace for users with custom lidar jobs archive, job stats, user privileges

9. The primary point of contact at OpenTopography for interactions with the CZOs will be Christopher Crosby, OpenTopography Facility Manager.

IV. Period of Agreement:

A. Effective Period

This MOU becomes effective on the date of the last signature.

B. Review Cycle

OT and the CZOs will meet at the request of either party, to review this MOU and to make possible modification, as appropriate.

C. Modifying or Terminating Clauses

The agreement may be modified at any time by mutual consent. All modifications to this agreement will be incorporated as written amendments to the agreement. The agreement may be terminated by either group after a sixty (60) day notification of intent to terminate. If either party terminates this agreement, all work in progress will be completed.

V. Points of Contact:

OpenTopography:

Chaitan Baru
OpenTopography PI
San Diego Supercomputer Center
University of California, San Diego, MC 0505
9500 Gilman Drive
La Jolla, CA 92093-0505
baru@sdsc.edu

Christopher Crosby
OpenTopography Facility Manager
San Diego Supercomputer Center
University of California, San Diego, MC 0505
9500 Gilman Drive
La Jolla, CA 92093-0505
858.822.5458
ccrosby@sdsc.edu

J Ramón Arrowsmith
OpenTopography Co-I
School of Earth and Space Exploration
Arizona State University
Tempe, AZ 85287-1404, U.S.A.
(480) 965-5081

ramon.arrowsmith@asu.edu

II.

CZOs:

VI. Approvals

Signed:
Chaitanya Baru, OpenTopography PI

Date:

Signed:
XXXX, , XXX

Date:

Integrated Data Management System for Critical Zone Observatories
Utah State University

Statement of Work

Utah State University (USU) will assist in the design and implementation of the Critical Zone Observatory (CZO) Integrated Data Management System. More specifically, USU will complete the following tasks:

1. Assist in the overall design of the architecture for and implementation of the CZO Integrated Data Management System
2. Provide coordination between the proposed project and the existing CUAHSI Hydrologic Information System project
3. Host, populate, and perform maintenance and upgrades to the CZO Shared Vocabulary System
4. Assist in the identification and documentation of metadata requirements for CZO datasets
5. Assist in design of templates to be used by CZO Data Managers for creation of metadata describing CZO datasets
6. Assist in the design and development of tools to support the creation and publication of display files by CZO Data Managers
7. Assist in the development of training materials for CZO Data Managers and attend and assist with CZO Data Manager trainings and workshops
8. Assist in developing an information management working group for CZO Data Managers
9. Attend at least one annual project meeting and annual data manager's meetings

Plan and Timeline

Requested funds will be used throughout years one and two to support the work described above. Hosting and maintenance of the CZO Shared Vocabulary System will continue in years one and two. Work toward identification and documentation of metadata requirements for CZO datasets will be completed in year one. Work on the design of templates for metadata creation and tools to support the creation and publication of display files by CZO Data Managers will begin in year one and extend to year two. It is anticipated that training materials for CZO Data Managers will be developed as these tools become available. It is also anticipated that the information management working group for CZO Data Managers will be established near the beginning of the proposed project and will be supported throughout years one and two.

STATEMENT OF WORK

GfG Contributions to the Development of CZO Data Management for Geochemical and Sample Data

The Geoinformatics for Geochemistry (GfG) Program at Columbia University will contribute to the development of CZOData by developing data systems and services that support the management of geochemical and sample data. The GfG Program is part of the NSF-funded Integrated Earth Data Applications (IEDA) facility (http://www.iedadata.org). GfG develops and maintains geochemical information systems and repositories, such as the EarthChem data network and the Geochemical Resource Library, and operates the System for Earth Sample Registration SESAR that provides registration and identification services for geological samples. Since 2005, GfG Director Kerstin Lehnert has participated in discussions and planning efforts of the CZO Data Management System, attending various CZO meetings and workshops. She has provided guidance and advice for the management of sample-based and geochemical data to CZO investigators, and worked with the CZO PI Dr. S. Brantley and her group at Pennsylvania State University on the development of the CZchemDB database.

With funding provided through this subcontract, we will complete a suite of developments as described below that are aimed to (a) ensure proper formatting and documentation of geochemical data for preservation, display, and inclusion in the CZO Portal, aligned with community-based data reporting standards, (b) enhance access to CZO geochemical data by developing a web-based search interface for CZchemDB and incorporating CZchemDB data in the EarthChem portal, (c) ensure registration and unique identification of CZO samples by developing the CZO Registration Agent in the International Geo Sample Number (IGSN) system.

