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 13-1

NSF 13-525 **03/19/13**

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

OCI - Software Institutes

FOR NSF USE ONLY

NSF PROPOSAL NUMBER

1339834

DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	DUNS# (Data Universal Numbering System)	FILE LOCATION
03/19/2013	**1**	**01110000 OCI**	**8004**	**047918474**	**03/19/2013 4:51pm**

EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN)

522081073

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

Stroud Water Research Center

AWARDEE ORGANIZATION CODE (IF KNOWN)

5300012638

ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE

Stroud Water Research Center
970 Spencer Road
Avondale, PA. 19311

NAME OF PRIMARY PLACE OF PERF

Stroud Water Research Center

ADDRESS OF PRIMARY PLACE OF PERF, INCLUDING 9 DIGIT ZIP CODE

Stroud Water Research Center
970 Spencer Road
Avondale ,PA ,193119514 ,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 **Collaborative Research: SI2-SSI: The community-driven BiG CZ software system for integration and analysis of bio- and geoscience data in the critical zone**

REQUESTED AMOUNT	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PRELIMINARY PROPOSAL NO. IF APPLICABLE
$ **2,949,985**	**48** months	**12/15/13**	

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)
_____

PI/PD DEPARTMENT
Stroud Water Research Center

PI/PD FAX NUMBER
610-268-0490

PI/PD POSTAL ADDRESS
970 Spencer Road

Avondale, PA 19311
United States

NAMES (TYPED)	High Degree	Yr of Degree	Telephone Number	Email Address
PI/PD NAME **Anthony K Aufdenkampe**	**PhD**	**2002**	**610-268-2153**	**aufdenkampe@stroudcenter.org**
CO-PI/PD **Robert M Cheetham**	**MLA**	**1997**	**215-701-7713**	**cheetham@azavea.com**
CO-PI/PD **Jeffery S Horsburgh**	**PhD**	**2009**	**435-797-2946**	**jeff.horsburgh@usu.edu**
CO-PI/PD **Kerstin A Lehnert**	**PhD**	**1989**	**845-365-8506**	**lehnert@ldeo.columbia.edu**
CO-PI/PD **Emilio Mayorga**	**DPhil**	**2004**	**206-543-6431**	**mayorga@apl.washington.edu**

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 13-1

NSF 13-525 **03/19/13**

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

OCI - Software Institutes

FOR NSF USE ONLY

NSF PROPOSAL NUMBER

1339793

DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	DUNS# (Data Universal Numbering System)	FILE LOCATION
03/19/2013	1	01110000 OCI	8004	804355790	03/19/2013 4:52pm

EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN)

956006144

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 California-San Diego

AWARDEE ORGANIZATION CODE (IF KNOWN)

0013177000

ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE

Office of Contract & Grant Admin
9500 Gilman Drive, 0934
La Jolla, CA 92093-0934

NAME OF PRIMARY PLACE OF PERF

University of California-San Diego

ADDRESS OF PRIMARY PLACE OF PERF, INCLUDING 9 DIGIT ZIP CODE

University of California-San Diego
9500 Gilman Drive MC 0934
La Jolla ,CA ,920930934 ,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 **Collaborative Research: SI2-SSI: The Community-Driven Big-CZ Software System for Integration and Analysis of Bio-Geoscience Data in the Critical Zone**

REQUESTED AMOUNT	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PRELIMINARY PROPOSAL NO. IF APPLICABLE
$ 1,050,015	48 months	12/15/13	

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)
_____

PI/PD DEPARTMENT
San Diego Supercomputer Center

PI/PD POSTAL ADDRESS
9500 Gilman Dr., MC 0505
La Jolla, CA 920930517
United States

PI/PD FAX NUMBER
858-534-5113

NAMES (TYPED)	High Degree	Yr of Degree	Telephone Number	Email Address
PI/PD NAME **Ilya Zaslavsky**	**PhD**	**1995**	**858-534-8342**	**zaslavsk@sdsc.edu**
CO-PI/PD **David Valentine**	**PhD**	**2004**	**858-534-4896**	**valentin@sdsc.edu**
CO-PI/PD				
CO-PI/PD				
CO-PI/PD				

PROJECT SUMMARY

Overview:

Critical zone (CZ) scientists from 16 bio- and geoscience disciplines are at the front line of efforts to effectively compile and use the "dark data in the long tail" of Earth science and integrate it with the "Big Data". As such, many see the CZ data effort as a pilot for the EarthCube endeavor. This proposal is motivated by strong and nearly unanimous cyber-needs expressed at a CZ-EarthCube workshop in Jan. 2013 (Award #1332257). Our overall goal is to co-develop with the critical zone science community a high-performance web-based integration and visualization environment for joint analysis of cross-scale bio and geo processes in the critical zone (BiG CZ), spanning experimental and observational designs.

Project Objectives and Deliverables are to: (1) Engage the critical zone and broader communities to co-develop the BiG CZ software stack to meet their specific needs through a CZ community advisory committee and a series of two co-design workshops, four training & testing workshops, one CZ science synthesis institute and one workshop for citizen scientists and resource managers; (2) Develop the BiG CZ Portal web app for high-performance map-based discovery, visualization, access and publication of data on critical zone structure and function; (3) Develop the BiG CZ Toolbox to enable cyber-savy CZ scientists to directly access BiG CZ APIs to search, access, manage and publish data using a turnkey open-source scripting and database package; and (4) Develop the BiG CZ Central software stack as the backend for Portal and Toolbox functionality for single query discovery and access to data systems developed for multiple bio- and geo-science domains, including CUAHSI Hydrologic Information System (HIS), Integrated Earth Data Applications (IEDA), the System for Earth Sample Registration (SESAR), DOE's System Biology Knowledgebase (KBase), Ecoinformatics MetCat holdings including all Long Term Environmental Research (LTER) data, DataOne, OpenTopogaphy, and others.

Intellectual Merit :

The CZ science community has identified that the central scientific challenge of the critical zone science community is to develop a grand unifying theory of the critical zone through a theory-model-data fusion approach to answer: (1) how do tectonics, lithology, climate and biology co-determine the evolution of critical zone structure and function?; (2) What are the drivers of energy and material fluxes (i.e. water, sediment, carbon, nutrients, solutes, etc) moving through the critical zone?; (3) How will critical zone structure, function and evolution respond to human and natural disturbances and over various time and spatial scales? CZ-EarthCube participants also unanimously envisioned -- independently in each breakout group -- that software similar to what we propose would enable such transformative understanding.

We will develop the BiG CZ software system upon widely-used, high performance, open source, multi-platform software components. We believe that an engaging user interface with very fast response times is a fundamental requirement for user satisfaction and therefore software sustainability.

Broader Impacts :

We believe that a key aspect of increasing the computational capabilities of the CZ community requires increasing the computational skills of the CZ community. We will partner with the Software Carpentry project to teach 80 early-career Earth and environmental scientists computational science skills through four Bootcamps and one CZ Science Synthesis Institute, which will serve the dual purpose of testing and evaluating the BiG CZ system.

From our previous experience with the Model My Watershed project (Award #0929763), we believe that the BiG CZ Portal can hold exceptional value toward engaging resource managers, citizen scientists and even secondary school teachers and students. We will thus hold a workshop for resource managers and citizen scientists to introduce them to the deliverables outlined above and to solicit feedback on the usability of those tools to meet their needs. We will also develop curricular materials around the portal designed specifically for secondary teachers to assist them to integrate research data into their classrooms.

TABLE OF CONTENTS

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

*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.

SI2-SSI: The community-driven BiG CZ software system for integration and analysis of bio- and geoscience data in the critical zone

C. PROJECT DESCRIPTION

C.1. Scientific Motivation and Software Needs

C.1.1. Needs of Critical Zone Science Community

The Critical Zone (CZ) scientist community takes as their charge the effort to integrate theory, models and data from the multitude of disciplines collectively studying processes on the Earth's surface. The Critical Zone is Earth's permeable near-surface layer – from the atmosphere at the vegetation's canopy to the lower boundary of actively circulating groundwaters. The Critical Zone was a term coined by the National Research Council's *Basic Research Opportunities in the Earth Sciences (BROES)* Report (2001) to highlight the imperative for a new approach to massively multi-disciplinary research on the zone of the Earth's surface that is critical to sustaining terrestrial life on our planet. The National Science Foundation responded to that call by launching the national Critical Zone Observatory (CZO) program in 2007 (http://criticalzone.org/), which stimulated the development of an international network of CZOs (http://www.czen.org/). Today the term, concept and approach of "critical zone" science has firmly taken hold in the broad geoscience community, with several CZ sessions at each of the major annual geoscience meetings.

Critical zone scientists – from 16 bio- and geoscience disciplines – are at the front line of efforts to effectively compile and use the "dark data in the long tail" of Earth science (Heidorn 2009) and integrate that data with the "Big Data" produced by hydrologists, atmospheric scientists, geospatial modelers and molecular biologists. The CZ community has begun the process of building a shared community data system with the recent "*Integrated Data Management System for Critical Zone Observatories*" project (CZOData, NSF Award #1332257, with all PI/coPIs also participating in this BiG CZ SSI proposal), and many see the CZ data effort as a pilot for the EarthCube endeavor (http://earthcube.ning.com).

In January 2013, 103 members of the CZ community met for the CZ-EarthCube Domain Workshop to prioritize the CZ community's key science drivers, key cyber challenges and key cyber needs (NSF Award #1252238, http://criticalzone.org/national/events/event/2013-01-21-earthcube-czo-workshop/, CZ-EarthCube workshop Executive Summary Draft). Their response was strong and nearly unanimous. They identified that the central scientific challenge of the critical zone science community is to develop a "grand unifying theory" of the critical zone through a theory-model-data fusion approach to answer: (1) how do tectonics, lithology, climate and biology co-determine the evolution of critical zone structure and function?; (2) What are the drivers of energy and material fluxes (i.e. water, sediment, carbon, nutrients, solutes, etc) moving through the critical zone?; (3) How will critical zone structure, function and evolution respond to human and natural disturbances and over various time and spatial scales? Workshop participants also unanimously envisioned -- independently in each breakout group -- that a key need to enable such transformative understanding was a future cyberinfrastructure for seamless 4D visual exploration of the knowledge (data, model outputs and interpolations) from **all** the bio and geoscience disciplines relevant to critical zone structure and function, similar to today's ability to easily explore historical satellite imagery and photographs of the earth's surface using Google Earth.

Although map-based data discovery portals do exist for some geoscience domains, none come close to serving as a "one-stop-shop" for datasets from the breadth of disciplines and multitude of data providers needed by CZ scientists. Furthermore, for many of the most important CZ dataset types -- such as soil geochemistry or near surface geophysics -- no national repository exists to even catalog datafiles, let alone integrate granularly at data field and data value levels. In other cases, such as for microbiology or molecular biology, highly-developed cyberinfrastructures exist, but those databases do not have a means to connect that data to environmental settings through precise sampling location and time or to their geochemical and mineralogical matrix. Therefore, the CZ community needs a top-to-bottom software stack that provides an extremely powerful yet easy-to-

use web user interface as a client to a high-performance back-end that enables interoperability of data search, access and publication across a variety of bioscience and geoscience domains, across a large range of spatial and temporal scales, and across observational and experimental designs.

C.1.2. Vision for the Proposed BiG CZ Software System

Our **Overall Goal** is to co-develop with the critical zone science community a web-based integration and visualization environment for joint analysis of cross-scale bio and geoscience processes in the critical zone (BiG CZ), spanning experimental and observational designs. This BiG CZ software stack would consist of the BiG CZ Portal web application, the BiG CZ Toolbox, and the BiG CZ Central software infrastructure that together enable interoperability for cross-scale multi-modal discovery, visualization, access, and publication of data through a simple yet powerful user interface and open APIs. Developing such a cyberinfrastructure would not only meet pressing multi-disciplinary science needs, but will also present interesting yet achievable software challenges to bridge the barriers that currently exist between science domains, large ranges in spatial and temporal scales and observational versus experimental

We organize our efforts toward our overall goal through four **Project Objectives**:

Objective 1. Engage the CZ and broader community to co-develop and deploy the BiG CZ software stack to meet their specific needs through a CZ community advisory committee and a series of co-design and training & testing workshops.

Objective 2. Develop the BiG CZ Portal web application for intuitive, high-performance map-based discovery, visualization, access and publication of data on critical zone structure and function by scientists, resource managers, citizen-science volunteers, and the general public.

Objective 3. Develop the BiG CZ Toolbox to enable cyber-savy CZ scientists to directly access BiG CZ APIs to search, access, manage and publish data using a turnkey open-source scripting and database package.

Objective 4. Develop the BiG CZ Central software stack to bridge data systems developed for multiple critical zone domains into a single metadata catalog for single query search, visualization and access from the Portal web app or using more powerful APIs, and to serve as a public repository for data published via the Toolbox or via web forms.

We will develop all of the proposed software on public repositories as a modular suite of fully open source software projects, building upon widely used existing free and open source software (FOSS). As such, all of our software components could be adapted or forked to serve other uses or projects, on any operating system. However, we will focus our efforts on developing a single Portal web application and Central software stack on an Openstack cloud in order to maximize search, access and visualization performance. The Toolbox will, on the other hand, be an easy-to-install, multi-platform (Windows, Mac OSX, Linux) package for local data management prior to publishing to the Central repositories and for local access to Central search, access and visualization functionality through Application Programming Interfaces (APIs).

The entire BiG CZ Software system will be built around a new information model that our team is currently designing with the NSF-funded project "*Developing a Community Information Model and Supporting Software to Extend Interoperability of Sensor and Sample Based Earth Observations*" project (Award #1224638, with all PI/coPIs also in the BiG CZ SSI). This new information model, which we call **Observations Data Model Version 2.0** (ODM2), builds upon the extensive experiences from the CUAHSI Hydrologic Information Systems (HIS) and their ODM 1.1 (Horsburgh et al. 2008), the Integrated Earth Data Applications (IEDA) databases such as SedDB, PetDB and EarthChem (Lehnert et al. 2000), the Integrated Ocean Observing System's (IOOS) NANOOS Visualization System (NVS) (Mayorga et al. 2010), and the recent efforts of the "*Integrated Data Management System for Critical Zone Observatories*" (CZOData) project (NSF Award #1332257, Zaslavsky et al. 2011). In short, we are applying lessons learned from these diverse experiences to design ODM2 to enable fully granular integration of data and rich-metadata from sensor-based and sample-based observations of numerical and categorical data types at sampling point locations, along sampling curves (1D profiles and transects), along sampling surfaces (2D coverages) and within sampling volumes (3D). ODM2 profile of Open Geospatial Consortium's Observations and

Measurements (OGC O&M) standard will have broad applicability to information exchange with central data systems such as those used by DataOne and the Ecoinformatics.org community. From concept to several working prototypes, the ODM2 will include encodings for metadata catalogs, relational database management systems, and for dataset archival functions. The ODM2 project is a third of its way through a 2 year timeline (Aug. 2012 to Jul. 2014), has held its first community engagement workshop, and is beginning prototype development now. We propose for the BiG CZ SSI is that ODM2 prototypes implementations become sustainable, enterprise class software that would form the backend foundation for CZO and other communities. We will develop into the BiG CZ Portal, Toolbox and Central software stacks to provide a standards-based cross-domain interoperability functionality, and user interfaces that address the needs of the CZO community.

Our vision for integrating successful, existing software elements into a new, sustainable framework that addresses clearly articulated needs of the CZ research community aligns well with NSF's vision of a Cyberinfrastructure Framework for 21st Century Science and Engineering (NSF CIF21, 2012).

C.1.3. Project Deliverables

Our vision can be stated as a number of **Project Deliverables** for each Objective:

Objective 1 Deliverables: Community Engagement in Software Design.

- **CZ Community Advisory Board** (CZCAB), to advise feature development priorities, provide feedback on releases, recruit workshop participants and promote the BiG CZ software system.
- **Co-design workshops**, two in year 1, to prioritize requested functionality and give specific feedback on our detailed plans and mockups for the BiG CZ software stack.
- **3-day BootCamp Training & Testing workshops**, four total in years 2-4 for 20 attendees each, both to elevate the overall computational capabilities of CZ scientists while teaching them to be power users of the BiG CZ system (Portal, Toolbox, APIs), and also to provide feedback on usability and desired new features. Modeled after Software Carpentry hands-on bootcamps (http://software-carpentry.org/bootcamps/), the BiG CZ will include a 3rd day focused on the BiG CZ usage. We will target students, postdocs, and early-career investigators.
- **Citizen-science & Resource Manager workshop**, one in late year 3, to evaluate the value of our integrated earth surface data and visualization system to professionals and volunteers who manage and monitor natural resources.
- **7-day CZ Science Synthesis Institute**, one in early in year 4 and immediately after a 3-day BootCamp. A cohort of 14-16 BootCamp graduates will be mentored by 3-6 senior CZ investigators and 3-6 project team members to utilize their newly developed skills with the BiG CZ software system address 2-3 high-priority synthesis objectives identified through the earlier community engagement. This institute is based on the very successful model developed in NSF's exploratory Hydrologic Synthesis program (Thompson et al., 2012), and will serve as the ultimate "Training and Testing" evaluation of the BiG CZ software system.

Objective 2 Deliverables: The BiG CZ Portal web application for high-performance map-based discovery, visualization, access and publication of data on critical zone structure and function.

- The **BiG CZ Portal web application client to the BiG CZ Central** system would enable a user to zoom to any location in the continuous USA (i.e. lower 48) to search, view, filter, select and download heterogeneous datasets, including:
  - Points with sensor and sample based observational data, with pop-ups showing 2D data series displays (i.e. time series or depth profiles) and metadata. This functionality will be based on the software behind the NANOOS Visualization System (NVS, http://nvs.nanoos.org);
  - 2D satellite and GIS imagery from many different agencies and sources. This functionality will be based on GeoTrellis ((http://www.azavea.com/products/geotrellis/), which forms the foundation of Model My Watershed (http://app.wikiwatershed.org/stroud/guest);

- 2D and 3D interactive visualization of select datasets, such as geophysical “fenceline” images of subsurface structure, topography (i.e. similar to Google Earth) or canopy structure from LiDAR point clouds. Interactive visuzalization functionalities will be based on a number of open source libraries, such as the Cesium WebGL Vitual Globe and Map Engine javascript library (http://cesium.agi.com/), VisIt (http://en.wikipedia.org/wiki/VisIt and https://wci.llnl.gov/codes/visit/home.html), and the Data Driven Document javascript library (http://d3js.org/).

