Rediscovering EarthCube: Collaborate. Or collaborate not. There is no I

Plato Smith II
Department of Research Data Services, University of New Mexico, Albuquerque, New Mexico, USA
Tanu Malik
University of Chicago, Chicago, Illinois, USA, and
Gary Berg-Cross
RDA US Advisory Committee, Washington, DC, USA

Received 7 September 2015
Revised 11 November 2015
Accepted 12 November 2015

Abstract

Purpose – The EarthCube Technology and Architecture Committee working groups needed current information on the development of existing EarthCube-funded projects (e.g. building blocks, conceptual designs, and research coordination networks) to fulfill the goals of the working groups (e.g. gap analysis, use cases, standards bodies and testbed). The aims of this study include a compilation of planned outcomes, an assessment of current work and an investigation of interests in research collaboration among select EarthCube-funded projects.

Design/methodology/approach – Twenty-four principal investigators of 24 different EarthCube projects completed the Funded Projects Questionnaire composed of 35 questions in March and April 2015.

Findings – The survey response rate was 100 per cent and included a diversity of results ranging from planning stages to early development to final development. The funded projects in this study received awards in 2013 and 2014.

Research limitations/implications – The results are EarthCube-specific and are not generalizable. Suggestions for future research include integration of crosscutting disciplines and perspectives, best practices, guidelines and standards for broader impact.

Practical implications – This study identified potential collaboration opportunities, use cases and gaps (e.g. unmet architectural, functional, operational, organizational and/or technical needs).

Social implications – The impact on society include an improved understanding of the various EarthCube-funded projects and potential for collaboration within and across multiple disciplines.

Originality/value – This study contributed to the development of select outputs for EarthCube-funded projects' presentations, Tech Hands Meeting, 2015 All Hands Meeting, select

This study was made possible by contributions of the PIs, co-PIs and senior personnel of all of the EarthCube (EC)-funded projects who participated in this study. Thanks to the EarthCube Technology and Architecture Committee and Engagement Team, Jay Pearlman, Yolanda Gil, Marjorie Chan, Basil Gomez, Ilya Zaslavsky, Phil Yang, Matty Mookerjee, the Gap Analysis Working Group, Anna Kelbert, Emily Law, Mike Daniels, Janet Fredericks, Rachael Black and Steve Diggs. Special thanks to Karl Benedict, Director of Research Data Services, and Laura Soito, Physical Sciences Librarian, at the University of New Mexico.

2059-5816
DOI 10.1108/DLP-09-2015-0017

working groups' outcomes and EarthCube Strategic Technology Plan and is of value to stakeholders, scientists and users.

Keywords Research, Collaboration, Assessment, EarthCube, Funded projects, Geosciences

Paper type Research paper

Introduction

There is growing demand for scientists and researchers to develop data management plans (DMPs) and research data management (RDM) workflows that manage data over the data's life cycle (NSF, 2010; Tenopir *et al.*, 2011). Academic institutions, research organizations and stakeholders must develop campus-wide research data services resources that assist scientists and researchers in meeting funding agencies' DMP and data sharing requirements. The National Science Foundation (NSF) DMP Requirements[1] and the NSF Public Access Plan[2] supported by the Fair Access to Science and Technology Act of 2015[3] influenced by the 2013 White House Office of Science and Technology Policy (OSTP) Memorandum[4] on "Increasing Access to the Results of Federally Funded Scientific Research" are major examples of such requirements. These evolving federal data management and sharing policies apply to federally funded projects, including the NSF-funded EarthCube[5] project. The current and planned software infrastructure development of some EarthCube building block-funded projects seek to produce products that aid geoscientists in DMP planning, data reproducibility and sharing within the geosciences and across disciplines such as GeoSoft: Collaborative Open Source Software Sharing for Geosciences project and its Ontosoft[6] software application; CINERGI: Community Inventory of EarthCube Resources for Geoscience Interoperability; Enterprise Architecture for Transformative Research and Collaboration; EC3: Challenges of Field Data Collection, Management and Integration; and iSamples: The Internet Samples in the Earth Sciences.

EarthCube[7], Developing a Community-Driven Data Knowledge Environment for the Geosciences, is an NSF-sponsored partnership between the Directorate of Geosciences (GEO) and Directorate of Computer & Information Science & Engineering (CISE). GEO includes the Divisions of Atmospheric and Geospace Sciences, Earth Sciences, Ocean Sciences and Polar Programs. CISE includes the Division of Advanced Cyberinfrastructure. "EarthCube will enable geoscientists to address the challenges of understanding and predicting a complex and evolving solid Earth, hydrosphere, atmosphere, and space environment systems"[8]. As of May 2015, EarthCube includes four main types of funded projects for developing cyberinfrastructure and research practices for the geosciences: building blocks (BBs), conceptual designs (CDs), research coordination networks (RCNs) and integrative activities (IA) (Gil *et al.*, 2015). This study explored BB-, CD- and RCN-funded projects from March to April 2015.

The EarthCube Technology Architecture Committee (TAC) developed working groups[9] to gather current information on the developments and planned outcomes of existing EarthCube-funded projects. The TAC Gap Analysis Working Group (TGA WG) developed a preliminary framework and gap analysis model to contribute to such efforts.

The primary goal of the TGA WG is to consolidate understanding about these projects, especially the available interfaces, and document outputs and interactions that projects have with each other as they progress. The working group, in its preliminary discussions, has realized that principal investigators (PIs)/co-PIs and stakeholders for funded projects often do not have the resources to make inter-project connections, and so, the working group will serve as a coordinator for the funded projects, so that transparent, accurate, reliable and timely information on the project deliverables, outcomes, prototypes and websites is shared in a consistent manner. The focus of the TGA WG is to improve the flow of information about funded projects (e.g. the way information is presented, represented and shared across current and future funded projects), understand how funded projects interact and identify gaps within, between and across funded projects to promote an integrated implementation of the EarthCube CD(s). This project contributed to information gathering and sharing:

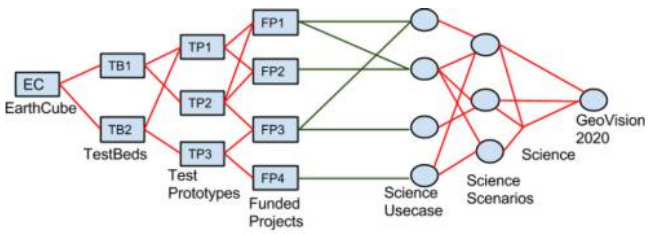
RQ1. How can the assessment of funded projects contribute to increased research collaborations, data management support and software infrastructure development that aids geoscientists in effective and efficient data reproducibility and sharing?

Definitions and key concepts relevant to the study

The TGA WG defined key concepts and created a model in which to conduct preliminary work and prepare for future gap analysis. During the course of preliminary work through working group meetings, observations of funded projects' conference presentations and review of outputs, the gap analysis group agreed on three broad, high-level dimensions for the gap analysis. The gap analysis model (Figure 1) conceptualizes an example of the three dimensions of gaps (features and functionality, integration requirements and operational requirements) in practice within funded projects and in theory within testbeds environments.