1) Develop Templates for Display of Geochemical data and Metadata

We will develop spreadsheet templates that allow investigators to post geochemical data on the web sites of CZO sites and to submit the data to CZchemDB for inclusion in this synthesis database. The templates should also allow easy submission of the data to the CZO Portal. We will design the templates to include all metadata that are relevant for proper evaluation of data quality and later re-use of the data, including information about the analyzed samples and about the analytical procedure, precision, and reproducibility. The GfG Program has been leading initiatives such as the Editors' Roundtable that established recommendations for the reporting of geochemical data, and we will ensure that the geochemical data templates of the CZO will comply with these recommendations.

Together with our collaborator Dr. S. Brantley at Pennsylvania State University, we will hold a 1-day workshop for CZO geochemists to discuss the design and usability of the templates and ensure the capture of CZO specific metadata.

2) Develop Web-based User Interface to CZchemDB

Currently, CZchemDB is not web-enabled and can only be accessed locally at PSU. We will develop an interactive, web-based interface to CZchemDB, similar to the interfaces that we developed for other GfG databases such as EarthChem, PetDB, SedDB, and VentDB 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, 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.

3) Maintenance of CZChemDB

We will work with the CZchemDB group at PSU to ensure that new data gets added to the CZchemDB, that these data get posted at the CZO sites' web sites, and is properly integrated at the CZO Portal.

4) Integrate CZChemDB Data into the EarthChem system

In order to enhance discovery of CZO geochemical data as part of a much broader geochemical dataset for other sample types, we will incorporate the CZO geochemical data into the EarthChem Portal. The EarthChem Portal serves as a "one-stop-shop" for geochemical data that gives users the ability to search for data across federated databases and obtain data from these databases integrated into a common output format. By including CZO geochemical data in the EarthChem Portal, the data be integrated with and compared to spatially or thematically related data.

The EarthChem Portal accepts and publishes geochemical data in a standard XML format (EXtensible Markup Language). We will work with the PSU group to develop scripts and procedures to convert the CZchemDB data into XML documents, so that the rock and regolith chemistry data can be seamlessly incorporated into the EarthChem system. The conversion of CZchemDB data into EarthChemXML documents requires that vocabularies used in CZchemDB are aligned with those in EarthChemXML, for example lithologies, geographical names, journal names, etc. We will update EarthChem vocabularies to include terminology used by CZchemDB, and ensure that CZchemDB uses existing vocabulary in EarthChem if applicable. We will also try to align these vocabularies with the CZOData shared vocabularies as far as possible.

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. Similar web services have already been developed for the PetDB and SedDB data collections, but will need to be modified as EarthChem is going to rebuild the EarthChem Portal site using a service-oriented architecture to provide a web-service aspect to all search and output functions, responding to the increasing demands for interoperability. The new web-service based architecture will offer a superior range of options and possibilities for interaction and data exchange with other developers and systems.

5) Develop the CZO Registration Agent for the IGSN

At the 2010 CZO data meeting in Boulder, consensus was reached to use the International Geosample Number (IGSN) to uniquely identify samples collected as part of the CZO program. The IGSN is a global unique identifier for samples that allows distributed data for individual samples generated in different labs and published in different papers to be unambiguously linked and integrated. The IGSN was developed by the GfG group (NSF-EAR 05-52123), which operates the System for Earth Sample Registration (SESAR) (http://www.geosamples.org) as the registry and metadata catalog for Earth samples. Registration of samples in the IGSN system also allows to track relationships between samples and sub-samples, for example, samples collected from a core as a depth series, or mineral concentrates separated from a soil specimen.

CZO investigators Ilya Zaslavsky and Tom Whitenack participated in the most recent SESAR workshop (February 2011) that refined the existing SESAR architecture to become a modular and scalable approach. In the new architecture, the IGSN Registry is separated from a central Sample Metadata Clearinghouse (SESAR), and 'Local Registration Agents' have been introduced that provide registration services to specific communities, including tools for metadata submission, management, and archiving. This new architecture will be developed over the next year.

In order to address the CZO specific requirements for sample registration, we will build the CZO Registration Agent that will provide registration and metadata management services as requested and defined by the CZO community. We will work with the CZO community to create CZO specific metadata profiles for all objects that the CZO data management group wants to register with the IGSN, and to define controlled vocabularies as needed. We will also develop web services for CZO sites to register objects from local client systems.

We will develop manuals to help investigators or Site Information Managers to register their samples.

Integrated Data Management System for Critical Zone Obervatories

Statement of Work for UCSD Subcontract

Purpose: As a participant in the proposal titled "**Integrated Data Management System for Critical Zone Obervatories "** being proposed to NSF, UCSD will conduct research and development activities focused on the development of central CZO data harvesting, archiving and dissemination system, and supporting infrastructure and protocols. This work is contingent upon the proposal being accepted for funding.