- **Map views will be filterable** by a time period, variable, medium, data provider, data creators (authors) and a large number of other parameters.
- **A suite of data publication/registration forms** will be integrated into the BiG CZ Portal web app, to assist “long tail” scientists to tag, publish and share their “dark data”. These forms will:
  - be readily accessible from the map-based discovery & visualization system through various buttons and quick-links;
  - allow users to map file-level and field-level metadata to the BiG CZ shared vocabulary system;
  - accept a wide variety of file types, including CSV, tab-delimited and other text files, MS Excel workbooks, GIS files and many other file types;
  - publish to one of our many partnering data repositories (i.e. IEDA, CUAHSI, KBase) or the BiG CZ Central ODM2 database selected through guided user choice (see Obj. 4, BiG CZ Central).

Objective 3 Deliverables: The BiG CZ Toolbox to enable cyber-savvy CZ scientists and data managers to manage and publish the data they produce through a single scientist focused toolkit. The BiG CZ Toolbox will incorporate access multiple BiG CZ APIs from a single package for searching, accessing, transforming and analyzing heterogeneous data using readily deployed open-source scripting and database packages, and to load cross-discipline data into the BiG CZ infrastructure. Training and Testing workshops/bootcamps (Obj. 1) will focus on using the BiG CZ Toolbox. The BiG CZ Toolbox includes two main capabilities:

- **Facilitation of local data management** and publication to data repositories via Big CZ Central. Components will include:
  - A ready-to-use cross-platform relational database schema based on ODM2 implemented in PostgreSQL, to facilitate local data management of both sensor and sample based datasets. This database will build on the capabilities of ODM 1.1, which are focused on site-based time series for particular variables. ODM2 information model is being designed to address results associated with a variety of experimental protocols, such as specimen-based laboratory experiments; observations from sensor arrays, moving sensors, and ex situ sample analyses; and tracking of observation result provenance through processing chains.
  - ODM Tools 2.0, a cross-platform update to ODM Tools for managing sensor networks and their data (http://his.cuahsi.org/odmtools.html), using the ODM2 information model. ODM Tools 2.0 will be a set of Python modules designed to function out of the box with a local ODM2 database or configurable to work with a number of other information repositories.
  - Streaming sensor middleware, possibly based on Data Turbine (http://www.dataturbine.org/) or the CUAHSI HIS Streaming Data Loader, configured out of the box to load data from a stream into a local ODM2 database or configurable to work with a number of other data systems.
  - Web service interfaces for publishing near real-time or historical data to BiG CZ Central. Project tools will facilitate service deployment associated with ODM2 databases, with flexibility for mapping to other data models. We anticipate using recently developing functionality, such as WebSockets (Fette, 2011) or CoAP (Shelby et al, 2012, Dingee, 2013) to address real time performance requirements. We will expand the "Water One Flow in Python" library (WOFpy, http://pythonhosted.org/WOFpy/) and prototype OGC SOS service implementations.

- A publication coordination tool to maintain common identifiers and spatial-temporal registration when publishing genomic and environmental data to appropriate repositories (KBase and ODM2-related repositories, respectively).

- **API for direct access to BiG CZ Central** to programmatically search, ingest, transform and analyze heterogeneous data using the powerful, widely-used Python language and computing environment. Components will include:
  - Web services interfaces for searching and fetching data from BiG CZ Central and associated catalogs for local, programmatic and interactive analysis. This capability will build on existing Python libraries for web-service based data access (OWSLib, ulmo, pyoos), prototype implementations being developed for ODM2, and data access and parsers for KBase that use the new, "MG-RAST" REST API currently under development. These data services will be wrapped into a more consistent object presentation for easier use, relying on core Python scientific libraries for data handling and efficient local storage and transformation (NumPy, SciPy, Pandas, PyTables).
  - Consistent data exploration and visualization capabilities based on Python-based or enabled scientific libraries, including matplotlib, Pandas and VisIt.

Objective 4 Deliverables: The BiG CZ Central software stack to bridge data systems developed for multiple critical zone domains using linked data principles (Berners-Lee, 2007; Heath and Bizer, 2011) BiG CZ Central would provide capabilities to:

- Register and catalog data-series level metadata from domain-specific data services for sensor- and sample-based Earth observation data managed in ODM2 information model, and other information models adopted by the community. Our data repository partners will include:
  - CUAHSI Hydrologic Information System (HIS, http://his.cuahsi.org/) and it's daughter systems, such as HydroShare (http://www.cuahsi.org/HydroShare.aspx), which have already integrated hyrological and meteorological data from dozens of federal and state agencies and hundreds of academic projects.
  - Integrated Earth Data Applications (IEDA, http://www.iedadata.org/) and its many catalogs, such as the EarthChem data repository (http://www.earthchem.org/) and the System for Earth Sample Registration (SESAR, http://www.geosamples.org/).
  - DOE's System Biology Knowledgebase (KBase, http://kbase.science.energy.gov/), which is a large, emerging software and data environment bridging molecular and systems biology of microbes, plants, and their communities.
  - OpenTopogaphy (http://www.opentopography.org/) is a repository and access point for LIDAR data.
  - U.S. and State Geological Surveys through U. S. Geoscience Information Network (USGIN, http://usgin.org)
- Develop a BiG CZ data repository to collect and integrate datasets at multiple spatial and temporal granularities for the CZ community, along with sensor and sample management information that is not presently captured along with datasets by any data repository that we know of. This data repository will be built on a scalable architecture for high performance to support the mapping and visualization capabilities of BiG CZ Portal.
- Develop web services for observation data based on the ODM2 information model.
- Provide a web-service interface to provide single stream/query access to data from multiple, different data systems using controlled vocabularies from ODM2, and mappings to vocabularies for other systems that BiG CZ community adopts for clients such as BiG CZ Portal and others.

C.1.4. Comparison of BiG CZ Software to Existing Software.

We present within our deliverables the software elements that we are likely to integrate into the BiG CZ software system. We selected these software component candidates based on criteria that each is:

- Free and Open Source Software, released with a relatively permissive Berkeley Software Distribution (BSD) or GNU General Public License (GPL) styled licenses (see the software sustainability section);
- Highly performant and scalable with parallel cloud computing, in order to deliver sub-second response times on large datasets;
- Cross-platform deployable in tightly integrated fashion with other software components using Python bindings;
- Software with which our team has prior development experience.

We believe that very fast response times within user interfaces (UI) are a fundamental requirement for user satisfaction and therefore software sustainability. Achieving this will likely require BiG CZ Portal software and BiG CZ Central software, catalogs and data repositories to be nearly all co-located in a common infrastructure. We have selected the open-source Python programming language as a common binding for software components (many of which are written in C++ or Java), particularly those involving data and metadata flows, because Python is flexible, easy to learn, easy to use, and has powerful, high-performance libraries for data manipulation and analysis that are being increasingly used by scientists.

There are a number of other software packages that deal with integration and visualization, but do not meet our requirements. Other data integration software include iRODS, MetaCat, and DataOne API, but these are primarily document management systems that would not sufficiently support granular data-field and data-value level integration across bio-geo boundaries. Other data visualization software include ESRI ArcMap and ArcScene, Google Earth Engine, Microsoft Layerscape, but none of these are open source and thus not adaptable. The Unidata & UNAVCO Integrated Data Viewer (IDV) is one open source visualization library that we may explore further, although it is unclear whether it can be readily adapted as a component of our BiG CZ Portal web application. CUAHSI's HydroDesktop and IEDA's GeoMapApp both provide map-based data discovery, but both are desktop applications not adaptable to a web browser. Several map-based data discovery web pages exist, such as Delaware Environmental Observing System (DEOS), but most are deployed upon the Google Maps API, which is not open source and severely limits adaptability, and few provide as capable a user interface as NVS or integrate GIS data with point observation data. Wunderground.org WunderMaps is an exception, but it only provides weather data and primarily from citizen-science sources (i.e., limited federal data).

Although these and other potential components exist, none are presently integrated into seamless workflows that enable researchers to focus on science. Our plan is to build on the best available existing packages to implement the unique requirements of Critical Zone science, including integration of diverse bioscience and geoscience information models, observation scales and granularity, and visualization capabilities in 3-D and 4-D.

Proposed BiG CZ software functionality that is not currently available includes:

- Services using Azavea GeoTrellis geoprocessing capabilities, demonstrated by the Model My Watershed application, to perform grid-based data-integration and display processing on the fly at lightning speed.
- Integration of existing catalog service functionality for dataset and data series discovery and transfer with Observation Data Model (Versions 1 and 2), WaterML (Versions 1.1 and 2), and OGC Sensor Observation Service (SOS) based query and data access services for visualization and data acquisition.
- Seamless data acquisition and loading into users' native software environments like ArcGIS, R, Matlab, Excel, and Python (using the core scientific packages, NumPy, SciPy, and Pandas).
- Deployable BiG CZ Toolbox as easy-to-install virtual environments for various operating systems (likely based on the customizable Anaconda Python distribution) and virtual machine images for various host environments (Virtual Box, VMware, Microsoft HyperV).

C.2. Research and Development Plan

C.2.1. Project Team

The core of the BiG CZ SSI team includes the same investigators driving critical zone data integration efforts through the CZOData project (Award #1332257), the ODM2 project (Award #1224638), and the CZ-EarthCube Domain Workshop (Award #1252238). This core team has been joined by additional bioscience and geoscience investigators and software developers to expand on the team's extensive experience in community building, software development, and distributed system architecture. Team members and their roles are described in detail in the supplementary Management and Coordination Plan.

Aufdenkampe (SWRC, lead-PI) is an aquatic geochemist who is co-leading the Christina River Basin CZO and leading the CZOData project. He will oversee overall project coordination, including mentorship of the post-doctoral project coordinator, and will be responsible for Objective 1 community engagement and broader impacts. **Mayorga** (UW-APL, co-PI) is a biogeochemical modeler and co-developer for the NANOOS Visualization System. He will oversee Objective 2 BiG CZ Portal and Objective 3 BiG CZ Toolbox software development. **Zaslavky** (SDSC, PI) is a cyber-scientist and co-developer of the CUAHSI Hydrological Information System (HIS) cyber-infrastructure. He will oversee Objective 4 BiG CZ Central software development. These three together will form the project's executive team.

Objective 1 team. Aufdenkampe will lead, including a small team of technical and data management staff for SWRC and the CRB-CZO who will test each software release. **Packman** (NU), a hydrologist, and **Aronson** (UCR), a microbiologist, will co-chair the CZ Community Advisory Board. Packman will lead the CZ Science Synthesis Institute. **Gill** (SWRC Director of Education), a soil scientist, will engage resource managers, citizen science NGOs and secondary school educators throughout the project by inviting key representatives to co-design workshops and by organizing a workshop at the end of year 3.

Objective 2 Team. Mayorga will lead, including a team of developers at UW-APL. **Cheetham** is founder and CEO of Azavea, Inc., a systems design and software development firm specializing in web and mobile software that combine geospatial technology, engaging user interfaces and high performance computing. As lead developer for the GeoTrellis geoprocessing engine, he and the Azavea team will work closely with the UW-APL team to develop the core high-performance web visualization framework. **Versteeg** (Subsurface Insights) is a geophysicist and cyber-developer who will integrate the VisIt 3D and 4D visualization package into the BiG CZ Portal web app. **Lehnert** (Columbia) is a marine geologist and lead for the Integrated Earth Data Applications (IEDA) facility. She and her team will integrate a suite of data publication/registration forms into the BiG CZ Portal web app.

Objective 3 Team. Mayorga will lead, including a team of developers at UW-APL. **Horsburgh** (USU) is a hydrologist, the developer of ODM 1.1 database and tools, and the PI of the ODM2 project. He and his team will work closely with the UW-APL team to develop the BiG CZ Toolbox.

Objective 4 Team. Zaslavsky will lead. **Valentine** (SDSC) is a developer for CUAHSI HIS and many other projects, using the Microsoft DotNet platform, Python, OData, and various web services, and experience with tuning geospartial databases and digital library systems and metadata. **Richard** (AZGS) and his team (Ryan Clark, Genhan Chen) developed the Geoscience Information Network (USGIN) (Award #0753154) and the National Geothermal Data System 'node-in-a-box' stack (Clark et al., 2013). They have experience with Django, Python, Node.js, and Javascript, and a variety of other FOSS components. **Berg-Cross** (SOCoP) brings expertise in metadata integrating and vocabulary mapping. Together, Zaslavsky, Valentine, Richard and Berg-Cross will develop the BiG CZ Central software stack. **Meyer** and **Henry** (UC) serve as the Microbial Communities and the Microbial Science team leads for the US DOE's Systems Biology Knowledgebase (KBase). They will work with the Obj. 4 team to implement or expose features within KBase to enable cross-domain interoperability with BiG CZ Central.

C.2.2. Community Driven Approach

For a software product to be successful, it must provide functionality with direct value to the intended user community, with user interfaces that are intuitive and easily learned. Our user-centered development process will start with community input through the co-design workshops to identify top-priority functionality, and develop user interfaces based on user-testing and feedback integrated into the development process. Close collaboration between active researchers in the community and the technology-development team will be used to keep development focused on actual user experience and workflow priorities, and the Training and Testing workshops will provide periodic review and assessment from the community.

The CZ Community Advisory Board (CZCAB) will attract domain scientists to user workshops and bootcamps by presenting posters and talks on the software and its uses within individual domains, as BiG CZ develops. The users will then be welcomed to workshops where they will be led through demonstrations. In the first workshops, the software development and CZCAB will gather to work on co-development of the platform, first just the project team and then again with domain scientists interested in having a lasting impact on the project. Next, after development phase 1, boot camps and testing workshops will be held with members of various CZ domains to test the beta version. Similarly, the CZCAB will gather domain scientists for workshops and trainings after the development of versions 1 and 2. The Board will lead discussions among CZ domain scientists at these workshops, in order to generate documents describing recommended changes to the software for scientific use. The CZCAB take the these recommendations directly to the developers, to work out any vocabulary inconsistencies between domain and computer scientific languages.

C.2.3. Software Engineering Process

The project team will use ‘Agile’ project-management and software-development methodologies in the design and development process to discover impediments early, and produce a well-tested and high-quality software system. An adaptive and cyclic software development methodology is appropriate in this project because user scenarios, workflows, and requirements for integration between biological and geospatial databases for the critical zone research community are diverse, and are not completely known from the start. Maintaining a continuously updated, web-accessible list of functional requirements and priorities based on user input will be an important component of project management. In addition to managing requirements, we need to determine the correctness of information transformations, and visualizations using human interaction from domain scientists. To do this, we will use Approval Tests (Falco and Gilkerson, 2013) to expand the testing domains to scenarios where automated evaluation is difficult or human evaluation is required.

We will utilize the Scrum project management process for the software engineering tasks, with one project PI acting as “product owner”, an experienced project manager serving as scrum-master, and development teams at Azavea, AzGS, Columbia, SDSC, USU, and UW-APL. The project backlog -- the prioritized list of features, accompanied by technical work and software exploration notes -- will be initiated at the initial co-design workshops, and updated continuously as the product evolves with input from the project participants and community. Following the definition of the functional scope development will proceed in 2-4 week sprints ending in a stable software release which serves as the start point of the next sprint. The use of test-driven development (tests are written for each function as the application software is written), frequent integration testing, and selective pair programming will be used to assure product quality. BiG software modules will be continuously tested at several levels, including build-integrated unit testing, and integration testing to verify interfaces between BiG components. Approval tests will validate software releases against science requirements, and for items such as data set transformations where precise unit testing is not ideal and domain scientists need to be involved in the evaluation of results.

We will use collaborative code management platforms and automatic build systems such as GitHub, Codeplex, Google Code, TeamCity, Alfresco, as well as project communication tools such as project wiki and issue tracking system (we have experience with all of these and choose based on current features). This will streamline communication for our geographically distributed development team, and has the potential to engage other developers to contribute to our open-

source code-base. Vagrant (or similar) software will be used to assure a reproducible and identical development environment for the software development teams in different locations.

Incorporating community input in all stages of development is critical for success of the project. The project web site, presentations and booths at professional meetings, and surveys of users and stakeholders will be used to ensure that the community is aware of the project and there are clearly defined ways to provide input and contribute to development. Community input will be solicited through an interactive online inventory of user stories with templates for community members to contribute and comment on scenarios, maintaining a question and answer system at the BiG portal, and focus group demonstrations of software releases and special attention to user satisfaction. A user interface designer will be a part of development team working on front-end components of the portal. This input will be used to update product requirements, product backlogs, and priorities as the project develops, leading to improved satisfaction that the tools are meeting the needs to the users.

The BiG software will be designed to support large volumes of data, and provide efficient visualization experience. Our project team includes software engineers with substantial experience developing highly scalable distributed data processing systems (GeoTrellis) and a variety of data visualization tools, including D3, Cesium and the VisIt framework. Software libraries will be developed using iPython supporting parallel execution. Scalability and extensibility will be also achieved by reliance on standard service protocols. We plan to build on existing, active open-source software projects whenever possible to increase the likelihood that support for the software will continue after the project funding ends.

C.2.4. Timeline and Metrics

Project Month: Activity, over 48 month project.

0-9: Co-design workshops/meetings (2 total)
9-21: Development phase 1, beta version.
18-27: Training & testing BootCamps on alpha & beta versions (2 total)
24-33: Development phase 2, version 1
30-39: Training & testing BootCamps on version 1 (2 total)
39-45: CZ Science Synthesis Institute and Citizen-Science & Resource Manager Workshop
39-48: development phase 3, version 2

We will develop a set of internal and external metrics for success. Internal metrics will be developed by the project team at a high level as measures of internal consistency, functionality vs specifications, code re-use across iterations, efficiency of support within the group and at a low level specific task accomplishments will be iteratively tallied for each milestone. External metrics will be based primarily on formative and summative evaluation, as well as publications (peer-reviewed) and other products generated based on BiG CZ, its integration with other CI efforts, curriculum use, invited talks, workshops, citations and eventually by adoption by a larger community of researchers working at the intersection of biological and environmental processes. At a low level we will quantitatively monitor usage: number of registered users and the communities they represent; diversity of users; number of software downloads; number of accesses to educational materials; number of students using them or exposed to the tools in the curriculum; number of code developers; number of trained users; volumes of data that are processed via services.

User input to the project deliverables is critical to assure that they are both intuitive to use and provide the necessary functionality. We will develop an online survey asking each tester to provide feedback on the ease of use and the effectiveness of each tool, especially after workshops. We will also provide a space for open-ended suggestions that will be fed back to the tool developers. In addition, we will develop questionnaires for those participating in each of the workshops and bootcamps that ask participants to respond to specific content and the format of information as well as an overall assessment of the learning experience. As the web applications are broadly available, we will also ask users to take an online survey to assess the quality of the information, the ease of use and requests for any additional information or tools. A document outlining the ongoing evaluation will be provided as part of the annual and final reports for the project.