As the survey was aimed at identifying gaps, survey questions were organized along areas of anticipated gaps. The gaps impede data integration, sharing and research collaboration as represented in the absence of use cases, prototypes and testbeds in which to develop capacities and infrastructure:

- *EarthCube (EC)*: EarthCube[10] is a community-driven, 10-year initiative aimed at transforming the conduct of geosciences research and education to develop the cyber-infrastructure for the geosciences to better enable transformational science within and across disciplines. EarthCube enables "fostering community-governed efforts that develop a common cyberinfrastructure for the purpose of collecting, accessing, analyzing, sharing, and visualizing all forms of data and related resources, through the use of advanced technological and computational capabilities"[11]:



Source: Developed by Tanu Malik

Figure 1.
A model for gap analysis

- *Building blocks (BBs)*: BB represented initial software component development for EarthCube.
- *Conceptual designs (CDs)*: CD represented broad architectural design for EarthCube. The idea is that understanding the CDs will facilitate the scoping, alignment and coordination of EC BBs and to help manage architectural complexity by describing component interdependencies in a usable, understandable fashion.
- *Research coordination networks (RCNs)*: "EarthCube RCNs are intended to advance geosciences cyberinfrastructure through interaction, discussion and planning between geoscientists and cyberinfrastructure experts. RCNs provide opportunities for academic geosciences communities to organize, seek input, come to consensus and prioritize data, modeling, and technology needs, as well as standards and interoperability within and across domains"[12].
- *Gap*: Identification of architectural, functional and technical gaps is a key aspect of enterprise architecture efforts (Saha, 2004) and identifiable with CDs. Within the scope of this paper, a gap is defined as an unmet need (e.g. architectural, functional and technical). An example is a capability gap which is the inability to achieve a desired effect, analysis or data process under specified standards and conditions through combinations of means and ways to perform a set of tasks.
- *Gap analysis*: Within the scope of this paper, gap analysis is the comparison of existing capabilities to a set of required capabilities (e.g. "What we have against what we need") as defined by EarthCube Science Domains and Technical Architecture Strategy relevant to appropriate working groups. For example, as part of EarthCube software and service development, analyses were used to document which services and/or functions have been accidentally omitted, have been deliberately eliminated and/or need to be developed.
- *Gap analysis model*[13]: To understand potential gaps, the TGA WG model was used to capture the current state of EarthCube, where progress is being made to fulfill gaps, gaps that are being identified and gaps that still need to be filled. Figure 1 shows one such model in which the squares represent funded projects and the circles represent the class of science use cases that they are currently developing the technology for. The green lines show that the gap is being fulfilled, in that some technology is being developed to address the chosen science use case. Outstanding gaps include the ability of funded projects to apply themselves to more realistic science and encompassing use cases to accomplish the EarthCube GeoVision 2020[14] and for funded projects themselves to expand their technical features (test prototypes, testbeds, etc.), so that they can address more realistic science use cases.
- *Science scenarios*: "Science scenarios and use cases capture specific contexts where technology can be inserted in science practice, and can therefore play a crucial role in facilitating communication and fruitful interactions between computer scientists and geoscientists"[15] (Gil *et al.*, 2015, p. 3).
- *Testbeds*: According to National Oceanic & Atmospheric Administration (NOAA) Testbed & Proving Grounds[16], testbeds "facilitate the orderly transition of research capabilities to operational implementation through development testing

in testbeds, and pre-deployment testing and operational readiness/suitability evaluation in operational proving grounds, as approved in the Guidelines and Performance Measure".

- *Use case*: According to the Unified Modeling Language, "a use case shows the interaction between the system and 'actors', which may be human or other systems" (Kelbert, 2014, p. 5). To understand researchers' needs, the science case, one may adapt the usual software-oriented use case model to describe more abstract things. These could be system/capability boundaries, how they interact, scientific workflows involved and some idea of services offered by the system as well as the data used.

Gap Analysis Working Group preliminary framework and model section

The EarthCube-funded projects provide a vital insight into the technologies that are important for EarthCube and a vision of the architecture of a future EarthCube. Thus, there is a critical need to aggregate, coordinate and articulate accurate and comprehensive information about the funded projects across the EarthCube community, and identify gaps in capabilities.

The TGA WG identified three dimensions of performing gap analysis, namely:

(1) features and functionality (e.g. Demo Science Use Case 1, etc.);

(2) integration requirements (e.g. Testbed 1, Test Prototype 1); and

(3) operational requirements (e.g. applying test prototypes and testbeds, etc.).

The three dimensions of gaps represented common challenges among the funded projects. The desired goals for some funded projects is to increase the features and functionality, integration and operational capacities of software APIs and tools to facilitate data access, discovery and integration of research data within and across disciplinary domains. An example is the effective execution and performance in the use of Network Common Data Form, Sensor Model Language and WaterML data integration crosswalks within and across domains.

The technical (non-functional) and architectural requirements gaps will include operations such as loose coupling, federation, standards (compliance) and performance requirements (response time, resource utilization and availability).

Additionally, to make the task of gap analysis easier, the TGA WG has made progress on two efforts:

(1) the TGA WG surveyed funded projects for gaps and opportunities for collaboration with other funded projects; and

(2) coordinated with other working groups (such as the science use case and testbed working groups), to identify existing gaps from other working groups' perspectives.

The TGA WG objective is to perform gap analysis that will allow the comparison of current capacities in existing applications, services, data, system and technical architecture against a set of targeted architectural needs for those applications, services, data and technical capabilities.

The gap analysis model (Figure 1) conceptualizes key components in the exploration of gaps stemming from the GeoVision 2020 plan to science drivers to science scenarios to developing use cases. One of the goals of BBs-funded projects is to create use cases, while one of the goals of the funded RCNs is to connect multiple funded projects through test prototypes as an outcome of testbeds. Through the processes and workflows from the GeoVision 2020 plan to testbeds supported through the collaborations of funded projects, CDs and RCNs, functional components of EarthCube are supported to advance the access, integration and management of data across the geosciences.

Methodology

The TAC developed (with contributions from select working groups) the survey to gather and study information on current efforts. Members of TAC and PIs of two EarthCube-funded projects tested the survey before distribution to the PIs. The online questionnaire survey was edited and updated based on feedback during testing and then disseminated to the PIs of EarthCube-funded projects in March 2015. The funded projects that participated in this study at the time this survey was disseminated included 15 BBs, 3 CDs and 6 RCNs. The survey was completed by all EarthCube-funded projects' PIs in April 2015 before the April 8-10, 2015, Tech Hands Meeting[17] and the May 27-29, 2015, All Hands Meeting (AHM)[18]. Information gathered from the survey was used in the development of material for both the Tech Hands and AHM presentations. The TGA WG-funded projects questionnaire and the 2015 EC3 Field Trip to Yosemite and Owens Valley are key components in this study used to gather information on some of the challenges and gaps encountered by EarthCube-funded projects.

The issues surrounding the development of the survey instrument included integration of working group-specific questions while lowering the overhead for PIs to complete the survey (e.g. creating one survey vs multiple surveys from all working groups). The survey comprised questions developed from multiple working groups and perspectives. The questions were aggregated, compiled, organized and streamlined into one survey through a series of reviews, feedback and edits. During the survey pretest, researchers and PIs provided constructive feedback on clarifying unclear and/or redundant questions. The final developed survey gathered useful information on the current and planned outcomes of EarthCube-funded projects in support of the geosciences.

To improve the overall understanding of research data services in relation to the required functional areas of EarthCube (e.g. data access, data discovery, data integration, data management, modeling, ontologies and other things supported by the components of the BBs, CDs and RCNs projects), a collaborative assessment of EarthCube-funded projects' components, infrastructure and outcomes was necessary for baseline information and knowledge needed to facilitate future research collaborations. The study of this baseline was a collaboration between the EarthCube Technology and Architecture Committee[19], working groups, a research data services unit within an academic research library and the Gap Analysis Working Group.

Funded projects questionnaire

This study used an electronic, Qualtrics questionnaire survey composed of a total of 35 open-ended, closed-ended, single-choice and multiple-choice questions. The PIs were given the options of (A) completing the survey online or (B) completing the survey off-line in a Word document (off-line Word document surveys were later entered into the online survey database by a member of a research data services team). The survey covered several categories. Questions included identification of scientific EarthCube targeted goals and objectives. The goals were broken down into component parts of the projects needed to achieve objectives. For example, a cloud-based brokering framework was identified as needed to mediate "query and access across diverse data sources". This in turn was described as, "serving science use cases in hydrology, oceanography meteorology and cryospheric research". The survey categories included the following topics:

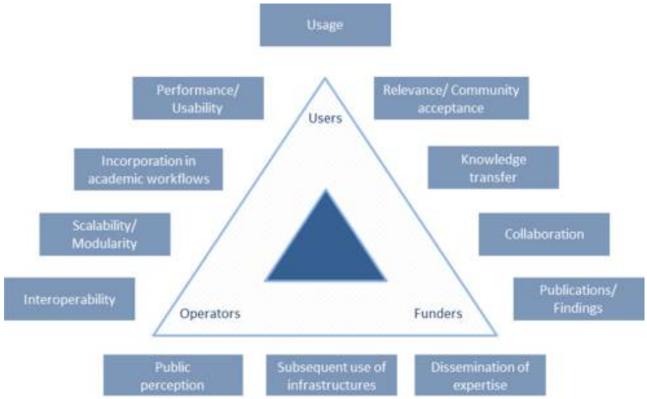
- Project Description including Planned Outcomes (Questions 1-8).
- System and Technology Building Blocks and Concept Architecture Projects (Questions 9-18).
- Standards and Interoperability (Questions 19-23).
- User Interface and Science Testing (Questions 24-28).
- Testbed – Environment and (Systems) Interface (Questions 29-35).