Period of Performance: 2 years (October 2011 – September 2013)

UCSD Personnel:
Ilya Zaslavsky (PI), Thomas Whitenack (Programmer/Analyst)

Specific tasks include:

- 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. 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. Ensure that the data are periodically re-harvested, and the information provided via services is appropriately updated.
- In coordination with CZO PIs and project leadership team, 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.
- In coordination with the USU team, 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.
- In collaboration with the other project teams, develop solutions for registering CZO-collected data of other data types (not time series). This includes, primarily, geochemical samples and spatial data layers, and can be extended based on project priorities.
- Ensure that the harvested data become available via standards-based service interfaces, and update the interfaces when new standards are adopted.
- Host CZO online content management system, and, in collaboration with the UC-Boulder team, integrate it with the CZO data discovery portal.

- 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.
- Conduct workshops for data managers, and establish continuous exchange of ideas, technologies and innovations through regular phone/web conferences and one-on-one communication.
- Work with the OpenTopography project to utilize the LiDAR project resources for archiving and processing of CZO LiDAR data, and making them available via the CZO data portal.
- 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.
- Enhance the CZO online system that assists data managers in registering and managing their observation networks (password protected)
- Maintain a document and file sharing portal for data managers (password protected).
- Assist the Columbia University team in development of standards-based web services for geochemical data and registering them to the CZO data portal
- Work with the Columbia University team to develop protocols for posting of site geochemistry data and metadata on site web pages, to be harvested by the central CZO system.
- Work with the University of Washington team to adapt the web-based time series discovery client for CZO, adjust it to working with the CZO central data repository, and integrate it with the CZO data portal
- Develop code to automatically register the harvested display files to the online catalog, and make them downloadable via a simple user interface.
- Integrate CZO data portal with other available assets, including CUAHSI HIS data, and, via CUAHSI, data from USGS, EPA and other relevant federal repositories

Statement of Work

Integrated Data Management System for Critical Zone Observatories

Emilio Mayorga, mayorga@apl.washington.edu

University of Washington Applied Physics Laboratory

July 15, 2011

The Applied Physics Laboratory (APL) of the University of Washington will develop an online data access and visualization tool for CZO (Critical Zone Observatories) data that is interactive, feature-rich, user-friendly and responsive to CZO needs. APL will deploy this capability by adapting and extending the regional data integration and visualization system used by the Northwest Association of Networked Ocean Observing Systems (NANOOS, http://www.nanoos.org). The NANOOS Visualization System (NVS, http://nvs.nanoos.org) was developed by NANOOS partners led by APL, with data-integration, web-services and user-interface development led by the APL senior personnel in the proposed project (Mayorga and Tanner). NVS has undergone rapid evolution since its public release in November 2009 (Mayorga, E., T. Tanner, et al., 2010, The NANOOS Visualization System (NVS): Lessons learned in data aggregation, management and reuse, for a user application, *Proc. MTS/IEEE Oceans'10*) and has received considerable attention within the U.S. Integrated Ocean Observing System (IOOS) community for its user-friendly yet rich interface, its rapid development pace, and its consistent data presentation that is independent of the original data source.

The CZO network has developed the initial phase of a data cataloguing, integration and distribution system focusing on sensor-based time series data from all CZO sites; the CZO Central system (http://central.criticalzone.org) is hosted at the San Diego Supercomputer Center (SDSC). The CZO Integrated Data Management System project will greatly expand these capabilities at SDSC and will advance the development of similar capabilities for sample-based geochemical data integration and distribution based at the Columbia University Lamont-Doherty Earth Observatory (LDEO).

APL will develop the "CZO VS" in close collaboration with SDSC, LDEO and the CZO network. The CZO VS will focus on ingestion of time series data generated by CZO sites and distributed via the SDSC system using CUAHSI HIS-based web services (http://his.cuahsi.org) and Open Geospatial Consortium (OGC) standard web services. The inclusion of CZO geochemical data aggregated by LDEO will be assessed and included only if the LDEO system is sufficiently advanced, but is likely to only encompass catalog and metadata information. The CZO VS system will be deployed and maintained on APL servers, with optimization schemes developed jointly with SDSC.

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. The capability to search across CZO sites and explore the resulting network-side data will be developed jointly with SDSC and its CZO GeoPortal catalog, in parallel with the site-focused approach of the CZO VS. 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 ontologies implemented by SDSC (e.g., measurement variables, platform types, asset types). 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. This functionality will represent a substantial component of the APL contribution to this project that will involve both user-interface development by APL and data-discovery optimizations developed and implemented jointly with SDSC.

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 (http://hiscentral.cuahsi.org), 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 SDSC. Other application functionality may be developed for the CZO VS depending on a needs-assessment and prioritization with the CZO network, and available resources. These include a generic capability for generating depth vs. value "profile" plots (e.g., soil moisture with depth); and the ability to select and download asset data for multiple sites and multiple variables through one focused search.