C.2.5. Project Tasks

Four major tasks are identified; the first of these (community engagement and software design) will be the first activity to be initiated on the project timeline, but the other three tasks will be executed in parallel, with software development sprints focused on aspects of those tasks as dictated by evolving priorities in the Agile Scrum process.

Objective 1 Tasks: Community Engagement and Software Design

The first objective is development of a white paper scoping the system architecture, major user scenarios to be supported, and the software development environment. This will be a living document during the duration of the project, subject to update and revision as a function of the Agile development process; at project completion, it will serve as the foundational system documentation. The software development teams will need to establish and test the development environment, with any training necessary to coordinate the distributed development team.

Task 1.1. Organize and conduct co-design workshops. This will involve arranging a location, inviting participants, logistics, developing the workshop syllabus, running the workshop, and reporting on results. The results will include user scenarios and requirements to inform software development, design for system architecture, identification of existing components and software projects that can be leveraged.

Task 1.2. Integrate workshop results into an initial project backlog. This backlog will identify tasks necessary to implement the user scenarios and requirements identified in the co-design workshops. Software development will be based on the use cases

Task 1.3 Organize software development teams and management process. Identify scrum master.

Task 1.4 Select and test the software development environment. Establish software version control repositories, test script-based building of the development environment. If code from existing projects will be used as a starting point, get the source code and test building and integration of existing software.

Task 1.5 Ongoing community engagement, testing and feedback. The user training and testing bootcamps (interleaved in Tasks 2,3,4), CZ Science Synthesis Institute, and Citizen-science & Resource Manager workshop together form a critically important component of software design and refinement. The Institute in particular will serve as the capstone evaluation of our success. Prior experience with the similar Hydrologic Synthesis program demonstrated the potential bottleneck and enormous effort required to organize data in advance of the workshop. If we can successfully guide BiG CZ Institute participants to integrate and analyze data on the fly, our project will be deemed incredibly successful by the community. In addition to workshops, the project team will conduct demonstrations and workshops on the BiG CZ system at professional meetings and other targets of opportunity to engage the community and solicit feedback.

Objective 2 Tasks: BiG CZ Portal Web Application Development

The BiG CZ Portal component of the project will be aimed at a more general audience than the ToolBox. The Portal will support contemporary web browsers and will rely on open web standards, enabling usage without plugins or other specialized software. Beginning in Year 1, the Azavea user interface design team will develop a series of personas to represent audience types and then wireframe screen designs for review by stakeholders. User scenarios developed in Task 1 will be deconstructed into specific software development tasks to be implemented in a series of sprints over a 12 month period, beginning late in Year 1. The first training and testing workshop will be planned in the middle of this period. Following this initial period of development, we do not expect the development to be a continuous process. Rather, it will be interrupted by feedback sessions with project partners and stakeholders. Workshops and other feedback will shape subsequent development periods in Years 2 and 3. We expect the software development completed prior to the first CZ Science Synthesis Institute in early year 4.

Task 2.1. UI/UX Design -- The user interface and user experience (UI/UX) design process will begin with development of a series of personas representing key target audience members. Based on these personas and the material developed in Task 1, the UI/UX design team will implement

wireframe screen designs that will combine the personas and features. The initial wireframes will not be functional, but following review with the project team, they will be translated into functional HTML/CSS/Javascript scaffolds to which data will be connected by the development team.

Task 2.2. Software Development -- Using the functional scaffolds delivered in Task 2.1, the software development team will implement the features based on the various data source and server-side visualization services. It is anticipated that an initial implementation will rely significantly on the existing capabilities in the Nanoos Visualization System and the GeoTrellis technology, but that subsequent iterations will depart from this starting point in order to adapt to the specific user workflow required by the BiG CZ Portal.

Task 2.3. Software Deployment - Leveraging contemporary software engineering best practices, the software development teams will implement a continuous integration (CI) service, test-driven development practices, and a continuous deployment process that will use open source tools such as Vagrant, Chef and Puppet. The software team will continue to adapt these tools as the development and deployment environments change.

Task 2.4. Documentation and training materials for use of the Portal. Development of software documentation will be an integral part of the development process, but training materials for research users will need to be developed on a separate track.

Objective 3 Tasks: BiG CZ Toolbox Development

The BiG CZ Toolbox is aimed at investigators who have data that they want to share using the BiG CZ Software System or who wish to programmatically access BiG CZ Central to search, ingest, transform and analyze data using Python, which can in turn be readily linked to or wrapped around other statistical or modeling applications written in R, C++, Fortran, Java, Matlab, and others. Development of the Toolkit will follow standard services oriented architecture (SOA) design, but will be fully open source and use only freely available components. Beginning in Year 1, the BiG CZ Toolbox team will establish a scope for and begin to design and prioritize software tools to be included in the BiG CZ Toolbox as necessary to implement data publication and access user scenarios. The sequence of tasks would be similar to that identified for Objective 2, with a series of sprints based on established priorities, leading up to a user workshop for testing and feedback. Toolbox development will begin after the first training and testing workshop, and continue through follow up on the final Citizen-science & Resource Manager and CZ Science Synthesis Institute workshops.

Task 3.1. Initial Software Design -- The relational database schema for ODM2 and the software prototypes developed by the ODM2 project will serve to initiate our software designs for the Toolkit. Given the goal of lowering the activation energy for CZ scientists to contribute data and use the system, the designs for Toolkit software will focus on operating system agnostic tools that ease the data management and sharing process.

Task 3.2. Development of Data Management and Publication Software -- Using the initial protoypes and designs developed in Task 3.1, we will operationalize a relational database implementation of ODM2 for PostgreSQL, a Python-based version of the ODM Tools software application for managing data stored in an ODM2 database, streaming sensor data middleware for automating the loading of data streams into an ODM2 database, and standardized web service interfaces for publishing data stored in an ODM2 databases.

Task 3.3. Development of Toolkit Software Deployments -- To facilitate ease of installation and use, we will develop executable installers and virtual machine clones for Windows, Mac OSX, and Linux platforms for the software tools developed under Task 3.2. Based on the community requirements and community stories, gather sample datasets, develop tests scripts and automated scripts to allow for testing of the components, individually, and as part of a deployed environment or virtual machine stack. Implement the scripts in a continuous build environment, and work to create semiannual stable builds. The virtual machine will include an operating system, data persistence components, service interface implementations, vocabulary services, and processing services.

Task 3.4. Develop documentation and training materials. Use the community stories that have been developed into automated tests as the basis for documentation and user training. Work with the community to expand the scope of the training to encompass emerging needs

Although not listed here, we would initiate a community engagement, iteration, documentation, and feedback process for development of the BiG CZ Toolbox that is similar to that described for development of the BiG CZ Portal (i.e., see Tasks 2.4-2.8 above).

Objective 4 Tasks: BiG CZ Central Development

Implement server stack to host BiG CZ repository, metadata harvesting and search, and data services and other APIs to support the Portal web application (Obj.2) and Toolbox (Obj. 3).

Task 4.1. Design system architecture integrating all Big CZ components. This task includes collecting community expectations, and gathering user stories which will be used to develop architectural requirements and product backlogs. Data persistence components, service interface implementations, vocabulary services, and processing services to manage and integrate cross-scale cross-disciplinary data, will be essential design components. The existing implementations to be leveraged include the USGS WaterSmart services stack (Department of the Interior, 2012), the AUScope Spatial Information Server Stack (Fraser et. al, 2010), and OSGeo Live distribution (OSGeo 2012). Based on the review, we will select components that meet requirements determined through the backlog process, and determine which of the components needs to be expanded and integrated in the BiG CZ central software stack.

Task 4.2. Develop tools to build an integrated catalog from CZO and other community information systems. To support user queries against both geoscience and bioscience data sources, we will develop an integrated catalog as a federation of CUAHSI, IEDA, KBASE, CZOData, OpenTopography and USGIN sources. The catalog will follow the ODM2 schema (ODM2 catalog extension), and expose at least CSW service interface and the interfaces designed through the ODM2 project. Records from domain catalogs will be either periodically harvested into the BiG CZ catalog or accessed at query time, depending on the types of interfaces exposed by these catalogs. This design will allow us to enable data discovery queries for diverse types of CZ-related information and support visualization of data availability, while metadata returned on data discovery queries will be used to formulate data requests against different data sources registered in the domain catalogs.

Task 4.3. Develop semantics components to support data publication, discovery and annotation across multiple data systems, and transformations to support cross-domain cross-scale integration. This task will address semantic and scale discrepancies across information systems being federated in the BiG CZ Central. We will use APIs exposed by KBase, CUAHSI HIS, IEDA, USGIN, OpenTopography and CZOData to develop information model mappings between them, using ODM2 as the target schema implemented at the BiG CZ Central. Development of improved conceptual models that handle a range of granularity & scaling to bridge across domains will be accompanied by implementing transformations of domain metadata into ODM2 formalisms. The latter functionality will be organized as a Python library and also made available via a standard service interface such as OGC WPS. We will ensure that the protocols and information models we employ are standards compliant, integrating OGC standards and best practices with standard markup being developed through the Genomics Standards Consortium (in particular, the Genomic Contextual Data Markup Language).

Task 4.4. Development of BiG CZ Central Web Service API, using designs from Task 3.1, we will implement a web service interface for discovering data via the BiG CZ Central system and that facilitates data retrieval for local analysis.

C.3. Software Sustainability

C.3.1. Usability as the Foundation for Sustainability

Delivering responsive, easy-to-use, powerful software is the key to building a critical mass of users and is thus a foundation for sustainability. To gain community input, we will solicit reviews from the CZ Advisory Board, which will advise on feature development priorities, provide critical feedback on software releases, and assist in the recruitment of workshop participants. The Co-Design

Workshops and the Training and Testing workshops will be used to engage users, show them how to use the system to be more efficient in their work, and obtain feedback to maximize the utility of the software. We view the cultivation of an active user and developer community, which will initially occur through these workshops, as a key part of developing sustainable software. Allocation of resources for training and documentation is thus an essential part of this project.

C.3.2. Sustainability Plan

Sustainability is a significant issue for software developed by a research community. We will build on existing, active open-source software projects. Doing so increases the likelihood that support for the software will continue after the project funding ends. For example, our development plan involves building on top of existing capabilities of mature systems such as the Nanoos Visualization System and the GeoTrellis technology. Our project team includes the software engineers that developed these systems and that have substantial experience developing highly scalable distributed data processing systems and a variety of data visualization tools. New functionality developed for this project can be leveraged for future implementations of these tools.

Much of our proposed development will use Python, which has seen tremendous growth in popularity in the past few years and has seen widespread deployment, excellent documentation, and a comprehensive library of already available code and tools. Scalability and extensibility of developed software will be achieved by reliance on standard service protocols and data encoding formats (see Data Management Plan). The unanimous envisioning of the proposed BiG CZ Software System by attendees at the CZ-EarthCube domain workshop means that there is great need for such a system and that it has high potential for being a significant part of the EarthCube cyberinfrastructure that is currently being envisioned by the greater geosciences community. The evaluation metrics described in this proposal will be used to quantitatively and qualitatively evaluate software releases and to ensure that community expectations are met.

Our commitment to ensuring accessibility of developed source code via open source distribution in a public repository and emphasis on useful documentation of the codebase will foster adoption of the code by other projects, which is necessary for sustainability after the initial project duration. Indeed, development of code as contributions to existing open source projects provides an approach to sustainability. Where possible, all new **source code and software developed by this project will be released under the New BSD (BSD 2) Open Source license** (Open Source Initiative, 2008). Given that we are incorporating and modifying some existing applications that already use different open source software licenses (e.g., GeoTrellis uses GPL3), the final BiG CZ Software System may use multiple open source software licenses.

C.4. Broader Impacts and Outreach Plan

C.4.1. Broader Impacts to Earth & Environmental Sciences

"*Computers are as important to modern science as telescopes and test tubes. Unfortunately, most scientists are never taught how to use them effectively: most scientists have to figure out for themselves how to build, validate, maintain, and share complex programs. This is as fair as teaching someone arithmetic and then expecting them to figure out calculus on their own, and about as likely to succeed*." From: http://software-carpentry.org/ (Hannay et al. 2009; Wilson 2009).

We believe that a key aspect of increasing the computational capabilities of the CZ community requires increasing the computational skills of the CZ science community. The Software Carpentry project has shown that a 20-40 hour course can increase the productivity of typical scientist by about 20%, or a day a week, and that it is common for their bootcamps to enable participants to launch into research topics that would not have been previously possible. We will partner with the Software Carpentry project to teach 80 early-career Earth and environmental scientists computational science skills through four Bootcamps and one CZ Science Synthesis Institute. Combined with the easy-to-use Toolbox that we develop, we expect this will have a profound and rippling impact on increasing community capabilities.

C.4.2. Outreach to Resource Managers and Citizen-Scientist Volunteers

Humans have become a geologic force on this planet, moving more sediment to the sea by an order of magnitude than move through natural processes (Williams, 2005). In addition, human impacts on Earth's water and the atmosphere threaten overall environmental quality (e.g., Foley et al., 2005). For that reason, scientific literacy is necessary to support both personal and civic decision-making as well as overall economic productivity. It is, therefore, necessary for the public to: 1) understand basic scientific concepts and 2) use data to make informed decisions. The broader impacts of this project will provide data within the context of scientific concepts to enhance scientific literacy for the public and for resource managers. To do this, we will develop the following deliverables: 1) ***designing*** an intuitive public interface for viewing and downloading data; 2) ***developing*** "how to" training material on access to and the use of scientific data; 3) ***linking*** citizens to ongoing local research efforts; 4) ***providing*** information on basic scientific processes to assist users to understand the implications of the data; 5) ***offering*** data manipulation and visualization tools that will allow users to compare sites, chart time-series data, and compare conditions to their own local measurements; and, 6) ***teaching*** how to use data to enhance civic engagement and citizen science efforts. To accomplish those tasks, as stated above, we will hold a workshop for resource managers and citizen scientists to introduce them to the deliverables outlined above and to solicit feedback on the usability of those tools.

C.4.3. Impact to Secondary Education

We will also develop curricular materials around the portal designed specifically for secondary teachers to assist them to integrate research data into their classrooms. These materials will be included in our extensive library of educational curricular materials available to educators through our website.

C.5. Results from Prior NSF Support

Integrated Data Management System for Critical Zone Observatories, PIs, A. **Aufdenkampe**, I. **Zaslavsky**, E. **Mayorga**, K. **Lehnert**, J.S. **Horsburgh**, Award No. 1332257, $1,469,953, 12/1/2012-3/31/2014. The objective of the project is to develop a comprehensive, integrated data management system for the NSF-funded Critical Zone Observatory (CZO) program, called CZOData. The overall goal for CZOData is to support, empower, and broaden the impact of CZO science and maximize the return on investment of the CZO program by transforming capabilities to easily share, integrate, analyze and preserve the wide range of multi-disciplinary data generated within and across CZOs.

Developing a Community Information Model and Supporting Software to Extend Interoperability of Sensor and Sample Based Earth Observations, PIs J.S. **Horsburgh**, I. **Zaslavsky**, K. **Lehnert**, A. **Aufdenkampe**, E. **Mayorga**, Award No. 1224638, $281,112, 8/1/2012-7/31/2014. This EAR Geoinformatics Program grant supports development of the Observations Data Model 2.0 (ODM2) and related software to enable web based interoperability of earth observations derived from sensors and samples that span discrete data and informatics initiatives for multiple communities. The system targets data from specific existing repositories in order to demonstrate how the information model can support federation of earth observations data across multiple data publication systems.

Zaslavsky is the PI of **EAGER: Readiness of Disciplinary Data Systems for Cross-Domain Interoperability within a Standards-Based EarthCube Reference Framework** (NSF-EAR, award 1238420, 4/1/2012-3/31/2013, $217,932.) The project has developed an EarthCube roadmap focused on cross-domain interoperability (version 1.1, August 2012, http://earthcube.ning.com/group/interop/page/roadmap), and a framework for evaluating interoperability readiness, and created a range of inventories - of EarthCube domain resources, architectures, and geoscience models – which are accessible online (http://earthcube.ning.com/group/interop/page/community-inventories).

A. **Packman** was the PI of NSF EAR-0810270, **Linking scales of geomorphology and solute transport in river corridors**, $248,852, 2008-2012. This project demonstrated that interactions

between river flow and fractal stream morphology produce non-linear multi-scale surface-groundwater interactions, and resulted in 10 publications (Stonedahl et al., 2010, 2012, 2013; Drummond et al., 2012; Patil et al., 2012; Boano et al., 2013; Harvey et al., 2012; Zhang et al., 2012; Guan et al., 2011; Cullis et al., 2012). We conducted two summer institutes focused on synthesis of hydrological, morphodynamical, biogeochemical, and ecological processes, providing cross-cutting training to ~40 early-career scientists; plus K-12 and public outreach events.

Model My Watershed: Developing a Cyberlearning Application and Curricula to Enhance Interest in STEM Careers (NSF DRL #0929763, $1,199,609 total, 2009-2012) PI: S. **Gill**; Co-PIs: **Aufdenkampe**, Dow, Marcum-Dietrich, Newbold and Tomlin. R. **Cheetham** (Azavea) was the lead developer. The MWW Project (http://app.wikiwatershed.org/stroudfx/bin-debug/Main.swf) is a place-based, hydrologic modeling web application that allows students, teachers and citizens to model storm-water runoff for their own neighborhoods using a professional-grade model and real data. Project modules were extensively tested by teachers and students, with evaluations confirming that the application was easy to use and led to broad knowledge gains and linkage to STEM careers. Project results were presented at many meetings and two publications (Gill et al., in press and Gill and Marcum-Dietrich, in review). The project was highlighted at the Sept. 2011 NRC/NSF STEM Smart event in Philadelphia as an example of effective STEM education.

Christina River Basin Critical Zone Observatory (NSF EAR #0724971, $4,300,000 total, 2009-2013, PI: D. Sparks; Co-PI **Aufdenkampe**, http://criticalzone.org/christina/) has the two-fold goal to (1) create a community resource for studying earth surface (i.e. critical zone) processes through sampling, sensing and cyberinfrastructure, and (2) test a set of hypotheses on the whole-watershed links between water, mineral and carbon cycles over a range of modern and historical land uses. The project team includes 4 post-docs, 11 graduate students, a sensor engineer and 17 faculty at many institutions including **Zaslavsky**. Results to date have been presented in 17 peer-reviewed publications, more than two-dozen invited talks and dozens of other presentations.

INTEROP--Geoscience Information Network (USGIN) (NSF EAR #0753154, $624,974 total, 2008-2012. PI Allison, Co-PI S. **Richard**, http://usgin.org). This project demonstrated data interoperability through initial partnerships with GEON, EarthChem, and USGS Community for Data Integration. After the first year of design and testing, USGIN was selected as the model for the more than $35 million investment by the U.S. Department of Energy for the Geothermal Technologies Program's National Geothermal Data System (NGDS – http://www.geothermaldata.org). The USGIN project converted the broad, generalized agreement between federal and state Surveys into a functional, operational data integration framework that is scalable and transportable across the geosciences, other scientific disciplines, and the geospatial data infrastructure.