Please note that all the questions were not relevant to all the funded projects (e.g. BB, CD or RCN) and a "no response" does not necessarily represent missing and/or incomplete information. Some PIs completed the questionnaire with the understanding that response to the questionnaire was preliminary and that the answers were subject to change. All of the PIs completed the questionnaire within a short, March and April 2015, time frame.

This study is intended as an introduction to some of the results from selected questions and is not a comprehensive representation and/or analysis of all of the responses from the funded projects. The Gap Analysis Working Group determined that science drivers, use cases and testbeds of funded projects in development require observations, identification and analysis of gaps in testbeds and science use cases for effective, meaningful and useful gap analysis. This study is a preliminary exploration of funded projects progressing toward the development of use cases and testbeds. During literature review, Figure 2 provided an example of a good reference model to use in developing future success factors and criteria metrics for funded projects and virtual research environments (VREs)[20]. The PIs, co-PIs, senior personnel and project team members contributed to the information captured in the projects' responses and the completion of the surveys:

> The criteria for the success of a virtual research environment involves starts with defining success of a virtual research environment in general and in relation to stakeholders as illustrated by success factors and measurable success criteria (Buddenbohm *et al.*, 2015, sec. 4.1, 4.2).

Figure 2 is an example of conceptualizing many of the complex components that researchers and scientists must plan, develop and prototype with limited resources and in a highly competitive marketplace while meeting the funders', operators' and users' expectations. The success factors and measurable success criteria for VREs (Buddenbohm *et al.*, 2015) articulate some of the challenges, gaps and desired goals of some of the funded projects based on the survey responses, direct interactions and conference meetings and can be used as a reference model when developing success factors and measurable criteria for funded projects including the VRE process of the



Source: Buddenbohm *et al*. (2015) (used with permission)

Figure 2. Success factors and measurable success criteria, weighted by relevance for the different stakeholder categories

research life cycle. "Research communities typically develop their own standard methods and approaches for managing and disseminating data" (Michener, 2015). Identifying and assessing the varied data management practices and workflows within research communities while integrating evolving data management requirements present constant challenges requiring interdisciplinary collaborations in which academic research libraries can serve as a partner in developing capacity for campus-wide data management.

Despite the perceived fragmentation perspective (Martin, 1992, p. 130) and varied public perceptions of some colleagues, researchers and scientists based on direct conversations, EarthCube-funded projects continue with transformative research efforts to "enhance the relationships between academic social researchers, [scientists] and their broader constituencies beyond the university" (Greenwood and Levin, 2000, p. 85) including but not limited to the geosciences.

Results and discussions

Appendix 1 – building blocks (BBs), conceptual designs (CDs) and research coordination networks (RCNs – includes the types of funded projects, name of the funded projects with hyperlink to the NSF Award Abstracts (e.g. proposed project summaries), numbers of PI/co-PI/senior personnel, project start date and project end date. The projects with expired end dates are highlighted in gray. Some of the project end dates entered by PIs did not match the end date from the NSF Award Abstract information.

Building blocks projects

With suggested input from a co-PI of one of the funded projects and based on understanding of the planned outcomes, the BBs have been assigned into one or more of five key categories based on their projects' planned outcomes, including technical, science and/or community outcomes. The BBs' planned outcomes are distinguished by the following categories.

(1) framework or testbed;

(2) standard;

(3) tool;
(4) report; and
(5) community engagement.

The BBs' planned outcomes were distinguished into one or more of the key categories based on understanding of the planned outcomes such as BB1-BCube "a cloud-based brokering *framework* […]" (Tables I and II). This process was applied to the remaining BBs, resulting in categorization of BBs to compare and contrast the similarities, differences and interrelationships of planned outcomes for future gap analysis of actual outcomes. Table I provides the first four of the BB planned outcomes from Appendix 2 – building blocks (BB) categories, planned outcomes, use and availability – that includes responses to survey Questions 4, 5 and 7:

- Question 4 – Please list up to five planned outcomes of your project including technical, science and/or community outcomes.
- Question 5 – Which outcomes can be used directly by the scientific community, for example specific tools?
- Question 7 – What outcomes will be available for use by the AHM in May from the list in Question 4?

Planned outcomes use by the scientific community

Below are the results of all BBs' planned outcomes for use by the scientific community and available for use by the 2015 AHM.

Outcomes available for use by the All Hands Meeting in May 2015

In all, 73 per cent of the BBs reported at least one planned outcome available for use by the 2015 AHM. However, the planned outcomes available varied greatly from 1 to 5 or test.

The full list of the BB planned outcomes, use by scientific community and available for use by the 2015 AHM responses to Questions 4, 5 and 7 are available in Appendix 2.

Question 8 – Please give specific examples of benefit to the scientific community.

Rather than list all of the qualitative responses to Question 8, the top five most selected BBs in which PIs from other funded projects are interested in interfacing were selected (Table VI):

(1) *BB1 – BCube*: "The broker is now capable of accessing data sets from a USGS OPeNADP server and making them directly accessible from within desktop GIS applications. The functionality is being incorporated into a widely-used hydrological modeling tool with thousands of users around the world. The broker is also accessing a variety of oceanographic datasets from different sources and making them available via an established community portal (the MARACOOS Asset Map)".

(2) *BB3 – CINERGI*: "Ability to organize resources within their domains (RCNs using community resource viewers for this: C4P, SEN, also EcoGEO and possibly InlandWaters); Ability to find information resources across domains and catalogs; Ability to add resources to the catalog and make them discoverable

Table I.
Sample of BBs' planned outcomes distinguished by framework or testbed, standard, tool, report or community engagement

Project	Category	Question 4 – Please list up to five planned outcomes of your project including technical, science and/or community outcomes.
BB1-BCube	Framework or testbed	A cloud-based brokering *framework* that is mediating query and access across diverse data sources, serving science use cases in hydrology, oceanography meteorology and cryospheric research. A queryable catalog of Earth Science resources (data sets, and services by which data are made accessible), and the cloud-based Web crawler/reasoner that is used to populate and update this catalog. *A report on the current state of metadata and service description standards* and the impacts these have on interoperability, including recommendations for improvements. A prototype tool for scientists who seek an alternative to depositing their data in a repository and wish to advertise their data in a manner that is discoverable on the Web. A cloud-based interoperability testbed
BB2-CHORDS	Community Engagement	*Engagement* with real-time geosciences community, including real-time data providers and consumers. Demonstration of a prototype system that enables cloud-sourced, standards-based, easy access to real-time sensor data and otherwise proprietary or unavailable real-time sensor data streams. Specifications and documentation for scientists to use the prototype system
BB3-CINERGI	Framework or testbed, Standard, Tool, Community Engagement	*Searchable inventory of geoscience resources* across domains and methodology for extending it to additional use cases, with particular focus on resources that are intended to be part of EarthCube. Metadata processing and curation pipeline, including tracking provenance of metadata records as they are processed. Validation mechanisms for metadata documents. Information model (metadata profile and resource ontology) for documenting EarthCube resources. *Engaging* geoscientists into inventory compilation and discovery use cases, via community resource viewers
BB4-Cyberconnector	Standard, Tool	Significantly increase research productivity in the Earth science modeling community. Enable the effective use of the existing Sensor Web data and Earth Observations through *open Web interfaces and metadata standards*. Foster collaborations among Earth system modelers, geospatial information scientists and information technologists. Enhance infrastructure for Earth science research and education

(several projects have added resources) – essentially 'becoming part of EarthCube'".

(3) *BB12 – GeoLink*: The "data.geolink.org URIs can be embedded in existing data repositories/search portals to facilitate discovery of additional information about a given program, expedition, dataset, person, funding award, etc.".

(4) *BB13 – GeoSoft*: "Sharing and reusing software can save effort, promote the use of better quality software, and help disseminate scientific knowledge captured in code."

(5) *BB14-GeoWS*: "Simple discovery and access to data across several domains. Simplified more understandable representations of formats. Integration of multi domain data within a single client".

Responses to Question 8 clearly articulate framework (brokering), standard, tool and community engagement along with enhancing data access, discovery, sharing and reuse across domains.

Table III provides the first three responses of BB responses to survey Questions 9, 10 and 12 from Appendix 3 – building blocks (BB) key technologies, application areas and components:

(1) Question 9 – What are the key technologies being used or developed in your project?
   - Key technologies being used or developed include but not limited to brokering, semantic technologies, GeoServer, GIS database, CINERGI ontology, Open Geospatial Consortium (OGC) Web service and sensor Web specifications, MongoDb, USGS ScienceBase, Tethys Platform, VIVO software, Python, PROV, Resource Description Framework (RDF), Ontology Web Language (OWL), RDF Query Language (SPARQL), Apache Tika, Apache Solr, Apache Tomcat, Apache CXF, apache Nutch, apache Jena, Restful web services, Swagger and Open DAP.

(2) Question 10 – What are the application areas (include science disciplines, if appropriate)?
   - Applications areas, including science disciplines, include but not limited to hydrology, oceanography, meteorology, permafrost research, geosciences, atmospheric sciences, information technology, geochemistry, stratigraphy, sedimentology, critical zone science, metagenomics, paleogeosciences, marine, hydrogeology, paleontology, arctic sciences and geodesy, space science, seismology, hydroecology, marine biology, volcanology, Earth

BBs planned outcomes	Responses (%)
Planned outcomes 1, 2, 3, 4, 5 can be used directly by the scientific community	20
Planned outcomes 1, 2, 3, 4 can be used directly by the scientific community	13
Three of more planned outcomes can be used directly by the scientific community (Note: varied across funded projects and not sequential)	66
Planned outcome 2 can be used directly by the scientific community	93
Planned outcome 1 can be used directly by the scientific community	60

Table II. BBs planned outcomes that can be used directly by the scientific community

Table III.
Sample of BBs key technologies, application areas and components

Project	Question 9 – What are the key technologies being used or developed in your project?	Question 10 – What are the application areas (include science disciplines, if appropriate)?	Question 12 – Which components/products will you deliver at the end of the project? Next to each component in the list, indicate, if appropriate, its maturity (mature product, final development or early development) and the type of interface it could have with external products. Please give URLs if easier for you
BB1-BCube	Brokering, Web crawling using Nutch and semantic technologies	Hydrology, oceanography, meteorology, permafrost research	Broker *neither early nor final development* – evolutionary development – crawler/catalog *early development* – advertising tool *early development*
BB2-CHORDS	Cloud, GeoServer, GIS databases, virtual machine and Docker images, OGC Web services	Geosciences, to include Atmospheric Science, Hydrology, Solid Earth Sciences, Oceanography, Information Technology, etc.	Initial requirements, design and specification of the CHORDS architecture – *Initial prototype* demonstration system – Demonstration of ingest of example real-time data streams
BB3-CINERGI	Comprehensive modular pipeline for enhancing metadata from across geoscience domains and information systems. A set of modules for improving metadata quality. Unified CINERGI ontology for resource tagging and search across domains. Standards-based services for metadata access, provenance management and validation. Catalog visualization interfaces based on HTML5. SOLR & Geoportal – based search UI	Basic inventory and search: across all geoscience domains – Advanced search use cases: for specific use cases, in geochemistry, stratigraphy, sedimentology, critical zone science, metagenomics, paleogeoscience, marine	Community resource viewers (*final development*; Odata, Google Spreadsheet api/JSON); Metadata curation pipeline and several metadata enhancers (*final development*; REST services, SISO 19115/19139 XML); Cross-domain ontology for discovery (*early development*; RDF, REST services); Provenance services (*early development*; REST services, W3C PROV); Validation services (*early development;* REST services, ISO 19115 XML); Searchable inventory (*final development*; CSW/ISO 19115)

surface dynamics, climate modeling, storm-surge prediction and frontal analysis of sea-surface temperatures.

(3) Question 12 – Which components/products will you deliver as the end of the project? Next to each component in the list, indicate, if appropriate, its maturity (mature product, final development or early development) and the type of interface it could have with external products. Please give URLs if easier for you:
   - 60 per cent of the BBs articulated early development/initial prototype for some components;
   - 33 per cent of the BBs did not articulate or state whether the components were mature, final or early development;
   - 13 per cent of the BBs articulated final development of some components;
   - 6 per cent of the BBs articulate mature development of one component; and
   - 6 per cent of the BBs articulated middle development of some components.

Many of the projects were in early development at the time of this study and may have updated information not present in this article. However, there were no working URLs submitted for Question 12 at the time this survey was completed.

Conceptual designs projects
Table IV provides responses to Questions 4, 5, 8, 9 and 10 on the CD-funded project most selected by PIs from different funded interested in interfacing and/or collaborating with CD1-Enterprise Architecture for Transformative Research and Collaboration across the Geosciences (Table VI). This CD project articulated that a CD document could be used directly by the scientific community to improve understanding of EarthCube via pathways, convergence and use through the development of an enterprise architecture framework leveraging unified modeling language information modeling.

The responses to all CD for survey Questions 4, 5, 8, 9 and 10 are available in Appendix 4 – Conceptual design (CD) planned outcomes, use, benefits, key technologies and application areas.

Research coordination networks projects
Table V provides responses to survey Questions 4, 8 and 13 from the top RCNs, RCN 4 – EC3, most selected by PIs from different funded interested in interfacing and/or collaborating (Table VI). See Appendix 5 for full list of RCN responses.

Data management, standards and interfacing
The following aggregate responses to Questions 13, 15 and 17 articulated components' functional areas of support, level of standards compliance and interfacing interests (Figure 3).

Data management. Figure 4 shows 17 funded projects support data access, 19 support data discovery, 17 support data integration, 15 support data management, 12 support modeling, 10 support ontologies and 11 support other (e.g. Web processing services & workflow engine; Analysis & Decision making; Direct Support for Access, Discovery Ontologies – Indirect Support through improved metadata content; Method for enabling reproducibility; Interoperability between model-coupling frameworks; Logic programming; All & Overall Integration and References). Further research is

Table IV.
Sample of CD planned outcomes, use, benefits, key technologies and application areas

Project	Question 4 – Please list up to five planned outcomes of your project including technical, science and/or community outcomes.	Question 5 – Which outcomes can be used directly by the scientific community, for example, specific tools?	Question 8 – Please give specific example(s) of the benefit to the scientific community.	Question 9 – What are the key technologies being used or developed in your project?	Question 10 – What are the application areas (include science disciplines, if appropriate)?
CD1-Enterprise Architecture for Transformative Research and Collaboration Across the Geosciences	Summary of user needs and system requirements. Information model for EarthCube. Conceptual design document. Contributions to EarthCube governance and other developments. Feedback on the design from geoscientists and cyber scientists	Conceptual design document	The goal of conceptual design is to improve understanding of EarthCube development pathways, foster convergence to a common vision and help geoscientists find their place in the EarthCube process, in particular: how to influence EarthCube direction, use EarthCube resources and contribute resources	Enterprise architecture frameworks, RM-ODP, information modeling with UML	Entire geosciences

Project	Question 4 – Please list up to five planned outcomes of your project including technical, science and/ or community outcomes.	Question 8 – Please give specific example(s) of the benefit to the scientific community	Question 13 – What functional area(s) of the EarthCube will your components support? This is not a complete list and you may add additional functions as "Other"
RCN4-EC3	Run field trips that foster collaborations between the field-based geoscience community and the cyberinfrastructure community; Document the effectiveness of bringing these two communities together in a field-based setting; Catalog the use of existing digital field tools; Make a list of recommendations for the development of future digital tools and associated metadata standards; Maintain close connections between existing RCN projects and other EarthCube activities	Facilitating the community convergence on field data standards by having these dialogues both in the field and at various town hall meetings. Once our project is complete, a path forward for the development of a field based application with the appropriate metadata standards should be made clear and then researchers can actively seek out funding for this development	NA

Table V. Sample of RCN planned outcomes, benefits and functional areas of support necessary to explore exactly how the components support functional areas of EarthCube in relation to best practices, guidelines and standards.

Standards compliance. There were a total of 15 responses for Question 15. Figure 5 shows 47 per cent in compliance with published standards, 47 per cent in compliance with community-established standards, 27 per cent community-*de facto*, 13 per cent *ad hoc*, 7 per cent not evaluated and 7 per cent none. This figure shows a larger proportion of the funded projects' components in higher level of standards compliance than previously known.