C.S. **Henry** receives support from three NSF grants on: (A) Maize modeling and annotation (#1025398), requiring development of a large biochemistry database for plants and a new resource for the automated annotation and modeling of plant genomes (Plant SEED) resulting in 3 papers (Gerdes et al., 2012; Seaver et al., 2012; Krumholz et al., 2012); (B) microbial community modeling in the gut (#1137089), resulting in a Nature Communications paper on microbial interactions; and (C) annotation and analysis of the impact of metabolite repair (#1153357), which began last year and has a number of paper in preparation.

PI: **F. Meyer** –Second Argonne Soils Workshop**, Argonne National Laboratory, Argonne, IL, October 6-8. 2010.** Federal Award ID: 1059066. Project Period: 09/01/10 - 4/30/2012

E. **Aronson** and G. **Berg-Cross** have not served as a PI or Co-PI on an NSF grant in 5 years.

D. REFERENCES CITED

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

de la Beaujardiere, J., ed. (2006), OpenGIS Web Map Server Implementation Specification, OGC Implementation Specification OGC 06-042, Version 1.3.0. http://portal.opengeospatial.org/files/?artifact_id=14416.

Department of the Interior, 2012. WaterSmart Clearinghouse. http://www.doi.gov/watersmart/html/

Boano, F., Harvey, J.W., Marion, A., Packman, A.I., Revelli, R., Ridolfi, L., and Wörman, A., 2013, Hyporheic flow and transport processes, Reviews of Geophysics, submitted.

Bröring, Arne, Stasch, Christoph, and Echterhoff, Johannes. 2013. OGC® Sensor Observation Service Interface Standard OGC Implementation Specification OGC 12-006 https://portal.opengeospatial.org/files/?artifact_id=47599

Clark, Ryan, Kuhmuench, Christoph, and Richard, Stephen M. (2011): Developing the National Geothermal Data System Adoption of Ckan for Domestic and International Data Deployment. *Proceedings, Thirty-Eighth Workshop on Geothermal Reservoir Engineering,* Stanford University, Stanford, California, February 11-13, 2013, accessed at https://pangea.stanford.edu/ERE/pdf/IGAstandard/SGW/2013/Clark.pdf.

Cullis, J., Gillis, C., Bothwell, M. , Kilroy, C., Packman, A.I. , and Hassan, M., 2012, A conceptual model for the growth, persistence, and blooming behavior of the benthic mat-forming diatom Didymosphenia geminata in oligotrophic streams, Journal of Geophysical Research-Biogeosciences, 117, art. no. G00N03, doi: 10.1029/2011JG001891.

Drummond, J.D., Covino, T.P., Aubeneau, A.F., Leong, D., Patil, S., Schumer, R., and Packman, A.I., 2012, Effects of solute breakthrough curve tail truncation on residence time estimates: A synthesis of solute tracer injection studies, Journal of Geophysical Research-Biogeosciences, 117, art. no. G00N08, doi:10.1029/2012JG002019.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., Snyder, P.K., Global Consequences of Land Use. *Science 22 July 2005: 309 (5734), 570-574. [DOI:10.1126/science.1111772]*

Fraser, Ryan, Woodcockm, Robert, Simons, Bruce, and Wyborn, Lesley. 2010. AuScope Grid: Spatial Information Services Stack. Global Geospatial Conference 2010. http://www.gsdi.org/gsdiconf/gsdi12/papers/36.pdf

Guan, K., Thompson, S.E., Harman, C.J., Basu, N.B., Rao, P.S.C., Sivapalan, M., Packman, A.I., and Kalita, P.K., 2011, Spatio-temporal scaling of hydrological and agrochemical export dynamics in a tile-drained Midwestern watershed, 47, W00J02, Water Resources Research, doi:10.1029/2010WR009997 [Editor's highlighted article].

Hannay, J. E., Langtangen, H. P., MacLeod, C., Pfahl, D., Singer, J., & Wilson, G. (2009). How do scientists develop and use scientific software?. In Software Engineering for Computational Science and Engineering, 2009. SECSE'09. ICSE Workshop on (pp. 1-8). IEEE.

Harvey, J.W., Drummond, J.D., Martin, R.L., McPhillips, L.E., Packman, A.I., Jerolmack, D.J., Stonedahl, S.H., Aubeneau, A.F., Sawyer, A.H., Larsen, L.G., and Tobias, C.R., 2012, Hydrogeomorphology of the hyporheic zone: Stream solute and fine particle interactions with mobile bedforms and floods, *Journal of Geophysical Research-Biogeosciences*, 117, G00N11, doi:10.1029/2012JG002043.

Heidorn, P. B. (2008). Shedding light on the dark data in the long tail of science. *Library Trends*, 57(2), 280-299.

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.

Lehnert, K., Y. Su, C. H. Langmuir, B. Sarbas, and U. Nohl. 2000. A global geochemical database structure for rocks. Geochemistry, Geophysics, Geosystems 1:Paper number 1999GC000026.

Mayorga, E., Tanner, T., Blair, R., Jaramillo, A. V., Lederer, N., Risien, C. M., & Seaton, C. (2010). The NANOOS Visualization System (NVS): Lessons learned in data aggregation, management and reuse, for a user application. In *OCEANS 2010* (pp. 1-9). IEEE.

National Science Foundation (2012): Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21). Washington, DC. http://www.nsf.gov/od/oci/cif21/CIF21Vision2012current.pdf.

National Research Council (2001): *Basic Research Opportunities in the Earth Sciences (BROES)*. Washington, DC: The National Academies Press. http://www.nap.edu/catalog.php?record_id=9981.

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.

Open Source Initiative, (2008), Open Source Initiative OSI - The BSD License, http://www.opensource.org/licenses/bsd-license.php, accessed 3/18/2013.

OSGeo. 2012. Open Source Geospatial Foundation Live DVD. http://live.osgeo.org/en/index.html

Patil, S., Covino, T., Packman, A.I., McGlynn, B., Drummond, J.D., Payn, R., and Schumer, R., 2012, Intra-stream variability in tracer breakthrough curves: Controls on solute retention behavior, J*ournal of Geophysical Research-Earth Surface*, doi:10.1029/2012JF002455.

Stonedahl, S.H., Harvey, J.W., and Packman, A.I., 2013, Nonlinear interactions between hyporheic flow produced by stream meanders, bars, and dunes, Water Resources Research, submitted.

Stonedahl, S.H., Harvey, J.W., Detty, J., Aubeneau, A., and Packman, A.I., 2012, Physical controls and predictability of stream hyporheic flow evaluated with a multi-scale model, Water Resources Research, 48, doi:10.1029/2011WR011582.

Stonedahl, S.H., Harvey, J.W., Wörman, A., Salehin, M., and Packman, A.I., 2010, A multi-scale model for integrating hyporheic exchange from ripples to meanders, Water Resources Research, 46, W12539, doi:10.1029/2009WR008865.

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. *Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation*, July 2009, p.988-994..

Taylor, P. (Ed.), (2012), OGC WaterML 2.0: Part 1 – Timeseries, OGC Implementation Standard. OGC 10-126r3, https://portal.opengeospatial.org/files/?artifact_id=48531

Taylor, P., Cox, S., Walker. G, Valentine, D, and Sheahan, P. 2013.WaterML2.0: Development of an open standard for hydrological time series data exchange. Journal of Hydroinformatics In Press.

Thompson, S.E., Harman, C. J., Schumer, R., Wilson, J.S., Basu, N.B., Brooks, P.D., Donner, S.D., Hassan, M.A., Packman, A I., Rao, P.S.C., Troch, P.A., and Sivapalan, M., 2011, Patterns, puzzles and people: Implementing hydrologic synthesis, Hydrological Processes, 25, 3256–3266.

Vretanos, P.A., ed. (2010), Open GIS Web Feature Service 2.0 Interface Standard. OGC Implementation Standard OGC 09-025r1 and ISO/DIS 19142, Version 2.0.0. http://portal.opengeospatial.org/files/?artifact_id=39967.

Whiteside, A. and J.D. Evans (2008), Web Coverage Service (WCS) Implementation Standard. OGC Implementation Standard OGC 07-067r5. Version 1.1.2. http://portal.opengeospatial.org/files/?artifact_id=27297.

Williams, Bruce, 2005, Humans as Geologic Agents: A Deep Time Perspective. *Geology*; March 2005; v. 33; no. 3; p. 161-164; DOI: 10.1130/G21108.1

Wilson, G. (2009). How Do Scientists Really Use Computers?. American scientist, 97(5), 360-362.

Zaslavsky, I., D. Valentine, and T. Whiteaker (2007), CUAHSI WaterML, Open Geospatial Consortium Discussion Paper. http://portal.opengeospatial.org/files/?artifact_id=21743.

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. *Proceedings of the Environmental Information Management Conference*, Santa Barbara, CA, 28-29 September, EIM'2011, pp. 145-150.

Zhang, Y., Meerschaert, M.M., and Packman, A.I., 2012, Linking fluvial bed sediment transport across scales, Geophysical Research Letters, 39, L20404, doi:10.1029/2012GL053476.

Gerdes S., Lerma-Ortiz C., Frelin O., Seaver S.M., Henry C.S., de Crécy-Lagard V., and A.D. Hanson (2012). Plant B Vitamin Pathways and their Compartmentation: a Guide for the Perplexed. Journal of Experimental Botany, Epub.

Seaver, S.M.D., Henry, C.S., and A.D. Hanson (2012). Frontiers in Metabolic Reconstruction and Modeling of Plant Genomes. Journal of Experimental Botany, 63 (6): 2247-58.

Krumholz, E.W., Yang, H., Weisenhorn, P., Henry, C.S., and I.G.L. Libourel (2012). Genome-wide Metabolic Network Reconstruction of the picoalga Ostreococcus. Journal of Experimental Botany, 63 (6): 2353-62.

Freilich, S., Zarecki, R., Eilam, O., Segal, E.S., Henry, C.S., Kupiec, M., Gophna, U., Sharan, R., and E. Ruppin (2011). Competitive and Cooperative Metabolic Interactions in Bacterial Communities. Nature Communications, 2:589.

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.

Adjunct Assistant/Associate Professor, 2004 – Present. Department of Oceanography, College of Earth, Ocean and Environment, University of Delaware, DE.

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) Products—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.

Zaslavsky, I., T. Whitenack, M. Williams, D. Tarboton, K. Schreuders, and **A. K. Aufdenkampe**. 2011. The Initial Design of Data Sharing Infrastructure for the Critical Zone Observatory. Pages 145-150 *in* M. B. Jones and C. Gries, editors. Proceedings of the Environmental Information Management Conference 2011 (EIM 2011). University of California, doi: 10.5060/D2NC5Z4X.

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.

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.

Aufdenkampe, A. K., E. Mayorga, J. I. Hedges, P. Quay, J. E. Richey, C. Llerena and A. 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.

Five Other Significant Products

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.

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.

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.

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.

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.

(iv) Synergistic Activities

Advisor/Supervisor. 5 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, 2004-2013.
Isotope Geochemistry with W. Ullman, University of Delaware, 2006.
Introduction to Field Ecology with J. Thorne, University of Pennsylvania, 2003.
Various 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.
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
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. 2010-2011.
Linda Carter	Post-Doc	Stroud Water Research Center. 2005-2006.
Michelle C. Cogo	M.S.	Universidade de São Paulo, Brazil. Co- w A. Krusche

Robert Cheetham, Azavea

Professional Preparation

1994–1997 **MLA Masters of Landscape Architecture** University of Pennsylvania
1989–1990 **Stanford Japan Center** Kyoto Center for Japanese Studies, Kyoto, Japan
1986–1991 **BA with distinction: Japanese Studies** University of Michigan, Ann Arbor

Appointments

2000–Present **Azavea Inc.**
Founder and President

1995–2003 **University of Pennsylvania**
Senior Software Developer, Cartographic Modeling Lab, 1998 – 2003
Lecturer, Dept. of Landscape Architecture, Univ. of Pennsylvania, 1997 – 1999
Research Associate, Cartographic Modeling Lab, 1996 – 1997
Project Manager, West Philadelphia Landscape Project, 1996 – 1997
Research Assistant, West Philadelphia Landscape Project, 1995 – 1996

1999–2001 **City of Philadelphia**
Senior GIS Developer, GIS Services Group, 1999 – 2001

1997–1999 **Philadelphia Police Department**
Crime Analyst, Philadelphia Police Department, 1997 – 1999

1997–2000 **Private Practice**
Consultant in Geographic Information Systems and GIS Software Development

1991–1994 **City of Konan, Japan**
Coordinator for International Relations

Five Recent Products

GeoTrellis: High performance geoprocessing engine (v 0.6.0 – 0.8.0) [Software]. (2012) Philadelphia, PA; Azavea, Robert Cheetham. Retrieved from https://github.com/azavea/geotrellis

HunchLab: Crime analysis and risk forecasting solution (v 2012-10) [Software]. (2011) Philadelphia, PA; Azavea, Robert Cheetham. Retrieved from www.azavea.com/HunchLab

Cheetham, R and M.L. Johnson. 2011. Using Geography and Math to Find Patterns in Crime. NSF Live Science.

Cheetham, R and M.L. Johnson. 2011. Using GIS and Statistical Analysis to Understand Crime Patterns. Public Safety IT.

Boyer, D; R Cheetham and M.L. Johnson. 2011. Using GIS to Manage Philadelphia's Archival Photographs. The American Archivist.

Five Other Significant Products

OpenTreeMap: Collaborative, geography-enabled urban tree inventory software (v. 1.0 - 1.2) [Software]. (2010). Philadelphia, PA; Azavea Inc., Robert Cheetham. Retrieved from http://opentreemap.github.com/

DistrictBuilder: Collaborative redistricting software (v. 1.0 - 2.0) [Software]. (2009 - 2012) Philadelphia, PA; Public Mapping Project, Azavea. Retrieved from https://github.com/PublicMapping/DistrictBuilder

Heckert, M., R. Cheetham, 2009, Bringing Multi-Criterion Siting Decisions to the Web, In: PhillydotMap: The Shape of Philadelphia, Cartographic Modeling Lab Papers.

Branas, C. B. G. Carr, M. Heckert, and R. Cheetham. 2010. Trauma Center Siting, Optimization Modeling, and GIS. In: R. Skinner (Ed.) GIS in Hospital and Healthcare Emergency Management. Taylor & Francis/CRC Press, Publishers.

Johnson, M., L. Krueger-Braneky, and R. Cheetham, 2010, Leveraging Geographic Indicators to Site Sustainable Business. Canadian Sustainability Indicators Network.

Synergistic Activities

2011–Present **Urban Forestry Simulation and Modeling Tools** Manage design and development of web software for prioritization and simulation modeling of urban forest ecosystems (USDA # 2011-33610-30511 and 2012-33610-19997).

2009–2012 **Model My Watershed** Direct software development activities to build, test and disseminate watershed-modeling tool set for secondary students (NSF # 0929736).

2009–Present **Philadelphia Stormwater Billing Project** Manage design and development of a public web application that supports stormwater billing based on impervious surface cover.

2007–2011 **HunchLab Crime Analysis and Risk Forecasting Solution** PI for NSF Phase I, Phase II, and Phase IIB SBIR projects (NSF #0637589 and IIP-0750507),to design and develop software tools for automated geographic pattern discovery and notification.

2010 **GPU-Based Raster Processing Algorithms** PI for NSF Phase I SBIR project (NSF # IIP-0945742), that tested the feasibility of graphics processing units (GPUs) to substantially increase the performance of GIS software processing.

Collaborators and Other Affiliations (Past 48 Months)

Adams, Jeff – Google; Adams, Michele, Meliora Design; Altman, Micah – MIT; Aufdenkamp, Anthony – Stroud Water Research Center; Baylson, Todd – City of Philadelphia; Becker-Klein, Rachel – PEER Associates; Berry, Elizabeth – TreeKIT; Blair, Lucas – Little Bird Games; Boyer, Deborah – Azavea; Branas, Charles – University of Pennsylvania; Calihan, Patrick – Tech Impact; Caplan, Joel – Rutgers University, Newark; Carr, Brendan – University of Pennsylvania; Cavalier, Darlene – SciStarter; Cohen, Marcia – Development Services Group; Cohen, David – Ben Franklin Technology Partners; Dahme, Joanne – City of Philadelphia; de la Torre, Javier – Vizzuality; Decker, Joan – City of Philadelphia; DiBiase, David – Esri; Elesh, David – Temple University; Elkis, Patty – Delaware Valley Regional Planning Commission; Evans, Barry – Penn State University; Ferdana, Zach – The Nature Conservancy; Fraser, John – Institute for Learning Innovation; Gill, Susan – Stroud Water Research Center; Gillen, Kevin – University of Pennsylvania; Glicksman, Allen – Philadelphia Corporation for the Aging; Gold, Michael – FarSite; Golas, Matt – University of Pennsylvania; Heckert, Megan – Swarthmore College; Hillier, Amy – University of Pennsylvania; Huber, William – Quantitative Decisoins; Jennings, Craig – Center for Effective Government; Johnson, Mary L. – Azavea ; Kennedy, Les – Rutgers University, Newark; Kinnevy, Susan – City of Philadelphia; Kloper, Eric – MIT; Krepicz, Matthew – City of Toronto; Krueger-Braneky, Leanne - BALLE; Leahey, Duke – Nidus Partners; Low, Sarah – US Forest Service; Maco, Scott – Davey Tree Experts; Marcus, Josh – Azavea; McDonald, Michael – George Mason University; McGinnis, Allison – Bureau of Alcohol, Tobacco and Firearms; McGinnis, Sean – State of New Jersey; Merchant, Raina – University of Pennsylvania; Michelson, David – City of Asheville; Mondoro, Dennis – US Department of Justice, Office of Juvenile Justice and Delinquency Prevention; Moskowitz, Bruce – Biomedical Research and Educational Foundation; Camden; Nesbit, Scott – University of Richmond; Nowak, David – US Forest Service; O'Neil, Daniel X.- Smart Chicago Collaborative; O'Neil-Dunne, Jarlath, University of Vermont; Querry, James – City of Philadelphia; Ratcliffe, Jerry – Temple University; Reina, Vincent – New York University; Ring, Lauren – Philadelphia Corporation for the Aging; Robinson, Anthony – Penn State University; Rodbell, Philip – US Forest Service; Satullo, Chris – WHYY; Schlagel, Joel – US Army Corps of Engineers; Silva, Philip – TreeKIT; Switala, Kevin – GeoDecisions; Taylor, Ralph – Temple University; Tomlin, Dana – University of Pennsylvania; Vargas, Kelaine – Urban Ecos; Wang, Gerry – Consultant; Watson, Anthea – Google; Zganjar, Chris – The Nature Conservancy; Zhu, Xiao-Wei, Consultant; Zwarg, David – Azavea.