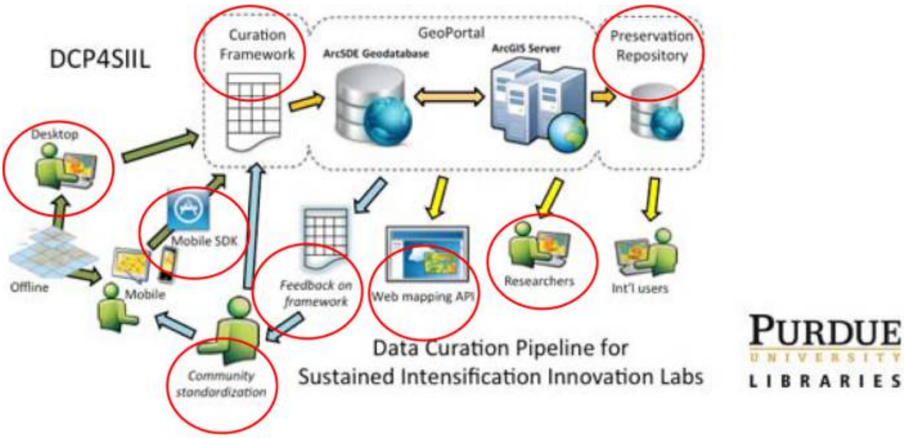
Interfacing. Table VI represents the EarthCube-funded project, and the number of a specific funded project was selected by a PI of another funded project who is interested

#	Answer	Response	%
1	Data Access	17	71%
2	Data Discovery	19	79%
3	Data Integration	17	71%
4	Data Management	15	63%
5	Modeling	12	50%
6	Ontologies	10	42%
7	Other (Please specify)	11	46%

Figure 3. Question 13 – What functional area(s) of the EarthCube will your component support?

#	Answer	Response	%
1	Published Standard	7	47%
2	Not evaluated	1	7%
3	Community-Established	7	47%
4	Community-Defacto	4	27%
5	Ad Hoc	2	13%
6	None	1	7%
8	Not Implemented	0	0%
9	Deprecated Standard	0	0%

Figure 4. Question 15 – What is the level of standards compliance of your components/ resources you are creating?



Source: (Developed by GIS Specialist, Nicole Kong) (used with permission)

Figure 5. DCP4SIIL

EarthCube funded project	Response
BB13 – GeoSoft: Collaborative Open Source Software Sharing for Geosciences	23
BB3 – CINERGI: Community Inventory of EarthCube Resources for Geoscience Interoperability	19
BB12 – GeoLink: Leveraging Semantics and Linked Data for Data Sharing and Discovery in the Geosciences	15
BB14 – GeoWS: Deploying Web Services Across Multiple Geoscience Domains	14
BB1 – BCube: A Broker Framework for Next Generation Geoscience	13
BB8 – Enabling Scientific Collaboration and Discovery through Semantic Connections	12
BB7 – Earth System Bridge: Spanning Scientific Communities with Interoperable Modeling Frameworks	11
BB9 – A Geo-Semantic Framework for Integrating Long-Tail Data and Models	11
BB4 – CyberConnector: Bridging the Earth Observations and Earth Science Modeling for Supporting Model Validation, Verification and Inter-comparison	10
BB11 – GeoDeepDive: A Cognitive Computer Infrastructure for Geoscience	8
BB6 – DisConBB: Integrating Discrete and Continuous Data	8
BB10 – GeoDataspace: Simplifying Data Management for Geoscience Models	7
BB5 – Digtial Crust: An Exploratory Environment for Earth Science Research and Learning	7
BB15 – ODSIP: Specifying and Implementing ODSIP, A Data-Service Invocation Protocol	6
BB2 – CHORDS: Cloud-hosted Real-Time Data Services	6
CD1- Enterprise Architecture for Transformative Research and Collaboration Across the Geosciences	8
CD2 – Developing a Data-Oriented Human-centric Enterprise Architecture for EarthCube	7
CD3 – A Scalable Community Drive Architecture	5
RCN4 – EC3: Challenges of Field Data Collection, Management and Integration	9
RCN5 – iSamplES: The Internet Samples in the Earth Sciences	8
RCN1 – C4P: Collaboration and Cyberinfrastructure for Paleogeosciences	7
RCN6 – SEN: A Sediment Experimentalist Network	7
RCN3 – ECOGEO: Oceanography and Geobiology Environment Omics	6
RCN2 – CReScNT: Coral Reef Science and Cyberinfrastructure Network	4

Table VI. Question 17 – are you interested in interfacing with one or more of the EarthCube projects?

in interfacing and/or collaborating with that specific funded project (e.g. BB13 – GeoSoft: Collaborative Open Source Software Sharing for Geosciences was selected by 23 different PIs representing 23 different EarthCube-funded projects as a funded project in which to interface and/or collaborate). Many PIs, co-PIs, and EC TAC working groups found the responses to Question 17 very interesting and informative in identifying current and future research collaborations in addition to potentially targeting future funding effort.

The 2015 EC3 field trip with some outcomes supported by the survey results

During the August 2-7, 2015, Earth-Centered Communication for Cyberinfrastructure (EC3) Field Trip (as one of the RCN4 activities) discussions with geologists and researchers, some researchers stated the following needs that support some of the survey results:

- the need for representative metadata and application of available relevant metadata and data citation standards (e.g. curation framework, data citation standard and principles http://best-practices.dataverse.org/data-citation/);

- the need for data access, management, storage and tools integration between platforms, systems, desktop-to-mobile applications with community support (e.g. collaborations);
- the need for a trusted digital data repository (e.g. standards-based general and discipline-specific trusted repository that is Open Archival Information Systems [21], Trusted Repository Audit and Checklist[22], ISO 16363/Trusted Digital Repository Checklist[23], Data Seal of Approval[24] or ORNL DAAC[25], Goddard Earth Sciences Data and Information Services Center[26] complaint, to name a few);
- the need to build mobile applications in parallel with desktop (e.g. some desktop applications are not user-friendly on mobile applications – reduced features/ functions); and
- the need for flexible APIs that allows data integration (e.g. commercial and/or proprietary applications limited/restrict modifications to fit researchers' needs).

The above listed needs discovered from the field interactions with the EC3 researchers are supported by Purdue Libraries' Data Curation Pipeline for Sustained Intensification Innovation Labs diagram (Figure 4) that illustrates important functional components in support researchers' and scientists' RDM needs. The RDM needs of the funded projects' researchers/scientists and the functional data management components (Table I) in support of EarthCube are recommended for further study. The components of the Data Curation Pipeline for Sustained Intensification Innovation Labs (Figure 2) circled in red represent components in which some geologists and researchers from the 2015 EC3 Field Trip to Yosemite and Owens Valley expressed challenges in the aggregation, dissemination and preservation of useful research data processed in the field, labs and research.

The EC3 (as one of the RCN4 activities) 2015 Field Trip allowed additional information gathering in support of some of the survey responses as confirmed through personal interactions with several PIs/co-PIs in working groups and meetings.

Types of collaborations

Collaboration across disciplines. The BBs, CDs and RCNs survey represented collaborative research within the geosciences and across multiple disciplines. The PIs, co-PIs and senior personnel are engaged in multiple EarthCube-funded projects research collaborations that are multidisciplinary, interdisciplinary and transdisciplinary:

(1) *Multidisciplinary*: "Researchers work on a problem within their own discipline-based perspectives in parallel with others, fitting their results together as the end of the project" (Heitman and Litewka, 2015).

(2) *Interdisciplinary*: "Research is more intentionally collaborative, with researchers working together on a common problem from their respective disciplinary perspectives" (Heitman and Litewka, 2015).

(3) *Transdisciplinary*: "Researchers approach a common problem from an integrated conceptual framework, identifying the discipline-based methods and approaches that they can take together and redefining both the problem and its solution accordingly" (Heitman and Litewka, 2015).

This study can contribute to the development of the criteria for the success of EarthCube-funded projects (e.g. actual outcomes vs planned outcomes, prototype development and operational success along with sustainability) in addition to compliance to developing federal data management mandates. As many of the EarthCube-funded projects incorporate VREs, which are part of the research life cycle[27], the criterion for success for VREs (Buddenbohm *et al.*, 2015) is a useful model for researchers and scientists developing research life cycle success criteria, factors and metrics workflows in alignment with funders, stakeholders and users.