Graduate Advisors and Postdoctoral Sponsors: N/A
Thesis Advisor and Postgraduate-Scholar Sponsor: N/A

Jeffery S. Horsburgh, PhD
Utah Water Research Laboratory
Utah State University
Logan, UT
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

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

Most Relevant Products

Ames, D.P., J.S. Horsburgh, Y. Cao, J. Kadlec, T. Whiteaker, D. Valentine (2012), HydroDesktop: Web services-based software for hydrologic data discovery, download, visualization, and analysis, Environmental Modelling & Software, doi:10.1016/j.envsoft.2012.03.013.

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.

Other Significant Products

Spackman Jones, A., J.S. Horsburgh, N.O. Mesner, R.J. Ryel, D.K. Stevens (2012), Impact of sampling frequency on annual total phosphorus and total suspended solids loads, Journal of the American Water Resources Association, 48(6), 1258-1275, doi:10.1111/j.1752-1688.2012.00684.x.

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.

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.
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.

Synergistic Activities

- *I lead a team of computer programmers and technicians* in the development of cyberinfrastructure and hydroinformatics databases, software and systems.
- *I am part of the CZOData team* participating in the design and development of an integrated data management system for the Critical Zone Observatories.
- *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 with the CUAHSI Hydrologic Information System (HIS) team* participating in the design of the Observations Data Model, the HydroServer software stack, the HydroShare system, in developing software applications that use technology created by the CUAHSI HIS, and in deploying the HIS software.
- *I operate and maintain continuous hydrologic monitoring infrastructure* for the Little Bear River Experimental watershed.

Collaborators and Other Affiliations

Collaborators and Co-Editors

Daniel P. Ames (Idaho State University); Anthony Aufdenkampe (Stroud Water Research Center); Matthew E. Baker (U. of Maryland, Baltimore County); Michelle Baker (USU); Christina Banderagoda (Silver Tip Solutions); Phaedra Budy (USU/USGS); Steven Corbato (U. of Utah); Alva Couch (Tufts U.); Todd Crowl (USU); Ernesto de la Hoz (USU); Jeff Dozier (U. of California, Santa Barabara); Jim Ehleringer (U. of Utah); Joanna Endter-Wada (USU); Susan Gill (Stroud Water Research Center); Jonathan L. Goodall (U. of South Carolina); Michael Halling (USU); Richard Hooper (Consortium of Universities for the Advancement of Hydrologic Science, Inc.); Ray Idaszak (RENCI); Douglas Jackson-Smith (USU); Norm Jones (Brigham Young U.); Kerstin Lehnert (Columbia U.); David R. Maidment (U. of Texas at Austin); Emilio Mayorga (U. of Washington); Jamie P. McEvoy (USU); Nancy O. Mesner (USU); William Michener (U. of New Mexico); Reagan Moore (RENCI); Bethany T. Neilson (USU); Sarah Null (USU); Fred L. Ogden (U. of Wyoming); Michael Piasecki (The City College of New York); Diane Pataki (U. of Utah); 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 (U. of Texas at Austin); Thomas Whitenack (San Diego Supercomputer Center); Mark Williams (U. of Colorado Boulder); Ilya Zaslavsky (San Diego Supercomputer Center)

Graduate Advisors

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 - (1) MS, (0) PhD, and (0) Postgraduate Scholars

Brant Whiting, Utah State University, Logan, UT (current MS student)

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

PRODUCTS RELEVANT 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 PRODUCTS

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

- Member of the NSF Advisory Committee for Cyberinfrastructure (term: 5/2012 – 4/2016)
- Chair of the Geoinformatics Division of the Geological Society of America
- President-elect of the ESSI Focus Group (Earth & Space Science Informatics) of the American Geophysical Union
- Member of the Policy Committee of UCAR's Unidata facility
- Organization of numerous short courses & workshops
    - *"EarthChem Community Workshop", Montreal, Canada, June 2012*
    - *"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.

COLLABORATORS AND OTHER AFFILIATIONS

Lee Allison (Arizona State Geological Survey), Anthony Aufdenkampe (Strouds Water Research Center), Susan Brantley (Pennsylvania State University), Jeffrey Horsburgh (Utah State University), Mark Ghiorso (OFM-Research, Seattle), Anthony Koppers (Oregon State University), Carla Moore (National Geophysical Data Center, NOAA), Michael Mottl (University of Hawaii), Anders Noren (University of Minnesota), David Tarboton (Utah State University), Douglas J. Walker (Kansas University), Mary Whitton (RENCI), Ilya Zaslavsky (UCSD, San Diego Supercomputing Center)

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, Seattle.

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

Haines, S., V. Subramanian, **E. Mayorga**, D. Snowden, R. Ragsdale, C. Rueda and M. Howard. 2012. IOOS vocabulary and ontology strategy for observed properties. *Proc. MTS/IEEE Oceans 2012, Hampton Roads, VA.*

Martin, D.L., J.C. Allan, J. Newton, D.W. Jones, S. Mikulak, **E. Mayorga**, T. Tanner, N. Lederer, A. Sprenger, R. Blair and S.A. Uczekaj. 2011. Using web-based and social networking technologies to disseminate coastal hazard mitigation information within the Pacific Northwest component of the Integrated Ocean Observing System (IOOS). *Proc. MTS/IEEE Oceans'11*

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*, doi:10.1109/OCEANS.2010.5663792

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

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 Products

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

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

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

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

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

(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.

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

Carbon in Tropical Rivers, Organization for Tropical Studies (OTS) NSF Research Coordination Network, 2010 – 2012.

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. *US Integrated Ocean Observing System (IOOS), Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI), Critical Zone Observatories (CZO), EPA Network Exchange (Pacific NW Water Quality Data Exchange), Ocean Biogeographic Information System (OBIS).*

(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
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

E M M A L . A R O N S O N , P H . D .

UC Irvine, 321 Steinhaus Hall, Irvine CA 92697;
earonson@uci.edu; 267.738.4855

(i) Professional Preparation

University of Pennsylvania, Philadelphia, PA — September 2005 – August 2011
Ph.D., Department of Biology, GPA 3.92
Ecology and Evolution, concentration in Microbial Ecology & Atmospheric Chemistry

- Emphasis on the soil microbial methane-cycling population and atmospheric impacts.

McGill University, Montreal, Quebec, Canada — September 2000 – June 2004
B.Sc. Degree with "Great Distinction" (equivalent Magna Cum Laude)
Environmental Science Major, concentration in Biodiversity & Conservation; English minor

- Thesis evaluated long-term effects of monoculture vs. diversity on forest carbon uptake.

(ii) Appointments

Research Specialist September 2011 – Present
NOAA Climate and Global Change Fellow at UC Irvine with Dr. Steven Allison,

- Field investigation of climate change impacts on GHG flux and soil microbial communities.

Assistant Professor July 2013+
University of California, Riverside, Department of Plant Pathology and Microbiology

(iii) Publications:

Peer Reviewed Publications:

Aronson, EL, EA Dubinsky, & BR Helliker. 2013. "Nitrogen amendment effects on total soil microbial biodiversity and functional capacity for methane-cycling depend on drainage conditions in a pine forest soil." In press by Soil Biology and Biochemistry.

Aronson, EL, DR Vann & BR Helliker. 2012. "Methane flux response to nitrogen amendment in an upland pine forest soil and riparian zone." Journal of Geophysical Research – Biogeosciences. 117, G03012, doi:10.1029/2012JG001962.

Aronson, EL & SD Allison. 2012. "Meta-analysis of environmental impacts on nitrous oxide release in response to N amendment." Frontiers in Terrestrial Microbiology. 3, 272.

Aronson, EL & SG McNulty. 2010. "Response to comment on: Appropriate experimental ecosystem warming methods by ecosystem, objective, and practicality." Agricultural and Forest Meteorology. 150, 499-500.

Aronson, EL & BR Helliker. 2010. "Methane flux in non-wetland soils in response to nitrogen addition: a meta-analysis." Ecology. 91, 3242–3251.

Aronson, EL & SG McNulty. 2009. "Appropriate experimental ecosystem warming methods by ecosystem, objective, and practicality." Agricultural and Forest Meteorology. 149, 1791–1799.

(iv) Collaborators

None

BIOGRAPHICAL SKETCH
Susan E. Gill

Address: Stroud Water Research Center, 970 Spencer Road, Avondale, PA 19311
Communication: Tel: (610-268-2153 (x297) FAX: (610-268-0490)
e-mail: sgill@stroudcenter.org

(i) Professional Preparation

University of Pennsylvania	Landscape Architecture	B.A., 1976
Arcadia University	Environmental Education	M.A., 1989
University of Pennsylvania	Geology	Ph.D., 1996

(ii) Appointments

Director of Education, April 2007-Present. Stroud Water Research Center, PA, USA.

Director of Professional Education, July 2003-March 2007, Department of Earth and Environmental Science, University of Pennsylvania, PA, USA

Executive Director, July 2003-March 2007, Institute for Environmental Studies, University of Pennsylvania, PA, USA

Director*, IT@Penn* 2000-2001, University of Pennsylvania, PA, USA

Assistant Dean, Dec. 1996-June 2003, University of Pennsylvania, PA, USA.

Director, PennAdvance Distributed Learning Program, 2002. University of Pennsylvania, PA, USA.

(iii) Selected Products

Gill, S. E., N. Marcum-Dietrich, (in review). Model My Watershed: Contextualizing Problem-based and Place-based Geoscience Learning within a Sociocultural Framework. J. Geosci. Educ.

Gill, S. E., Marcum-Dietrich, N. & Fraser, J. (2013). Developing a web application for the integration of real-world, scientific, problem-solving into the secondary classroom. In A. D. Ritzhaupt & Swapna Kumar (Eds.) *Cases on Educational Technology Implementation for Facilitating Learning.* Hershey, PA: IGI Global.

Reiter, M. A., Focht, W. J., Barresi, P. A., **Gill**, **S. E.**, Smardon, R. C., Baker, S., Reiter, K., Fitch, E., Rolfe, T., and Bumpous ,S. (2012). Making education for sustainability work on campus: The proposals of the Roundtable on Environmental Systems and Sustainability. In: W. Leal (Ed.) *Sustainable Development at Universities: New Horizons*. Peter Lang Scientific Publishers, Frankfurt Germany.

Marcum-Dietrich, N., Marquez, L., **Gill, S.E**., and Medved C. 2011. No Teacher Left Inside: Preparing a New Generation of Teachers. J. Geosci. Educ. **59**, 1.

(iv) Synergistic Activities

Member.

- Advisory Council, Geography and Planning Department, West Chester University.
- Education Committee, Schuylkill Center for Environmental Education, 2011-present.
- Panel on the Evaluation of Interdisciplinary Environmental Programs, 2007-present.

Instructor.

- Advisor to 150 students. Master of Environmental Studies Program, University of Pennsylvania, PA, USA. 1996-2007.

- Department of Earth and Environmental Science, University of Pennsylvania, PA, USA. 1996-2009.
- Chemistry and Physical Science Department, Philadelphia University**,** Philadelphia, PA, USA. 1993-present.
- PATHS/PRISM Science Resource Leader Program, The Philadelphia Partnership for Education**,** Philadelphia, PA, USA. 1993

Environmental Impact Planner, Writer and Editor,

- Portfolio Associates, Philadelphia, PA, USA. 1985
- McCormick, Taylor & Associates**,** Philadelphia, PA, USA. 1982-1984

Community Organizer and Planner. Eastwick Project Area Committee and Washington Square West Project Area Committee, Philadelphia, PA. 1980-1982.

Assistant Project Director, Schuylkill River Study, Pennsylvania Environmental Council**,** Philadelphia, PA, USA. 1975-77

(v) Collaborators

(a) Collaborators and Co-Authors

Jodi Asbel-Clark	EdGE@TERK, Cambridge, MA, USA
Anthony Aufdenkampe	Stroud Water Research Center, PA, USA
Nanette Dietrich	Educational Foundations Department, Millersville
David Dunbar	Science Department, Cabrini College, PA, USA
Louis Kaplan	Stroud Water Research Center. PA, USA
Delphis F. Levia	Department of Geography, University of Delaware, DE, USA
Lyn Marquez	Department of Earth Sciences, Millersville University of PA,
J. Denis Newbold	Stroud Water Research Center, PA, USA
Caroline Nielson	Science Department, Cabrini College, PA. USA
Barbara Riebling	Department of English, University of Pennsylvania, PA, USA
Carolyn Staudt	Concord Consortium, Concord, MA, USA PA, USA.
Janet Theophano	Center for Folklore and Ethnography, University of Pennsylvania, PA, USA.
Robert Tinker	Concord Consortium, Concord, MA, USA
C. Dana Tomlin	Department of Landscape Architecture, University of Pennsylvania, PA, USA
Mylissa Turlecki	Psychology Department, Cabrini College, PA, USA University of PA, PA, USA
K. Yemane	Currently unaffiliated

(b) Dissertation Advisor

A. H. Johnson	Department of Earth and Environmental Science, University of Pennsylvania, PA, USA.

CHRISTOPHER S. HENRY: BIOGRAPHICAL SKETCH

(a) Professional Preparation

Institutions	Major/Area	Degree	Years
Undergraduate: University of Dayton	Chemical Engineering	BS	1998-2002
Graduate: Northwestern University	Chemical Engineering	PhD	2002-2007
Postdoc: Argonne National Laboratory	Computational Biology	PhD	2007-2009

(b) Appointments

2012-	Adjunct Professor, Chemical and Biological Engineering, Northwestern University
2009-	Fellow, Computation Institute, University of Chicago
2009-	Assistant Computer Scientist, Argonne National Lab
2007-2009	Postdoctoral Scholar, Mathematics and Computer Science, Argonne National Lab
2002-2007	Graduate Student, Department of Chemical Engineering, Northwestern University

(c) PRODUCTS **5 most closely related to proposed project**

Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL (2010) High-throughput generation, optimization, and analysis of genome-scale metabolic models. *Nature Biotechnology* **Nbt.1672**: 1-6.

Freilich, S, Zarecki, R, Eilam, O, Segal, ES, Henry, CS, Kupiec, M, Gophna, U, Sharan, R, and Ruppin E (2011). Competitive and Cooperative Metabolic Interactions in Bacterial Communities. *Nature Communications* 2:589.

Henry CS, Zinner JF, Cohoon MP, Stevens RL (2009) iBsu1103: a new genome-scale metabolic model of *Bacillus subtilis* based on SEED annotations. *Genome Biology* **10:** R69

Henry CS, Overbeek R, Xia F, Best AA, Glass E, Gilbert J, Larsen P, Edwards R, Disz T, Meyer F, Vonstein V, DeJongh M, Bartels D, Desai N, D'Souza M, Devoid S, Keegan K, Olson R, Wilke A, Wilkening J, Stevens RL (2011) Connecting Genotype to Phenotype in the Era of High-throughput Sequencing. *BBA - General Subjects 1810 10).*

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BØ (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1261 ORFs and thermodynamic information. *Molecular Systems Biology* **3:** 121.

Other significant publications

Henry CS, Xia F, Stevens RL (2009). Application of high-performance computing to the reconstruction, analysis, and optimization of genome-scale metabolic models. *Journal of Physics: Conference Series* **180**: 12-25.

Faria JP, Rocha M, Stevens RL, and Henry CS (2010) Analysis of the Effect of Reversibility Constraints on the Predictions of Genome-scale Metabolic Models. *Advances in Soft Computing* **74**: 209-215.

Henry CS, Broadbelt LJ, Hatzimanikatis V (2010) Discovery of novel biosynthetic routes to 3-hydroxypropanoate from pyruvate. *Biotechnology and Bioengineering*. 106: 462-473.

Henry CS, Broadbelt LJ, Hatzimanikatis V (2007) Thermodynamics-based metabolic flux analysis. *Biophysical Journal* **92:** 1792-1805.

Henry CS, Jankowski MD, Broadbelt LJ, Hatzimanikatis V (2006) Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophysical Journal* **90:** 1453-1461.

(d) Synergistic Activities

- Honored with the Jay Bailey Young Investigator Best Paper in Metabolic Engineering Award
- Co-lead for the Microbial Science Domain on the DOE Knowledgebase for integration and

standardization of ‘omics data, biological models, and model predictions with Microbes, Plants, and Communities
- Developed Model Seed resource for high-throughput reconstruction, optimization, and analysis of genome-scale metabolic models utilized by numerous scientists and educators (www.theseed.org/models/)
- Session chair for National Meeting of the American Institute of Chemical Engineers
- Referee for PLOS One, Biotechnology and Bioengineering, Biophysical Journal, Nature Biotechnology, and BMC Bioinformatics
- Associate Editor for the Journal 3Biotech
- Trained undergraduate students and masters students at University of Chicago and Hope College in computational biology and bioinformatics research

(e) Collaborators & Other Affiliations

(i) Collaborators & Co-Editors (past 48 months)

Andrews, Beth (Northwestern University); Ann, Gary (University of Chicago); Arkin, Adam (Berkley National Laboratory); Best, Aaron A. (Hope College); Broadbelt, Linda J. (Northwestern University); Chivian, Dylan (Berkley National Laboratory); Christley, Scott (University of Chicago); Colasantic, Ric (University of Chicago); Cottingham, Bob (Oak Ridge National Laboratory); de Crécy-Lagard, Valérie (University of Florida); Dehal, Paramvir (Berkley National Laboratory); DeJongh, Matt (Hope College); Devoid, Scott (University of Chicago); Edwards, Robert (San Diego State University); Ehrlich, Dusko S. (INRA, France); Feist, Adam (Univeristy of California San Deigo); Folker, Meyer (University of Chicago); Gerdes, Svetlana (Fellowship for Interpretation of Genomes); Gilbert, Jack (University of Chicago); Glass, Elizabeth M. (University of Chicago); Hanson, Andrew (University of Florida); Hatzimanikatis, Vassily (EPFL, Switzerland); James Faeder (University of Pittsburgh); Jozce, Andrew R. (SRI International); Karp, Peter (SRI International); Krummenacker, Markus (SRI International); Maslov, Sergei (Brookhaven National Laboratory); Noirot, Phillipe (INRA, France); Novichkov, Paval (Burnham Institute); Olsen, Gary (University of Illinois Urbana Champagne); Osterman, Andrei L (Sanford-Burnham Medical Research Institute); Overbeek, Ross (Fellowship for Interpretation of Genomes); Palsson, Bernhard O. (Univeristy of California San Deigo); Pedro, Jose (University of Chicago); Price, Nathan (University of Illinois Urbana Champagne); Pusch, Gordon (Fellowship for Interpretation of Genomes); Reed, Jennifer L. (Univeristy of Wisconsin); Rodionov, Dmitry A (Sanford-Burnham Medical Research Institute); Saur, Uwe (ETH Zurich, Switzerland); Seaver, Sam (Argonne National Laboratory); Sorbal, Bruno (Virginia Biotechnology Institute); Stevens, Rick L. (University of Chicago); Thierry Emonet (Yale University); Tomb, Jean-francois (DuPont); Triplett, Eric W. (University of Florida); Vassieva, Olga (University o f Liverpool, United Kingdom); Vonstein, Veronica (Fellowship for Interpretation of Genomes); Ware, Doreen (Cold Spring Harbour Lab); Westin, Davin (Oak Ridge National Laboratory); and Xia, Fangfang (Argonne National Laboratory)

(ii) Graduate Advisors & Postdoctoral Sponsors: Graduate – LJ Broadbelt (Northwestern University, Evanston IL) and V Hatzimanikatis (EPFL, Lausanne, Switzerland); Postdoctoral – RL Stevens (ANL, Lemont, IL)

Career Totals: 4 post docs, 4 undergraduate students, 6 graduate students

Folker Meyer

i. Professional Preparation

Bielefeld University	Computer Science	M.S. 1996
Bielefeld University	Bioinformatics	Ph.D. 2002

ii. Appointments

2008 -	Assoc. Division Director, Institute for Genomics and Systems Biology, Argonne National Laboratory and University of Chicago
2006 -	Senior Fellow, Computation Institute, University of Chicago
2006 -	Scientist, Mathematics and Computer Science Division, Argonne National Laboratory
2003 - 2006	Head Bioinformatics, Resource Facility of the Center for Biotechnology, Bielefeld University
2000 - 2003	Head Bioinformatics, BMBF Competence Network, Bielefeld University
1998 -1999	IIT-BioTech GmbH, Bielefeld University
1997	Scientist, Center for Genome Research (ZfGB), Bielefeld University
1991-1996	Department of History and Philosophy, Bielefeld University

iii. Products

Five Most Relevant

1. J. A. Gilbert, F. Meyer, M. J. Bailey ,
   The Future of microbial metagenomics (or is ignorance bliss?)
   The ISME Journal , (25 November 2010) | doi:10.1038/ismej.2010.178, 2010
2. A. Wilke, J. Wilkening, E. M. Glass, N. L. Desai, and F. Meyer
   Porting the MG-RAST metagenomic data analysis pipeline to the cloud
   Concurrency and Computation2011, DOI: 10.1002/cpe.1799
3. J. Wilkening, A. Wilke, N. Desai, F. Meyer
   Using Clouds for Metagenomics: A Case Study
   IEEE International Conference on Cluster Computing (Cluster 2009), AUG 31-SEP 04, 2009 New Orleans, LA 2009
4. F. Meyer, D. Paarmann, M. D'Souza, R. Olson, E. M. Glass, M. Kubal, T. Paczian , A. Rodriguez, R. Stevens, A. Wilke, J. Wilkening, R. A. Edwards
   The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes
   BMC Bioinformatics 2008, 9:386.
5. F. Meyer.,
   "Genome Sequencing vs. Moores Law: Cyber Challenges for the Next Decade,"
   Cyberinfrastructure Technology Watch, 2006.

Five Other Relevant Publications

1. R. Overbeek, D. Bartels, V. Vonstein, F. Meyer, "Annotation of bacterial and archaeal genomes: improving accuracy and consistency, " Chem Rev. 2007 Aug;107(8):3431-47.
2. D. Field, L. Amaral-Zettler, G, Cochrane, J. R Cole, P. Dawyndt, G. M Garrity, J. Gilbert, F. O. Glöckner, L. Hirschman, I. Karsch-Mizrachi, H.-P. Klenk, R. Knight, R. Kottmann, N. Kyrpides, F. Meyer, I. San Gill, S.-A. Sansone, L. M Schriml, P. Sterk, T. Tatusova, D. W Ussery, O. White, J. Wooley
   The Genomic Standards Consortium
   PLoS Biology2011 June; 9(6): e1001088.

3. R. K. Aziz, D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko (2008)
The RAST Server: Rapid Annotations using Subsystems Technology
BMC Genomics, 2008 Feb 8;9(1):75
4. F. Meyer, A. Goesmann, A.C. McHardy, D. Bartels, T. Bekel, J. Clausen, J. Kalinowski, B. Linke, O. Rupp, R. Giegerich, A. Pühler (2003)
GenDB--an open source genome annotation system for prokaryote genomes.
Nucleic Acids Res 31(8): 2187-95
5. J. Stoye, D. Evers, F. Meyer (1998)
Rose: generating sequence families.
Bioinformatics 14(2): 157-63

iv. Collaborators and Other Affiliations

a. Collaborators and Co-Editors

A. Puhler (U of Bielefeld); R. Giegerich (U of Bielefeld); H. Kuster (U of Bielefeld); S.C. Schuster (Penn State); A.C. McHardy (Max-Planck-Institute); R. Overbeek (The Fig); A. Becker (U of Bielefeld); R. Reinhardt (Max-Planck-Institute); R.I. Amann (Max-Planck-Institute), R. Stevens (Argonne National Labs), K. P. White (University of Chicago)

b. Advisors

Robert Giegerich/Bielefeld University, Computer Science; Alfred Pühler/Bielefeld University, Biology

c. *Students*

Daniela Bartels, Thomas Bekel, Martin Bennemann, Michael Dondrup, Alexander Goesmann, Keywan Hassani-Pak, Julia Herold, Sebastian Kespohl, Nils Kleinboetting, Jan Kleinlützum, Thomas Kohl, Sebastian Konietzny, Lutz Krause, Burkhard Linke, Alice McHardy, Heiko Neuweger, Sebastian Oehm, Jörg Ostermann, Daniel Paarmann, Katharina Pernhorst, Jan Pieniak, Oliver Rupp, Robert Schmieder, Daniel Wetter, Jared Wilkening, Andreas Wilke, Martha Zakrzewski,

Biographical Sketch: Aaron I. Packman

Department of Civil and Environmental Engineering and Department of Earth and Planetary Sciences
Northwestern University

(i) Professional Preparation

B.S. Mechanical Engineering, cum laude. Minor in History. Washington University (St. Louis), 1991.
M.S. Environmental Engineering Science. California Institute of Technology, 1992.
Ph.D. Environmental Engineering Science. Minor in Geology. California Institute of Technology, 1997.

(ii) Appointments

2010-present:	Professor, Northwestern University
2004-2010:	Associate Professor, Northwestern University
2000-2004:	Assistant Professor, Northwestern University
1997-2000:	Assistant Professor, Drexel University
1994-1997:	Graduate Research Assistant, California Institute of Technology.
1991-1994:	ONR/NDSEG Fellow, California Institute of Technology.

(iii) Products- Five Most Closely Related to this Project

Zhang, W., Sileika , T., Packman, A.I, 2013, Effects of fluid flow conditions on interactions between species in biofilms, FEMS Microbiology Ecology, DOI: 10.1111/1574-6941.12066.

Zhang, Y., Meerschaert, M.M., and Packman, A.I., 2012, Linking fluvial bed sediment transport across scales, *Geophysical Research Letters*, 39, L20404, doi:10.1029/2012GL053476.

Battin, T.J., Kaplan, L.A., Findlay, S., Hopkinson, C.S., Marti, E., Packman, A.I., Newbold, J.D., and Sabater, F., 2008, Biophysical controls on organic carbon fluxes in fluvial networks, *Nature Geoscience*, 1, 95-100, doi:10.1038/ngeo101.

Wörman, A., Packman, A.I., Marklund, L., Harvey, J.W., and Stone, S.H., 2007, Fractal topography and subsurface water flows from fluvial bedforms to the continental shield, *Geophysical Research Letters*, 34, L07402, doi:10.1029/2007GL029426.

Arnon, S., Gray, K.A., and Packman, A.I., 2007, Biophysicochemical process coupling controls nitrogen utilization by benthic biofilms, *Limnology and Oceanography*, 52(4), 2007, 1665-1671.

FIVE OTHER SIGNIFICANT PUBLICATIONS

Stonedahl, S.H., Harvey, J.W., Detty, J., Aubeneau, A., and Packman, A.I., 2012, Physical controls and predictability of stream hyporheic flow evaluated with a multi-scale model, *Water Resources Research*, 48, doi:10.1029/2011WR011582.

Larned, S.T., Packman, A.I., Plew, D.R., and Vopel, K., 2011, Interactions between the mat-forming alga *Didymosphenia geminata* and its hydrodynamic environment, Limnology and Oceanography: Fluids and Environments, 1, 4–22, DOI: 10.1215/21573698-115208

Boano, F., Packman, A. I., Cortis, A., Revelli, R., and Ridolfi, L., 2007, A continuous time random walk approach to the stream transport of solutes, *Water Resources Research*, 43, W10425, doi:10.1029/2007WR006062.

Searcy, K.E., Packman, A.I., Atwill, E.R., & Harter, T., 2006, The capture and retention of *Cryptosporidium* oocysts in *Pseudomonas aeruginosa* biofilms, *Applied and Environmental Microbiology*, 72, 6242–6247 (Cover article).

Packman, A.I, and MacKay, J.S. 2003, Interplay of stream-subsurface exchange, clay particle deposition, and stream bed evolution, *Water Resources Research*, 39(4), 1097, doi:10.1029/2002WR001432.

(IV) SYNERGISTIC ACTIVITIES

I have regularly convened sessions at AGU and other technical meetings, primarily with the goal of encouraging interdisciplinary study of earth and environmental systems. I have substantially integrated my educational and research activities, including a) direct supervision of undergraduate students and visiting international scholars, and b) frequent use of research laboratories for tours and demonstrations to encourage pre-college students to become scientists and engineers. I explored pedagogical methods to improve synthesis of research and teaching activities

during my tenure as a Searle Junior Teaching Fellow, and have helped to develop a variety of innovative teaching and curricular programs at Northwestern (Environmental Sciences Program, Murphy Scholars Program, *et al.*)

(V) COLLABORATORS

J.F. Gaillard, N.P. Cianciotto, R.T. Thomson, N. Cianciotto, D. Chopp, G. Buscarnera, A. Jacobson & D.T. Keane (NU), J. Best, M.H. Garcia, K. Christensen, M. Sivapalan, & J.S. Wilson (UIUC), L.G. Larsen & S.E. Thompson (Berkeley), C.R. Tobias (UConn), A.H. Sawyer (Kentucky), D. Bolster and J. Tank (Notre Dame), C.J. Harman (Johns Hopkins), N.B. Basu (Waterloo), P.A. Troch & P.D. Brooks (Arizona), P.S.C. Rao (Purdue), T. Covino & R. Payn (Montana State), B. McGlynn (Duke), R. Schumer & Y. Zhang (DRI), D.J. Jerolmack & R.L. Martin (Penn), M.B. Cardenas (Texas), S.A. Bradford (USDA), V.L. Morales (Cornell), M.M. Meerscaert and W. Zhang (MSU), A. Mohanram (Hawaii), C. Welty (UMBC), M. Hausner (Ryerson), S. Larned, C. Kilroy, R. Davies-Colley, R. Stott, & D. Plew (NIWA, New Zealand), K. Vopel (Auckland Univ. Tech.), M. Bothwell (Env. Canada), F. Boano, L. Ridolfi & R. Revelli (Politecnico di Torino, Italy), C. Peterson & J. Kelly (Loyola), P. Pasten & G. Pizarro (Católica Univ, Chile), A. Aufdenkampe, L.A. Kaplan & J.D. Newbold (Stroud Water Research Center), D. Scott (VaTech), M. Parsek & B.-S. Tseng (U. Washington), G.A. Burton, Jr. (Michigan), J. Cullis (S. Africa), C.-A. Gillis (INRS) R.W. Harvey, J.W. Harvey, L.E., J. Detty & L.E. McPhillips (USGS), S. Patil (EPA), M.A. Hassan and S.D. Donner (UBC), A. Marion and M. Bellinello (U. Padova, Italy), T. Battin (U. Vienna, Austria), T. Harter, E.R. Atwill, S. Wuertz, L. Hou, P.G. Green, S. Cook, Y. Park, L. Hou, X. Li, & N. Watanabe (UC-Davis), A. Wörman (KTH).

ADVISEES (5 post-docs, 13 Ph.D., 8 M.S., 8 visiting scholars, plus 29 undergrads not listed individually)

Current Ph.D. Students: A. Aubeneau, A. Culotti, J. Drummond, J. Lee, X. Li, K. Roche, Postdocs: W. Zhang, M.S.: A. Ambekar, N. Chiangwong, A. Iyer, Visiting Scholars: Y. Bai (SW Jiaotong Univ.), M. Xie (SW Jiaotong Univ.)

Former Ph.D. Students: C. Chen (Haliburton), J. Ren (Texas A&M Univ.), R. Ryan (Temple Univ.), M. Salehin (BUET), K. (Searcy) Bell (Environ), S. Stonedahl (St. Ambrose Univ.), S. Waller (Sustainable Systems LLC), Post-Docs: S. Arnon (Ben Gurion Univ.), B. Lau (Baylor Univ.), Y. Liu (Univ. Alberta), W. Zhang (Northwestern Univ.), M.S.: M. Barnes (Applikon Biotechnology), J. Mackay (NTM Engineering), L. Marx (Abbott Labs), J. Miceli (NextEra Energy), R. Rajbanshi (Port Authority of NY & NJ), Visiting Scholars: F. Boano (Politecnico di Torino), D. Giuliani (Agip KCO), D. Jerolmack (Univ. Penn.), D. Peng (SW Jiaotong Univ.), G. Singer (Univ. Vienna), M. Zaramella (WET Engineering)

GRADUATE ADVISORS:

Norman H. Brooks and James J. Morgan, Dept. of Environmental Engineering Science, Caltech

ROELOF J. VERSTEEG

Subsurface Insights
62 Lebanon Street
Hanover, NH 03755

Telephone: Office (603) 443 2202
Home (802) 280-7161
roelof.versteeg@subsurfaceinsights.com

(i) Professional Preparation

Ph.D.	Geophysics	1991	University of Paris 7 (France)
M.S.	Sedimentology	1988	University of Utrecht (Netherlands)
M.S.	Geophysics	1988	University of Utrecht (Netherlands)
B.S.	Geophysics	1984	University of Utrecht (Netherlands)

(ii) Appointments

7/12-present Owner and chief scientist, Subsurface Insights, Hanover, NH
4/10-7/12 Senior Research Scientist, Sky Research, Etna, NH
11/01-4/10 Scientist IV, Idaho National Laboratory, Idaho Falls, ID
3/07-8/07 Visiting scientist, TNO Netherlands (currently Deltares) (3/2007-8/2007)
9/95-10/01 Associate Research Scientist/Adjunct Assistant Professor, Lamont Doherty Earth Observatory, Columbia University, NY, NY
9/94-8/95 Visiting Assistant Professor, Geology Department, University of Connecticut
6/92-7/94 Post Doctoral Scientist, Rice University, Houston, TX \
1/92-6-92 Post Doctoral Scientist, Amoco Research,Tulsa, OK

(iii) Products

Aziz, Z., A. Van Geen, R. Versteeg, H. A., Y. Zheng, G. S., M. Steckler, M. Stute, B. Weinman, G. I., H. M.A., S. M and A. K.M. (2008). "Impact of local recharge on arsenic concentrations in shallow aquifers inferred from the electromagnetic conductivity of soils in Araihazar, Bangladesh. ." Water Resour Res **44**(Water Resour. Res., 44, W07416, doi:10.1029/2007WR006000).

Few, D., R. Versteeg and H. Herman (2010). Semi Autonomous Mine Detection System. 2010 SPIE Defense, Security and Sensing Meeting. Conference 7664: Detection and Sensing of Mines, Explosive Objects and Obscured Targets XV, Orlando, Florida.

Glaser, D. R., D. D. Werkema, R. J. Versteeg, R. D. Henderson and D. F. Rucker (2012). "Temporal GPR imaging of an ethanol release within a laboratory-scaled sand tank (in press)." Journal of Applied Geophysics.

Johnson, T., R. Versteeg, H. Huang and P. Routh (2009). "A data domain correlation approach for joint inversion of time-lapse head, concentration, and electrical resistivity data." Geophysics **74**(F127 - doi:10.1190/1.3237087).

Johnson, T., R. Versteeg, M. Rockhold, L. Slater and J. Zachara (2012). "Characterization of a contaminated wellfield using three-dimensional electrical resistivity tomography implemented with geostatistical, discontinuous boundary, and known conductivity constraints." Geophysics (in press).

Johnson, T., R. Versteeg, A. Ward, F. Day-Lewis and A. Revil (2010). "Improved hydrogeophysical characterization and monitoring through parallel modeling and inversion of time-domain resistivity and induced polarization data." Geophysics **75**: 27-41.

Slater, L. D., D. Ntarlagiannis, F. Day-Lewis, K. Mwakanyamale, R. J. Versteeg, A. Ward, C. Strickland, C. D. Johnson and J. Lane (2010). "Characterization of Surface Water/Groundwater Exchange Regulating Uranium Transport Using Electrical Imaging and Distributed Temperature Sensing Methods." Water Resources Research **46**(W10533).

Slater, L. D., D. Ntarlagiannis, F. D. Day-Lewis, K. Mwakanyamale, R. J. Versteeg, A. Ward, C. Strickland, C. D. Johnson and J. W. L. Jr. (2010). "Use of electrical imaging and distributed

temperature sensing methods to characterize surface water–groundwater exchange regulating uranium transport at the Hanford 300 Area, Washington." WATER RESOURCES RESEARCH 46(W10533).
Versteeg, R., A. Richardson and T. Rowe (2006). "Web accessible scientific workflow system for performance monitoring." Environmental Science and Technology 10.1021/es0517421 S0013-936X(05)01742-6.
Versteeg, R. and T. Johnson (2008). "Using time-lapse electrical geophysics to monitor subsurface processes." The Leading Edge **27**(11): 1448-1497.
Versteeg, R., M. McKay, M. Anderson, R. Johnson, B. Selfridge and J. Bennett (2007). Feasibility study for an Autonomous UAV -Magnetometer system, Idaho National Laboratory - INL/EXT-07-13386 55 pages.