Conclusions

While the BBs, CDs and RCNs were quite diverse, the foci of each effort overlapped in cyberinfrastructure space. This cyberspace is populated by a series of tools – for data discovery and access, data integration and modeling along with the many facets of data management. Gluing these developments together is the next major step. Like all enduring interlocking structures, there needs to be mortar and also a complete inventory of components. Some additional challenges and gaps identified over the course of this study include but not limited to:

- lack of documentation (metadata), best practices, community standards;
- features/functionality, interoperability, data collection and integration, semantics issues;
- desktop/mobile applications, Web services (e.g. APIs, W3C, SOAP, RESTful, etc.);
- need for relevant researchers' use cases and science drivers scenarios (Figure 1); and
- end-to-end development (e.g. funding beyond prototype/end of funding).

Many of the researchers are aware of RDM policies, best practices, guidelines and standards but may not have the time or resources to effectively implement a full-life cycle RDM workflow that includes standards. Nevertheless, researchers are receptive and welcome collaborations with research data services in academic research libraries as long as overhead is low in terms of time, preparation and implementation. Research data services must provide researchers and scientists user-friendly collaborations/solutions that are easy, intelligent and relevant in support of their research and improve their initial return on investment.

Whether conducting an assessment of faculty research data or in this case conducting an assessment of planned outcomes, expected benefits, enumerated tools and identified functional areas of funded project work, the information gathered from such assessments is necessary to develop overall understanding of the plans, strategies and expected outcomes for developing projects across research data life and funding cycles. Taken together it serves to help development of an EC vision and roadmap for future work. This study provided current information on addressing gaps between funded projects while discovering opportunities for follow-up investigations into the data management components in support of EarthCube and the NSF data management public access plan.

The survey development and deployment was made possible with partnerships between research data services within an academic research library and EarthCube TAC and TGA WG. There are numerous opportunities for research data services operating as separate units or integrated within existing RDM workflows at academic research libraries, higher education institutions and data centers to engage and collaborate with PIs, co-PIs and senior personnel of on-campus and off-campus funded projects (e.g. EarthCube, DataONE, iDigBio, etc.). These partnerships can assist and facilitate researchers' compliance to developing federal data management and sharing requirements, publishers' data policy and recommended repositories (e.g. PLOS ONE [28], SHERPA/RoMEO[29]) and emerging RDM trends as exhibited by the Digital Curation Centre[30], and develop a gateway portal to funded projects' outputs, such as the Research Councils UK Gateway to Research[31]. General RDM training such as those available via MANTRA[32] and discipline-specific data management skills as reflected by the USGS Community for Data Integration Science Support Framework[33] and USGS Fundamental Science Practices[34] are possible success objectives.

Many researchers and scientists involved in EarthCube-funded projects are doing some great work, such as those involved with the GeoSoft (e.g. Geoscience Paper of the Future training series and Ontosoft[35]), CINERGI, EC3 and iSamples projects, to name a few, that address researchers' workflows, new data practices and education and training research data needs in the workforce. However, there still exist numerous opportunities for new interdisciplinary collaborations and partnerships between funded projects and between funded projects and research data services units. These units can connect researchers to potential collaborators, repositories and resources that exceed institutions' current capacities or contribute to these discussions within and outside the community of EarthCube-funded project to enhance collaborations and RCNs. The survey responses revealed that PIs are interested and willing to collaborate on synergistic projects if given testbeds, use cases and science scenarios that address current data, capacity and infrastructure challenges across the geosciences.

Many of the scientists and researchers of EarthCube-funded projects are faculty affiliated with universities, academic research libraries and information schools (iSchools)[36], and/or institutions in experimental program to stimulate competitive research (EPSCoR)[37] jurisdictions[38]. Faculty of funded projects are collaborating or leveraging strong/weak relationships in ways that advance curriculum development, project-based courses and action research commensurate with industry demands such as aggregating and connecting similar data management efforts across disciplines, projects and organizations to solve current and future data management problems. However, other researchers and scientists could benefit from improved collaborations and partnerships with universities' research data services (e.g. assessment, data management, informatics, outreach and science faculty librarians) in bridging the gaps between library and information science data management practices perspectives and domain science data management practices perspectives. Unobtrusive observations and/or participant observer interactions embedded in interdisciplinary working groups and representation in distributed and diverse community of practices (e.g. American Geophysical Union[39], Association of Research Libraries, Center for Open Science[40], Dataverse[41], Dryad, Humanities, Arts, Science, and Technology Alliance and Collaboratory (HASTAC), iDigBio[42], ORCiD, and W3C, to name a few) are some ways libraries can advance their role as partner in campus-wide data management initiatives.

Notes

1. www.nsf.gov/bfa/dias/policy/dmp.jsp
2. www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf
3. http://sparc.arl.org/news/sparc-applauds-senate-committee-action-on-public-access-legislation
4. www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
5. http://earthcube.org/
6. www.ontosoft.org/
7. www.nsf.gov/pubs/2013/nsf13529/nsf13529.htm
8. About EarthCube http://earthcube.org/info/about
9. http://earthcube.org/info/about/working-groups
10. www.nsf.gov/pubs/2013/nsf13529/nsf13529.htm
11. http://earthcube.org/info/about
12. www.nsf.gov/pubs/2013/nsf13529/nsf13529.htm
13. Gap Analysis Model and Figure 1 from the EarthCube Strategic Technology Plan, pp. 15-16.
14. Geoscience 2020 http://earthcube.org/sites/default/files/doc-repository/GeoScience%202020.pdf
15. http://earthcube.org/document/2015/earthcube-collaboratories
16. NOAA Testbeds & Proving Grounds www.testbeds.noaa.gov/
17. http://earthcube.org/group/tech-hands-meeting
18. http://earthcube.org/info/about/2015-all-hands-meeting
19. http://earthcube.org/group/technology-architecture-committee
20. http://dlib.org/dlib/september15/buddenbohm/09buddenbohm.html
21. http://public.ccsds.org/publications/archive/650x0m2.pdf
22. www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac
23. www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/iso16363
24. www.datasealofapproval.org/en/
25. http://daac.ornl.gov/
26. http://disc.sci.gsfc.nasa.gov/
27. http://webarchive.nationalarchives.gov.uk/20140702233839/http:/www.jisc.ac.uk/whatwedo/campaigns/res3/jischelp.aspx#ideas
28. http://journals.plos.org/plosone/s/data-availability
29. www.sherpa.ac.uk/romeo/browse.php?la=en&fIDnum=|&mode=simple
30. /www.dcc.ac.uk/
31. www.rcuk.ac.uk/research/gtr/
32. http://datalib.edina.ac.uk/mantra/

33. www.usgs.gov/cdi/cdi-ssf/cdi-ssf-components.pdf
34. www.usgs.gov/fsp/default.asp
35. https://my.usgs.gov/confluence/display/cdi/Geoscience+Paper+of+the+Future+Training
36. http://ischools.org/
37. www.nsf.gov/od/oia/programs/epscor/index.jsp
38. www.nsf.gov/od/oia/programs/epscor/images/cohortmap2013.pdf
39. http://sites.agu.org/
40. https://cos.io/
41. http://dataverse.org/
42. www.idigbio.org/

References

Buddenbohm, S., Enke, H., Hofmann, M., Klar, J., Neuroth, H., and Schwiegelshohn, U. (2015), "Success criteria for the development and sustainable operation of virtual research environments", D-Lib Magazine, September/October, Vol. 21 Nos 9/10, available at: http://dlib.org/dlib/september15/buddenbohm/09buddenbohm.html (accessed 3 November 2015).

Gil, Y., Arctur, D., Kelbert, A., Pearlman, J., Rubin, K. and Wood-Charlson, E.M. (2015), "Moving earth forward through science and technology collaboratories", available at: http://earthcube.org/sites/default/files/doc-repository/Collaboratories-13May2015.pdf (accessed 5 November 2015).

Greenwood, D.J. and Levin, M. (2000), "Reconstructing the relationships between universities and society through action research", in Denzin, N.K. and Lincoln, Y.S. (Eds), *Handbook of Qualitative Research*, 2nd ed., Sage, Thousand Oaks, CA, pp. 85-106.