(iv) Collaborators

Tim Johnson, Pacific Northwest National Laboratory
John Zachara, Pacific Northwest National Laboratory
John Lane, United States Geological Survey
Fred Day Lewis, United States Geological Survey
Lee Slater, United States Geological Survey
Ken Williams, Lawrence Berkeley National Laboratory
Susan Hubbard, Lawrence Berkeley National Laboratory

Ilya Zaslavsky

San Diego Supercomputer Center
University of California, San Diego
9500 Gilman Drive, MC 0505
La Jolla, CA 92093-0505

Telephone: 858 534-8342
Fax: 858 534-5113
E-mail: zaslavsk@sdsc.edu

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 Products

[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., T. Whitenack, M. Williams, D. G. Tarboton, K. Schreuders, and A. Aufdenkampe, (2011) “The Initial Design of Data Sharing Infrastructure for the Critical Zone Observatory”, in *Proceedings of the Environmental Information Management Conference*, Santa Barbara, CA, 28-29 September, EIM'2011, pp. 145-150.

[3] 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.

[4] 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).

[5] 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.

Other Significant Products

[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

- PI, NSF EarthCube Cross-Domain Interoperability Testbed Concept Award (http://earthcube.ning.com/groups/interop/)
- PI or Co-PI on multiple projects, with funding from NSF, NIH, US Department of State, state and local sources, foundations, companies, and international agencies. Key projects: CUAHSI Hydrologic Information System; Critical Zone Observatories; Biomedical Informatics Research Network; Chesapeake Bay Environmental Observatory; US-EU Cyberinfrastructure for Hydrometeorology; Superfund Research Program; Digital Brain Atlasing Infrastructure; Streaming Data Middleware; Digital Government
- Co-Chair, Hydrology Domain Working Group of the 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)

Collaborators

Chaitanya Baru (SDSC), David Maidment (UT Austin), Maryann Martone (UCSD), Keith Pezzoli (UCSD), Michael Piasecki (CUNY), David Tarboton (USU), Jeffrey Horsburgh (USU), Richard Hooper (CUAHSI), Kerstin Lehnert (Columbia U.), Alva Couch (Tufts U.), Stephen Richard (AZ Geological Survey), Philip Murphy (U. of Redlands), Beth Plale (Indiana U.), Ben Domenico (UNIDATA), Jerry Brotzge (Oklahoma U.), David Arctur (OGC), Ilkay Altintas (SDSC), Mark Williams (U. of Colorado), Anthony Aufdenkampe (Stroud Water Res. Center)

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

David W Valentine

Research Programmer
San Diego Supercomputer Center
University of California at San Diego
La Jolla, CA 92093-0505

Telephone: (858) 822-0923
Fax: (858) 534-5077
E-mail: valentin@sdsc.edu

Professional Preparation

San Diego State University San Diego, California	B.S. Engineering Geology,	1988
Humboldt State University Arcata, California	M.S. Environmental Engineering	1992
University of California Santa Barbara, CA	Ph.D. Geological Sciences	2004

Appointments

2005	Research Programmer, San Diego Supercomptuer Center
2000	System Architect, Alexandria Digital Library, Davidson Library, University of California, Santa Barbara
1999	GIS Analyst, US Geological Survey. Water Resources Division, Sacramento, California.
1998	Hydrographer, Kuwait Hydrographic Project. Tetra Tech, Inc.
1994	Research Assistant. Institute for Crustal Studies, University of California, Santa Barbara.

Related Products

[1] Peter Taylor, Simon Cox, Gavin Walker, David Valentine, Paul Sheahan. In press, 2013. WaterML2.0: Development of an open standard for hydrological time series data exchange. Journal of Hydroinformatics.

[2] Goodall, Jonathan, Valentine, David, Maidment, David. 2006. Exposing hydrologic databases using web services. Proceedings of the Seventh International Conference on Hydroscience and Engineering, Philadelphia, PA, September 2006. http://hdl.handle.net/1860/732

[3] I. Zaslavsky, D. Valentine, R. Hooper, M. Piasecki, A. Couch, A. Bedig. 2012. Community Practices for Naming and Managing Hydrologic Variables in Fox, Sandra (Editor), Geographic Information Systems (GIS) and Water Resources VII. AWRA's 2012 Spring Specialty Conference. American Water Resources Association, Middleburg, Virginia, TPS-12-1, CD-ROM. Online: http://www.awra.org/proceedings/Spring2012/doc/abs/IlyaZaslavsky_51e7c422_7956.pdf.

[4] Ilya Zaslavsky, David Valentine, Tim Whiteaker. 2007. CUAHSI WaterML.Open Geospatial Consortium Discssion Paper, no. 07-041r1. [URL: http://portal.opengeospatial.org/files/?artifact_id=21743] (accessed July 2, 2009).

[5] Valentine, David, Ilya Zaslavsky, Thomas Whitenack, and David R. Maidment. 2007. Design and implementation of CUAHSI WaterML and WaterOneFlow web services; geoinformatics 2007; data to knowledge. Scientific Investigations Report SIR 2007-5199, (2007): 81-83, [URL:http://pubs.usgs.gov/sir/2007/5199/] (accessed July 2, 2009).

Other Significant Products

[1] Valentine, D.W., Keller, E.A, Carver, G.A, Li, W., Manhart, C., and Simms, A.R. 2012. Paleoseismicity of the Southern End of the Cascadia Subduction Zone, Northwestern California,USA. Bulletin of the Seimological Society of America, v102, No. 3, pp. 1059–1078. doi:10.1785/0120110103 http://hdl.handle.net/10.1785/0120110103

[2] Valentine, D.W., Densmore, J.N.., and Galloway, D.L., and Amelung, F., 2001, Use of InSAR to identify displacements cause by aquifer system compaction in the Paso Robles area, San Luis Obispo county, California, March to August 1997. USGS Open-File Report 00-447.

[3] Keller, E.A., Valentine, D.W., and Gibbs, D.R., 1997. Hydrogeologic response of small watersheds following the California Painted Cave fire of 1992. HydrologicalProcesses, v.11(4) p.401-414..

[4] Carver, G. A., Jayko, A. S., Valentine, D. W., and Li, W. H.., 1994, Coastal Uplift Associated with the 1992 Petrolia Earthquakes, Northern California, Geology, v.22 p.195-198.

[5] Oppenheimer, D., Beroza, G., Carver, G., Dengler, L., Eaton, J., Gee, L., Gonzalez, F., Jayko, A., Li, W. H., Lisowski, M., Magee, M., Marshall, G., Murray, M., R. McPherson, B., Romanowicz, Satake, K., Simpson, R., Somerville, Stein, R., and Valentine, D., 1993, The Cape Mendocino, California Earthquake Sequence of April, 1992 - Subduction at the Triple Junction: Science, v. 261, p. 433-438.

Synergistic Activities

- Co-Chair. OGC WaterML 2 standards working group
- Author, CUAHSI WaterML specification (http://www.opengeospatial.org/standards/dp)
- Coordinator, Environmental Database Model, (http://envirodb.codeplex.com/)
- Technical Lead for National Digital Information Infrastructure and Preservation Program at Map and Imagery Library, University of California, Santa Barbara (2004-2005)

Collaborators & Other Affiliations

Collaborators

Couch, Alava (Tufts)
Goodall, Jonathan (University of South Carolina)
Hooper, Richard (CUAHSI)
Horsbberg, Jeff (Utah State University)
Idaszak, Ray (RENCI)
Keller, Edward (University of California, Santa Barbara)
Maidment, David (University of Texas, Austin)
Simms, Alex (University of California, Santa Barbara)
Tarboton, David (Utah State University)
Whiteaker, Tim (University of Texas, Austin)
Whitenack, Thomas (University of California, San Diego)
Zaslavsky, Ilya (University of California, San Diego)

Graduate and Post Doctoral Advisors:

Gary Carver, Humboldt State University (Emeritus)
Edward Keller, University of California, Santa Barbara

Biographical Sketch – Stephen M Richard

Arizona Geological Survey
416 W. Congress, #100
Tucson, AZ 85701

Tel: 520.209.4127
Fax: 520.770.3505
steve.richard@azgs.az.gov

Education:

Mass. Inst. of Technology, Cambridge, MA	Earth and Planetary Science; Electrical Engineering	B.S. 1978
University of Arizona, Tucson, AZ	Geosciences	M.S. 1983
University of California, Santa Barbara, CA	Geology	Ph.D. 1988
University of California, Santa Barbara, CA	Geology	PostDoc, 1988-1992

Professional Experience:

Senior Geologist, Arizona Geological Survey	2010-present
Chief, Geoinformatics Section, Arizona Geological Survey	2006-present
Research Geologist, Arizona Geological Survey	1992-2010
Adjunct Professor of Geosciences, University of Arizona	1992-present
Post doctoral researcher, Institute for Crustal Studies, Santa Barbara, CA	1988-1992
Electrical Engineer, Lawarence Livermore Laboratory	1978-1979

Products:

1. USGIN Standards and Protocols Drafting Team, 2011, USGIN URI Policies: Tucson, Az. Geo. Survey, accessed at http://repository.usgin.org/uri_gin/usgin/dlio/331.
2. Richard, S. M., compiler, 2010, Technical Contributions of the Geoscience Information Network for Design and Deployment of the National Geothermal Data system: Tucson, Arizona Geological Survey OFR10-04, 103 p.
3. USGIN Standards and Protocols Drafting Team, 2010, Use of ISO19139 xml schema to describe geoscience information resources: Tucson, Arizona Geological Survey OFR10-02, 134 p., accessed at http://repository.azgs.az.gov/uri_gin/azgs/dlio/109.
4. Richard, S.M., and Soller, D.R., 2009, Vocabularies for geoscience information interchange, *in* Soller, David R., Editor, Digital Mapping Techniques '08-Workshop Proceedings: U. S. Geological Survey Open-file Report.
5. Richard, S. M., 2006, Geoscience concept models, in Sinha, K, and Baru, C, Geoinformatics: Geological Society of America Special Paper 397, p. 81-107.

Other significant products:

1. USGS National Cooperative Geologic Mapping Program 2010, NCGMP09—Draft Standard Format for Digital Publication of Geologic Maps, Version 1.1, *in* Soller, D.R., Ed., Digital Mapping Techniques ‘09—Workshop Proceedings: U.S. Geological Survey Open-File Report 2010–1335, accessed at http://pubs.usgs.gov/of/2010/1335/pdf/usgs_of2010-1335_NCGMP09.pdf .
2. Cox, S.J.D., and Richard, S. M., 2005, A formal model for the geologic timescale and GSSP, compatible with geospatial information transfer standards: Geosphere, paper #GEOS-00022R.
3. Richard, S.M., Craigue, J.C., and Soller, D.R., 2004, Implementing NADM C1 for the National Geologic Map Database, *in* Soller, David R., Editor, Digital Mapping

Techniques '04-Workshop Proceedings: U. S. Geological Survey Open-file Report 2004-1451, p. 111-144.
4. North American Geologic Map Data Model Steering Committee (NADM C1), 2004, NADM Conceptual Model 1.0, A Conceptual Model For Geologic Map Information: U.S. Geological Survey Open-file Report 2004-1334, 58 p., http://pubs.usgs.gov/of/-2004/1334/.
5. North American Geologic-Map Data Model Science Language Technical Team (NADM SLTTm), 2004a, Classification of metamorphic and other composite-genesis rocks, Version 1.0 (12/18/2004), in D.R. Soller, ed., Digital Mapping Techniques '04—Workshop Proceedings: U.S. Geological Survey Open-File Report 2004–1451, Appendix B to SLTT report, 56 p., accessed at http://pubs.usgs.gov/of/2004/1451/nadm/.

Synergistic Activities:

1. Chair, Concept Definition Task Group of IUGS CGI Interoperability Working group 2007-present (This group is now a separate CGI Geoscince Terminology Workgroup)
2. Member, Data model team (GeoSciML), IUGS CGI Interoperability Working Group 2004-present; Co-Chair, GeoSciML Standards Working Group, Open Geospatial Consortium. Kick off meeting, Redlands, CA, January 2013
3. Member, Energistics Energy-Industry Metadata Working Group, 2010-present
4. Member, DOE Geothermal Data System Technical Working Group, 2011-present
5. Editor, ISOTC211 Workgroup for XML implementation of ISO19115-1.
6.

Collaborators and other affiliations:

Lee Allison, Arizona Geological Survey; Sky Bristol, U. S. Geological Survey; Ryan Clark, Arizona Geological Survey; David Danko, ESRI; Tamara Dickinson, Office of Science Technology Policy; Alan Doniger, Energistics Inc.; Kevin Gallagher, U.S. Geological Survey; Linda Gundersen, U.S. Geological Survey; Ted Habermann, NOAA; Ralph Haugarud, U. S. Geological Survey; Scott Hills, Chevron Inc; Kerstin Lehnert, Columbia University-Lamont; Kai Lin, San Diego Supercomputer Center; Walter Snyder, Boise State University; David Soller, U. S. Geological Survey; Jon Spencer, Arizona Geological Survey; Evan Thoms, U. S. Geological Survey; Douglas Walker, University of Kansas; Ilya Zaslavsky, San Diego Supercomputer Center.

Graduate Advisors:

Peter Coney, University of Arizona, Tucson (deceased); Richard H. Sibson, University of Otago, New Zealand; John C. Crowell, University of California, Santa Barbara

Graduate Students:

David Mahr, Bronco Creek Associates, Tucson, AZ.

Data Management Plan

Types of Data: Primary data collection through field or laboratory measurements is not anticipated as part of this project. However, the purpose of the Big CZ Software System is to enable web-based integration, visualization, and joint analysis for cross-scale biological and geoscience processes in the critical zone. As such, the Big CZ Software System is necessarily designed to support the full data lifecycle. We anticipate that Big CZ users will be extensively using and sharing existing data from many different sources as research products and as inputs to shared visualization and analyses. These data will include time series of observations from fixed monitoring locations (e.g., streamflow gages, weather stations), data resulting from analyses conducted on samples, time lapse geophysical data, and geospatial datasets (e.g., land use, digital elevation models).

Data Access: To facilitate shared access to the full range of anticipated data types and sources, significant development efforts on this project will be devoted to creating linkages between the BiG CZ Central System and data providers such as the CUAHSI Hydrologic Information System (HIS), the Integrated Earth Data Applications (IEDA) and related systems, the Critical Zone Observatory Integrated Data Management System (CZOData), DOE's System Biology Knowledgebase (KBase), OpenTopography, and federal data providers such as the USGS National Water Information system. Additionally, the BiG CZ Software System will facilitate sharing of individual investigator datasets and derived data products to relevant repositories through development of the BiG CZ Toolbox, as well as the BiG CZ Central System data repository. Development of the required repositories, protocols, workflows, and methods for doing so are part of the intellectual contribution of this project. Development efforts will be devoted to enhancing existing repositories, creating a BiG CZ repository, and creating web services that provide simplified programmatic access to existing datasets and systems by both individual users and the new BiG CZ Portal Web Application. Where possible, we will use existing, standards-based web services (e.g., those available already through the CUAHSI HIS and through the emerging CZOData system) or develop data services and workflows that interface with existing, Internet-based systems for retrieving the primary data that users will require (e.g., the USGS Seamless Data Server for the National Land Cover or National Elevation Datasets), increasing the discoverability and accessibility of these datasets to data consumers working within the BiG CZ Software System.

Other Research Products: The BiG CZ Team will engage the broader community to co-develop and deploy the BiG CZ Software System. Materials will be developed for the co-design workshops and the training and testing workshops to engage community members in the development of the system and to teach them how to use it. Workshop and user documentation aimed at faculty, graduate students, and technicians will be developed to maximize the usability of the BiG CZ Software System. Additionally, we will develop materials for the software bootcamps, science synthesis institute, and citizen science & resource manager workshop that can serve as a model for developing these types of activities for other efforts and projects. All of these materials will be made widely available on a public website affiliated with the project or published in appropriate journals and trade literature.

Data and Metadata Standards: The BiG CZ Software System will rely on existing and

emerging data and metadata standards for sharing environmental datasets. For example, the BiG CZ Portal Web Application will integrate data available through CUAHSI HIS WaterOneFlow web services for time series of point observations and that transmit data in Water Markup Language (WaterML) format (Zaslavsky et al., 2007), and utilise Open Geospatial Consortium (OGC) Sensor Observation Service 2.0 (Bröring, 2012) to distribute WaterML 2.0 format (Taylor et al., 2012, 2013). For geospatial datasets, we will use existing OGC standard interfaces such as Web Map Services (de la Beaujardiere, 2006), Web Feature Services (Vretanos, 2010), and Web Coverage Services (Whiteside and Evans, 2008). Metadata descriptions for all datasets will conform to appropriate ISO and/or OGC metadata specifications, with extensions for specific data types where necessary. A catalog of metadata at the dataseries/timeseries level presents unique issues because of the large number of records aggregated from existing catalogs available through the CUAHSI HIS, CZOData, IEDA, and KBase services. We will explore the use of OGC's Cataloging Services for the Web (CSW) standard (Nebert et al., 2007) for cataloging all granularities of metadata. The BiG CZ Central Software Stack will bridge across the data and metadata from the many critical zone domain data systems.

An important aspect of the proposed work is to operationalize Version 2.0 of the Observations Data Model ("ODM2") and to advance development of information models and metadata standards for the representation of the diverse data types to be integrated within the BiG CZ Portal Web Application. Here we will use the experience of our project team in developing standards – and our unique position at the forefront of developing the data portals we wish to integrate – to create web service interfaces that provide integrated access to data from multiple domain repositories for client applications like the BiG CZ Portal Web Application.

Policies for Data and Research Products: The vision for the BiG CZ Software System is that users could zoom to anywhere in the contiguous USA to search, view, filter, select, and download heterogeneous datasets for the critical zone. Although this project will not generate primary data through field or laboratory data collection, it is envisioned that the BiG CZ Toolbox would provide a set of tools that enable CZ scientists to manage and publish their data using a turnkey, open-source software solution. Given the focus on joint analysis of cross-scale biological and geoscience processes in the critical zone, the BiG CZ software system will focus on and promote freely and publicly available datasets.

As a general policy, all of the proposed software will be developed as a modular suite of fully open source software projects on public repositories and will be built upon widely-used, existing open source software. As such, all of our software components could be adapted or forked to serve other uses or projects. Where possible, all new source code developed by this project will be available under the New BSD (BSD 2) Open Source license. As we are incorporating and modifying some existing applications that already use different open source software licenses (e.g., GeoTrellis uses GPL3), the final BiG CZ Software System may use multiple open source software licenses. We will use publicly available, open source code repositories for our software development. This will enable us to coordinate our development activities across multiple partner institutions and engage developers and contributors from outside of the immediate project team who wish to contribute.