Heitman, E. and Litewka, S. (2015), CITI Program – Collaborative Institutional Training Initiative at the University of Miami, RCR Basic Course – Collaborative Research (RCR-Basic).

Kelbert, A. (2014), "Science and cyberinfrastructure: the chicken and egg problem", *Eos Trans. AGU*, Vol. 95 No. 49, pp. 458-459. doi: 10.1002/2014EO490006.

Martin, J. (1992), *Cultures in Organizations: Three Perspectives*, Oxford University Press.

Michener, W.K. (2015), "Ten simple rules for creating a good data management plan", *PLOS Computational Biology*, available at: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004525 (accessed 3 November 2015).

National Science Foundation (2010), "Scientists seeking NSF funding will soon be required to submit data management plans", Press release 10-077, available at: www.nsf.gov/news/news_summ.jsp?cntn_id=116928 (accessed 3 November 2015).

Saha, P. (2004), "Analyzing the open group architecture framework from the GERAM perspective", The Open Group, Tech. Rep, available at: www.opengroup.org/architecture/wp/saha/TOGAF_GERAM_Mapping.pdf (accessed 3 November 2015).

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L. Read, E., Manoff, M., and Frame, M. (2011), "Data sharing by scientists: practices and perceptions", *PLoS One*, Vol. 6, p. e21101.

Further reading

Brandt, D.S. (2015), "DLF E-Research Network 2015 Webinar on Introduction/Research Data Management Services in Academic Libraries for 2015 DLF E-Research Network Cohort", 13 May.

DLF e-Research Network (2015), *DLF e-Research Network 2015 Cohort*, available at: www.diglib.org/groups/e-research-network/ (accessed 3 November 2015).

Johnson, A.M. and Bresnahan, M.W. (2015), "DataDay! Designing and assessing a research data workshop for subject librarians", *Journal of Librarianship and Scholarly Communication*, Vol. 3 No. 2, p. eP1229. available at: http://dx.doi.org/10.7710/2162-3309.1229

Mookerjee, M., Vieira, D., Chan, M.A., Gil, Y., Pavlis, T.L., Spear, F.S. and Tikoff, B. (2015), "Field data management integrating cyberscience and geoscience", *Earth & Space Science News (EOS)*, Vol. 96 No. 20, pp. 18-21.

Research Councils, UK (2015), *RCUK Common Principles on Data Policy*, available at: www.rcuk.ac.uk/research/datapolicy/ (accessed 3 November 2015).

The Royal Society (2012), "Final report – Science as an open enterprise", available at: https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/ (accessed 3 November 2015).

Corresponding author
Plato Smith II can be contacted at: psmith4@unm.edu

Appendix 1

Table AI. Building blocks (BBs), conceptual designs (CDs) and research coordination networks (RCNs)

ID	Project	PI	Co-PI	Senior personnel	Start date	End date
BB1	BB – BCube: A Broker Framework for Next Generation Geoscience	1	4	6	9/15/13	8/31/15
BB2	BB – CHORDS: Cloud-hosted Real-Time Data Services	1	4	4	9/14/14	8/31/16
BB3	BB – CINERGI: Community Inventory of EarthCube Resources for Geoscience Interoperability	1	7	3	9/01/13	8/31/15
BB4	BB – CyberConnector: Bridging the Earth Observations and Earth Science Modeling for Supporting Model Validation, Verification and Inter-comparison	1	4	0	9/01/14	8/31/16
BB5	BB – Digtial Crust: An Exploratory Environment for Earth Science Research and Learning	1	4	2	9/01/14	8/31/16
BB6	BB – DisConBB: Integrating Discrete and Continuous Data	1	3	2	9/15/13	8/31/15
BB7	BB -Earth System Bridge: Spanning Scientific Communities with Interoperable Modeling Frameworks	1	6	0	9/15/13	8/31/15
BB8	BB – Enabling Scientific Collaboration and Discovery through Semantic Connections	1	3	5	9/01/14	8/31/16
BB9	BB – A Geo-Semantic Framework for Integrating Long-Tail Data and Models	1	1	0	9/01/14	8/31/16
BB10	BB – GeoDataspace: Simplifying Data Management for Geoscience Models	1	5	1	9/01/14	8/31/16
BB11	BB – GeoDeepDive: A Cognitive Computer Infrastructure for Geoscience	1	2	0	9/15/13	8/31/16
BB12	BB – GeoLink: Leveraging Semantics and Linked Data for Data Sharing and Discovery in the Geosciences	1	5	0	9/01/14	8/31/16
BB13	BB – GeoSoft: Collaborative Open Source Software Sharing for Geosciences	1	4	0	9/01/14	8/31/16
BB14	BB – GeoWS: Deploying Web Services Across Multiple Geoscience Domains	1	4	4	9/15/13	8/31/16
BB15	BB – ODSIP: Specifying and Implementing ODSIP, A Data-Service Invocation Protocol	1	4	0	9/15/13	8/31/15
CD1	CD – Enterprise Architecture for Transformative Research and Collaboration Across the Geosciences	1	4	0	9/15/13	8/31/16
CD2	CD – Developing a Data-Oriented Human-centric Enterprise Architecture for EarthCube	1	1	2	9/15/13	8/31/15
CD3	CD – A Scalable Community Drive Architecture	1	4	3	6/15/14	5/31/16
RCN1	RCN – C4P: Collaboration and Cyberinfrastructure for Paleogeosciences	1	3	4	9/15/13	8/31/16
RCN2	RCN – CReScNT: Coral Reef Science and Cyberinfrastructure Network	1	4	0	9/01/14	8/31/16
RCN3	RCN – ECOGEO: Oceanography and Geobiology Environment Omics	1	0	12	9/01/14	8/31/16
RCN4	RCN – EC3: Challenges of Field Data Collection, Management and Integration	1	0	0	9/15/2013	8/31/16
RCN5	RCN – iSamplES: The Internet Samples in the Earth Sciences	1	1	0	9/01/14	8/31/16
RCN6	RCN – SEN: A Sediment Experimentalist Network	1	3	0	8/01/13	7/31/16

Appendix 2

Table AII. Building blocks (BBs) categories, planned outcomes, use and availability

Project	Category	Question 4 – Please list up to five planned outcomes of your project including technical, science and/or community outcomes	Question 5 – Which outcomes can be used directly by the scientific community, for example, specific tools?	Question 7 – What outcomes will be available for use by the All Hands Meeting in May from the list in Question 4?
BB1-BCube	Framework or testbed	A cloud-based brokering *framework* that is mediating query and access across diverse data sources, serving science use cases in hydrology, oceanography meteorology and cryospheric research. A queryable catalog of Earth Science resources (data sets and services by which data are made accessible), and the cloud-based Web crawler/reasoner that is used to populate and update this catalog. A *report on the current state of metadata and service description standards* and the impacts these have on interoperability, including recommendations for improvements. A prototype tool for scientists who seek an alternative to depositing their data in a repository and wish to advertise their data in a manner that is discoverable on the Web. A cloud-based interoperability testbed	1, 2, 4	1-5
BB2-CHORDS	Community engagement	*Engagement* with real-time geosciences community, including real-time data providers and consumers. Demonstration of a prototype system that enables cloud-sourced, standards-based, easy access to real-time sensor data and otherwise proprietary or unavailable real-time sensor data streams. Specifications and documentation for scientists to use the prototype system	1, 2, 3	2
BB3-CINERGI	Framework or testbed, Standard, Tool, Community Engagement	*Searchable inventory of geoscience resources* across domains and methodology for extending it to additional use cases, with particular focus on resources that are intended to be part of EarthCube. Metadata processing and curation pipeline, including tracking provenance of metadata records as they are processed. Validation mechanisms for metadata documents. Information model (metadata profile and resource ontology) for documenting EarthCube resources. *Engaging* geoscientists into inventory compilation and discovery use cases, via community resource viewers	2, 3, 5	1, 2, 3, 5
BB4-Cyberconnector	Standard, tool	Significantly increase research productivity in the Earth science modeling community; **e**nable the effective use of the existing Sensor Web data and Earth Observations through *open Web interfaces and metadata standards*; **f**oster collaborations among Earth system modelers, geospatial information scientists and information technologists; and **e**nhance infrastructure for Earth science research and education	2	None

(*continued*)

Table AII.