Plans for Archiving Data: As envisioned, the BiG CZ Software System is primarily a system

for data consumers and not a data archival environment. Indeed, a significant focus of the project will be on developing functionality for exposing data from existing data repository partners through a common BiG CZ Portal Web Application. As stated above, however, the BiG CZ Toolbox will provide CZ investigators with a way to manage and publish their data in such a way that it can be cataloged by the BiG CZ Central system and discovered using the BiG CZ Portal Web Application. We anticipate developing arrangements with our data partners (e.g., the CUAHIS HIS and newly funded CUAHSI Water Data Center, the IEDA systems, and others) to facilitate long-term archival of datasets published using the BiG CZ Toolbox.

POSTDOCTORAL MENTORING PLAN

One post-doctoral scientist will be supported by the BiG CZ project. The post-doc on this project will actively collaborate across institutions with the faculty investigators. In addition to technical skills, we focus on: organizational skills, critical thinking; written and oral communication; time management; proposal writing; and public interactions with lay audiences.

The Post-Doc will take on a project coordination role in close collaboration with lead-PI, Aufdenkakmpe and to a lesser extent with Packman and Gill to assist with the broader impacts of the project, specifically with working with resource managers and citizen scientists. The post-doc will thus gain hands on experience managing collaborations on a large multi-disciplinary project. S/he will participate and in many cases coordinate the frequent workshops and WebEx meetings. Additionally,the Post-doc will assist in the coordination data and model testing and integration as the project develops.

The Post-doc will simultaneously be mentored to achieve his or her individual scientific and career goals and will be encouraged and funded to present their research at internal and at one or more national meeting annually. Whenever possible, the mentor will attend the same meeting to introduce the post-doc to scientific colleagues. The project team members will assist the post-doc with publishing post-doctoral research results in peer-reviewed journals. Our mentoring program involves making sure the post-doc has close interactions with students, and is assisted regarding potential job possibilities and guided in job applications. A continual series of discussions are routinely held to talk about career options and their requirements. Assistance is provided to help evaluate job openings and approaches to application and interviews.

The post-doc will take on the role of BiG CZ project coordinator, which is a position that is funded for 3.5 years at the high end of the postdoctoral salary range. S/he will likely have completed a previous Post-doc position and will take on organizational and leadership responsibilities to facilitate communication and collaboration among our research teams. This experience will prepare the Post-Doc for leadership in a similar, large-scale project and will serve as a strong stepping-stone to a faculty position.

The post-doc will be integrated into the research groups of Dr. Anthony Aufdenkampe (SWRC) and by inclusion in weekly group meetings and in weekly BiG CZ team WebEx teleconferences.

azavea

18 March 2013

Dr. Anthony Aufdenkampe
Stroud Water Research Center
970 Spencer Road
Avondale, PA 19311

RE: Collaborative Research: SI2-SSI: The community-driven BiG CZ software system for integration and analysis of bio- and geoscience data in the critical zone

Dear Dr. Aufdenkampe:

On behalf of Azavea, I am pleased that we are collaborating with the Stroud Water Research Center and other project partners on *Collaborative Research: SI2-SSI: The community-driven BiG CZ software system for integration and analysis of bio- and geoscience data in the critical zone*. Should this project be funded, Azavea agrees to provide the application development services outlined in the proposal and described in more detail attached to this letter. Azavea will collaborate with the Stroud Water Research Center (SWRC) and Dr. Emilio Mayorga's team at the Applied Physics Laboratory of the University of Washington (APL-UW) to develop the web portal component of the BiG CZ software. Azavea's specific focus will be on 1) co-design a software architecture design; 2) develop user interface and user experience designs; and 3) implement the web client software. The activities will leverage Azavea's previous work with distributed processing of geospatial data and UI/UX design.

Azavea specializes in creating web-based geographic analysis and modeling applications for government (local, state and federal), non-profit and academic clients. We focus on developing products and services that use both open source and commercial GIS technologies and geographic information science to solve complex, interesting and novel problems. The proposed project is particularly exciting for Azavea as we have a substantial background with land conservation and water management. The proposed spatial data integration project will leverage open source software tools and high performance computing frameworks on an innovative effort to integrate theory, models and data from a variety of Critical Zone disciplines to collectively study processes on the Earth's surface as well as human impact on those processes. As a certified B Corporation, the project also has clear social value and fulfills Azavea's social mission to apply geospatial data and technology to improving the communities we live in.

Please do not hesitate to call me at 215-701-7713, or e-mail cheetham@azavea.com if you have any questions about this partnership.

Sincerely,

Robert M. Cheetham
President and CEO

340 North 12th Street
Suite 402
Philadelphia, PA 19107

T (215) 925 – 2600
F (215) 925 – 2663
www.azavea.com

ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

Dr. Susan Hubbard
Deputy Division Director
Earth Science Division
Lawrence Berkeley National Laboratory
1 Cyclotron Rd MS 90-1116
Berkeley, CA 94720
March12, 2013

Anthony Aufdenkampe
Stroud Water Research Center
Re: Letter of commitment for NSF "Big CZ" proposal

Dear Dr. Aufdenkampe,
With this letter Lawrence Berkeley National Laboratory (LBNL) commits to the collaboration with your collaborative proposal to NSF titled "Collaborative Research: SI2-SSI: The community-driven BiG CZ software system for integration and analysis of Bio- and Geoscience data in the Critical Zone".
Specifically, we anticipate a collaboration between your project and the Department of Energy (DOE) funded LBNL Science Focus Area (SFA) project. The SFA project focuses on the development of a predictive understanding of multi-scale terrestrial ecosystems, with particular consideration of the role of coupled hydrological-biogeochemical and subsurface processes on system functioning. Specifically, it seeks to develop the understanding (which is critical to the DOE mission) of how climate induced changes in hydrology and vegetation affect subsurface carbon inputs, flow paths, biogeochemical cycling and metabolic potential, how these processes evolve over time, and what the effect of these processes is on watershed biogeochemical functioning.

As part of the SFA project, a significant volume of disparate microbiological, geochemical, geophysical and hydrological laboratory and field data are being collected by DOE funded scientists, and models are being developed which allow for understanding of coupled processes that play out from single cell to the watershed scales. While the SFA effort includes some modest support for cyberinfrastructure, the primary focus are paired process investigations and simulation capability developments. Given the volume and diversity of our datasets, we recognize the critical need for community tools for integration of biological and geoscience data, and feel that your proposed effort provides for an excellent opportunity for synergistic collaboration. Thus, if your proposal is successful, we will share data and models developed under the SFA with your project with the anticipation of using your developed cyberinfrastucture . The initial point of contact and collaboration would be Dr. Roelof Versteeg (a lead of the LBNL SFA data management and assimilation component), but I anticipate that other SFA scientists would also become involved through this collaboration.

Sincerely,
Susan S. Hubbard
Susan Hubbard
Lead, LBNL Sustainable Systems SFA

Lawrence Berkeley National Laboratory, One Cyclotron Road, MS 90-1116, Berkeley, CA 94720
Tel: (510) 486-5266 Fax: (510) 486-5686 E-mail: sshubbard@lbl.gov
http://esd.lbl.gov/about/staff/susanhubbard/

Ruth Duerr
National Snow and Ice Data Center
Cooperative Institute for Research in Environmental Science
University of Colorado at Boulder
Boulder, CO 80303

March 19, 2013

Anthony K. Aufdenkampe, Ph.D.
Associate Research Scientist - Isotope & Organic Geochemistry
Stroud Water Research Center
970 Spencer Road, Avondale, PA 19311

Dear Anthony,

I am very interested in the goals of the Big CZ SSI proposal and am willing to participate on the BiG CZ Community Advisory Board and to attend one or two of your co-design workshops as funding permits. In particular, I am interested in extending the systems you are proposing to enable discovery and access to remote sensing datasets and hope eventually to explore with your team methods to adapt high performance raster visualizations to NASA and NOAA's vast data holdings.

Sincerely yours,

Ruth Duerr

164 Ashdale Avenue
Toronto, Ontario
Canada M4L 2Y9
+1 (416) 435 9779
gvwilson@software-carpentry.org
http://software-carpentry.org

March 19, 2013

To Whom It May Concern:

Re: The community-driven BiG CZ software system for integration and analysis of bio- and geoscience data in the critical zone

I am writing to express my strong support for the above-mentioned Scientific Software Integration program proposal by Aufdenkampe et al. Software is as important to 21st Century science as telescopes were in the 19th Century and particle accelerators in the 20th. However, most scientists are still never taught how to design, build, maintain, validate, and share software productively. In our experience, the mixed delivery model included in this proposal, which combines an introduction to common-core skills with more targeted training in domain-specific material, is the most effective way for researchers to learn what they need to know as efficiently as possible.

Software Carpentry will therefore work with the proposers to deliver a series of intensive workshops on software skills for research scientists throughout the lifetime of the project. More specifically, we commit to teaching the “common core” of task automation, version control, structured program development, and software testing over two days, four times or more during two years, and to accommodate 20 or more learners per workshop. We will arrange the instructors and supply all the open source software and open license learning materials required; the only cost to the project will be the instructors' travel and accommodation.

I believe that projects like this one will not only advance our understanding of the Earth's Critical Zone, but also serve as an exemplar for scientists in other domains. I look forward to helping them accomplish their goals.

Sincerely,

Dr. Gregory V. Wilson
Project Lead, Software Carpentry

Texas Water Development Board

P.O. Box 13231, 1700 N. Congress Ave.
Austin, TX 78711-3231, www.twdb.texas.gov
Phone (512) 463-7847, Fax (512) 475-2053

March 19, 2013

Grant Manager
Software Infrastructure for Sustained Innovation - SSE & SSI (SI2-SSE&SSI)
National Science Foundation
4201 Wilson Blvd
Arlington, VA 22230

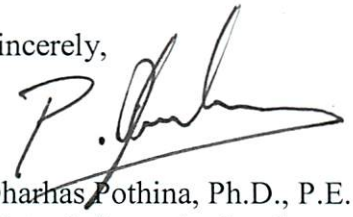
Dear Sir:

Re: TWDB support for BiG CZ SI2-SSI proposal

For the past several years, the Texas Water Development Board (TWDB) has led the development of the open source ulmo and wofpy python web service packages. TWDB is committed to further development of these packages and welcome contributions and collaborations from the BiG CZ proposal team to further extend their capabilities.

By this letter, we commit staff time to working with the BiG CZ team collaboratively to design, implement, and test extensions of the capabilities of these packages. The code is and will be licensed as an open source product.

Sincerely,



Dharhas Pothina, Ph.D., P.E.
Water Informatics Lead
Water Science and Conservation

512-936-0818 | Phone
512-936-0816 | Fax
dharhas.pothina@twdb.texas.gov

Supplement 3. Management and Coordination Plan -- BiG CZ SSI

S3.1. Project Team, Roles and Responsibilities

The core of the BiG CZ SSI team includes the same investigators driving critical zone data integration efforts through the CZOData project (Award #1332257), the ODM2 project (Award #1224638), and the CZ-EarthCube Domain Workshop (Award #1252238). This core team has been joined by additional bioscience and geoscience investigators and software developers to expand on the team's extensive experience in community building, software development, and distributed system architecture.

Aufdenkampe (SWRC, lead-PI) is an aquatic geochemist who is co-leading the Christina River Basin CZO and leading the CZOData project. He will oversee overall project coordination, including mentorship of the **post-doctoral project coordinator**, and will be responsible for Objective 1 community engagement and broader impacts. **Mayorga** (UW-APL, co-PI) is a biogeochemical modeler, cyber-scientist and co-developer for the NANOOS Visualization System. He will oversee Objective 2 BiG CZ Portal and Objective 3 BiG CZ Toolbox software development. **Zaslavsky** (SDSC, PI) is a cyber-scientist and co-developer of the CUAHSI Hydrological Information System (HIS) cyber-infrastructure. He will oversee Objective 4 BiG CZ Central software development. These three together will form the project's executive team.

Objective 1 team. Aufdenkampe will lead, including a small team of technical and data management staff for SWRC and the CRB-CZO who will test each software release. **Packman** (NU), a hydrologist, and **Aronson** (UCR), a microbiologist, will co-chair the CZ Community Advisory Board. Packman will lead the CZ Science Synthesis Institute. **Gill** (SWRC Director of Education), a soil scientist, will engage resource managers, citizen science NGOs and secondary school educators throughout the project by inviting key representatives to co-design workshops and by organizing a workshop at the end of year 3.

Objective 2 Team. Mayorga will lead, including a team of developers at UW-APL. **Cheetham** is founder and CEO of Azavea, Inc., a systems design and software development firm specializing in web and mobile software that combine geospatial technology, engaging user interfaces and high performance computing. As lead developer for the GeoTrellis geoprocessing engine, he and the Azavea team will work closely with the UW-APL team to develop the core high-performance web visualization framework. **Versteeg** (Subsurface Insights) is a geophysicist and cyber-developer who will integrate the VisIt 3D and 4D visualization package into the BiG CZ Portal web app. **Lehnert** (Columbia) is a marine geologist and lead for the Integrated Earth Data Applications (IEDA) facility. She and her team will integrate a suite of data publication/registration forms into the BiG CZ Portal web app.

Objective 3 Team. Mayorga will lead, including a team of developers at UW-APL. **Horsburgh** (USU) is a hydrologist, the developer of ODM 1.1 database and tools, and the PI of the ODM2 project. He and his team will work closely with the UW-APL team to develop the BiG CZ Toolbox. Versteeg will provide guidance with the integration of the VisIt Python API into the BiG CZ Toolbox.

Objective 4 Team. Zaslavsky will lead. **Valentine** (SDSC) is a developer for CUAHSI HIS and many other projects, using the Microsoft DotNet platform, Python, OData, and various web services, and experience with tuning geospatial databases and digital library systems and metadata. **Richard** (AZGS) and his team (R. Clark, G. Chen) developed the Geoscience

Information Network (USGIN) (Award #0753154) and the National Geothermal Data System 'node-in-a-box' stack (Clark et al., 2013). They have experience with Django, Python, Node.js, Javascript, and a variety of other FOSS components. **Berg-Cross** (SOCoP) brings expertise in metadata integrating and vocabulary mapping. Together, Zaslavsky, Valentine, Richard and Berg-Cross will develop the BiG CZ Central software stack. **Meyer** (ANL) and **Henry** (UC) serve as the Microbial Communities and the Microbial Science team leads for the US DOE's Systems Biology Knowledgebase (KBase). They will work with the Obj. 4 team to implement or expose features within KBase to enable cross-domain interoperability with BiG CZ Central.

In addition, we have four collaborative project contributors, who are not directly funded under this proposal but have all committed time, effort or resources. See supplementary Letters of Commitment. These include: **Greg Wilson** (Software Carpentry), who will assist with our four Training and Testing "BootCamp" workshops by following Software Carpentry's well-established training approach (http://software-carpentry.org/bootcamps/); **Ruth Duerr** (National Snow and Ice Data Center, NSIDC), who will participate inthe BiG CZ Advisory Board and attend co-design workshops to add in our integration of NASA remote sensing data; **Dharhas Pothina** (Texas Water Development Board), who is the leaf for the "Water One Flow in Python" (WOFpy) project and will assist in co-developing a web service package for our BiG CZ Toolbox; and **Susan Hubbard** (Lawrence Berkeley National Lab), who will collaborate with us on integrating the VisIt 3D visualization software into BiG CZ Portal.

S3.2. Cross-disciplinary and cross-institutional project management

The project's Executive Team (Aufdenkampe, Mayorga and Zaslavsky) will jointly oversee the coordination of tasks across partner institutions, within each of the four Objectives. The Executive Team will meet via telephone or webex at least once a month, and will be joined by at these meetings by most other project members. Overall governance and guidance for the project will be provided by the CZ Community Advisory Board (CZCAB), which will advise on feature development priorities, provide feedback on releases, recruit workshop participants and promote the BiG CZ software system. CZCAB membership will be composed of project team members Aufdenkampe, Packman and Aaronson, and additional CZ scientists to be drawn from the larger CZ community, the CZO Network community, and EarthCube CZO workshop (Jan. 2013) attendees.

The CZCAB will provide formal project guidance, review and recommendations once a year. Project management and coordination mechanisms will be established at an initial co-design kickoff workshop within the first six months, and in preceding regular communication among all project members, led by the executive team.

S3.3. Communication and coordination mechanisms

Excellent communication is a critical feature of any large collaborative effort, and our project leadership team has extensive experience with collaborative, multidisciplinary research involving large groups spread out over several campuses and in particular working together on the currently NSF-funded CZOData, ODM2, and CZ-EarthCube projects. Presently this core team has weekly WebEx video conferencing meetings, developing agendas, minutes and tasks lists collaboratively in Google Docs. In addition, we have recently started using Asana collaborative task management software (http://asana.com/) to further enhance these

commmunications. We will continue such practices as our group expands for the BiG CZ SSI project. However, because our group will more than double in size, we have budgeted for a post-doctoral level project coordinator, to assist in project management.

In addition to weekly web conferences, we will take advantage of the seven planned workshops for face to face interactions, with the first co-design workshop specifically designated for project team idea-sharing and communications.

Prioritization of functionality and system architecture choices resulting from the initial co-design workshops will be stored and maintained in a project backlog on Asana, to identify and assign tasks and status. Software development will be based on the use cases generated at these workshops.

The project team will use an 'Agile' project-management and software-development methodologies to accelerate the design and development process, effectively share knowledge among team members, discover impediments early, and result in a well-tested and high-quality software system. This includes selective pair programming, test-driven development and frequent iterations. We will use the GitHub platform's project communication tools such as project wiki, issue tracking system and version/source control. We will also utilize the Scrum project management process. Under Scrum, the project team organizes the work into 30-day 'sprints'. At the beginning of each sprint, the entire team reviews completed features and tasks are prioritized and assigned to team members. Our experience with the Scrum process has demonstrated that it improves alignment of project goals with development work, identifies blind alleys earlier, enables work to be quickly adapted to changing conditions and results in a cohesive team that feels responsible for delivering a high-quality product. All 30-day sprints will be organized within three larger, 3-month long major development sprints organized throughout the life of the project.

In addition to unit tests of the code we will utilize approval test strategies to allow for domain scientists to approve information transformations. BiG CZ software system users will also have ongoing opportunities to provide input via the user review and training workshops, the CZ Science Synthesis Institute, the Citizen-science & Resource Manager workshop, and demonstrations and workshops on the BiG CZ system at professional meetings and other targets of opportunity to engage the community and solicit feedback.

S3.4. Budget Support for Project Management and Coordination

The post-doctoral project coordinator is budgeted by SWRC, as are all participant support costs for invitees. Travel to annual project meetings and workshops by team members is budgeted under their respective travel budgets, along with travel for relevant scientific and cyberinfrastructure conferences and workshop in support of the project's training needs and outreach goals. The costs of hardware acquisitions and cloud deployments for production and development environments are listed in the budget justification documents for SDSC, UW-APL and Azavea.