Project	Category	Question 4 – Please list up to five planned outcomes of your project including technical, science and/or community outcomes.	Question 5 – Which outcomes can be used directly by the scientific community, for example, specific tools?	Question 7 – What outcomes will be available for use by the All Hands Meeting in May from the list in Question 4?
BB5-Digital Crust	Tool, Community Engagement	Demonstration of a *data system scaffolding* that can be used to deploy domain specific community databases in a federated information system. *Prototyping* of technology to integrate geospatial data represented in 2-D, 3-D, gridded, tessellated, or vector geometry; Development of infrastructure for registering information exchange models as a resource for cross domain concept mapping and data integration; Compilation of a continental scale permeability model that demonstrated usage of the Digital Crust components	1, 2	2, 3
BB6-DissConBB	Standard, tool	First land-surface hydrology workflow to model & forecast high-resolution horizontal flows holistically at the US national scale. Integration and downscaling of WRF-Hydro modeling output with RAPID streamflow routing model, across the USA, and integration of ECMWF forecasts with RAPID for comparison. Integration and downscaling of Europe's ECMWF and Global Flood Awareness System (GloFAS) with RAPID in the USA and other countries. *Data model conversion software to transform WaterML information content into netCDF/CF content, and to extract WaterML time series data from NetCDF layers* of satellite observations, demonstrated in a Snow Inspector tool. Downscaling tools for ArcGIS, to enable consistent downscale global/regional weather forecasts to local level	1, 2, 3, 4, 5	1, 2
BB7-Earth System Bridge	Framework or Testbed	Develop a standardized method for describing all of the features, limitations and capabilities of any modeling framework. To meet this goal we are developing an *Earth System–Framework Description Language (ES-FDL)* which identifies the key metadata items that are needed to fully describe an arbitrary modeling framework, something like an architectural blueprint. Develop general, cross-domain mechanisms for mediating between the different semantic descriptions of model variables and assumptions that are used within different modeling frameworks. Since different process model components use their own abbreviations and labels for their input variables, output variables and other metadata, modeling frameworks generally make use of controlled vocabularies or "standard names" Develop standardized metadata for technical descriptions of computational models. Develop and demonstrate a general mechanism that allows any model–after relatively simple, framework-agnostic modifications–to be used as a plug-and-play component in any major modeling framework. Demonstrate the new capabilities provided by Earth System Bridge technology by using them to couple operational/federal atmosphere-ocean models at the global-to-regional scale to academic (or operational)	1, 2, 3, 4, 5	None

(*continued*)

Table AII.

Project	Category	Question 4 – Please list up to five planned outcomes of your project including technical, science and/or community outcomes.	Question 5 – Which outcomes can be used directly by the scientific community, for example, specific tools?	Question 7 – What outcomes will be available for use by the All Hands Meeting in May from the list in Question 4?
		hydrologic/land/coastal models at the regional-to-local scale Earth System Framework Description Language (ES-FDL) for describing modeling frameworks. Model Coupling Metadata (MCM) schema for describing models. Browser-based, graphical MCM Tool for entering model metadata via MCM. Ontology and extensions to the CSDMS Standard Names. Crosswalk between the CSDMS Standard Variable Names and CF Standard Names. Crosswalk between CSDMS Standard Variable Names and CUAHSI VariableName CV. Basic Model Interface (BMI) adapters that allow any BMI-enabled model to be used as a component in multiple modeling frameworks (i.e. CSDMS, ESMF/NUOPC, OMS, OpenMI, Pyre)		
BB8-Enabling Scientific Collaboration	Tool, community engagement	Functional demonstration models that display the utility of linked data structures in the coordination and organization of virtual organizations and their products. *New tools and processes to better support information and linked data exchange in geoscience communities*. Recommendations for how the EarthCube community can engage user communities in linked data projects	2	Null
BB9-GeoSemantic Framework	Framework or testbed, standard, tool	Scientific outcomes *Geosemantics framework* will directly augment the multidisciplinary interaction between different geoscience communities by minimizing the human intervention in semantic mediation between resources (data and models) and their context ambiguity, and support the crosswalks among geoscience standard names. Technical outcomes Knowledge base: An RDF triple store for storing meta-information of ontologies, list of the registered data and models Web services and thesaurus of standard vocabularies. GeoSemantics Wiki: it is a combination of wiki and Semantic Web technology, for geoscience communities to develop mini-ontologies about their disciplines. Data Networks Service: It is a Web service for inferring the contextual relationships between given resources. Data Alignment service: It is Web service that can handle the *semantic mediation between two geoscience resources*. Semantic tagging service: It is a service for tagging data with ontology-based standard names and incorporating semantics in the development of simulation models	2, 3, 4, 5	1-5 (test)

(*continued*)

Project	Category	Question 4 – Please list up to five planned outcomes of your project including technical, science and/or community outcomes.	Question 5 – Which outcomes can be used directly by the scientific community, for example, specific tools?	Question 7 – What outcomes will be available for use by the All Hands Meeting in May from the list in Question 4?
BB10-GeoDataSpace	Tool, community engagement	A *hosted service for managing, curating and preserving your geounits*. A client for automatically creating geounits. A distributed search facility. Demonstration of 4 reproducibility challenges solved by using geounits. A *community agreement* on how reproducibility must be enabled	1, 2, 3, 4	4
BB11-GeoDeepDive	Tool	*TDM-ready digital library facility* that has pre-processed published scientific documents from all relevant publishers and content owners that are accessible for use by EarthCube community members. Working examples of domain-specific TDM applications (doi:10.1371/journal.pone.0113523) Cyberinfrastructure to support the rapid and large-scale deployment of additional document processing *toolkits* on the TDM library holdings (e.g. image processing and analysis, table parsing and reading software). Method to support the submission of geoscience-related content to the TDM library facility. Cyberinfrastructure to support the development of computing-intensive machine reading and learning applications for data extraction by EarthCube end-users	1, 2, 3, 4, 5	1, 2, 3, 4
BB12-GeoLink	Framework or Testbed	Outcomes will include a set of reusable *ontology design patterns (ODPs)* that describe core geoscience concepts, a network of Linked Data published by participating repositories using those ODPs and *Tools* to facilitate discovery of related content in multiple repositories	2	1
BB13-GeoSoft	Framework or testbed, tool, report, community engagement	*TurboSoft portal* to publish and browse software. GeoSoft portal for community software sharing. SoftCamp educational materials for scientists to learn about best practices in software sharing. The *OntoSoft ontology* to describe metadata for geosciences software. 5. Information retrieval approaches and software to automatically extract software metadata and perform code license analysis on software from Github and Apache and other repositories for *software communities in the geosciences*	1, 2, 3, 4	1, 2, 3
BB14-GeoWS	Framework, tool, community engagement	A good *collection of simplified restful user interfaces* to promote access to data in 6 funded domains and 8 unfunded collaborating centers or data collections. Standardized documentation. Similar URL builders to provide examples of data center interactions *Integration with CINERGI and B-Cubed Building Block Projects*. Simplified text based data formats that promote interoperability 6. Integration of metadata and scientific data (time series, point values) within existing clients (GeoMapApp, IDV, Brokering Portal)	1, 5, 6	

(*continued*)

Table AII.

Project	Category	Question 4 – Please list up to five planned outcomes of your project including technical, science and/or community outcomes.	Question 5 – Which outcomes can be used directly by the scientific community, for example, specific tools?	Question 7 – What outcomes will be available for use by the All Hands Meeting in May from the list in Question 4?
BB15-ODSIP	Framework or Testbed	*A draft specification for ODSIP (an Open Data Services Invocation Protocol)* as a DAP4 extension. A prototype implementation of ODSIP in open-source software and libraries. Demos of how the *prototype* ODSIP supports three geoscience scenarios: a. Accelerated visualization/analysis of model outputs on non-rectangular meshes (over coastal North Carolina). b. Dynamic downscaling of climate predictions for regional utility (over Hawaii). c. Feature-oriented retrievals of satellite imagery (focusing on satellite-derived sea-surface-temperature fronts)	2	3a. or 3b.

Table AII.

Appendix 3

Table AIII. Building blocks (BBs) key technologies, application areas and components

Project	Question 9 – What are the key technologies being used or developed in your project?	Question 10 – What are the application areas (include science disciplines, if appropriate)?	Question 12 – Which components/products will you deliver at the end of the project? Next to each component in the list, indicate, if appropriate, its maturity (mature product, final development or early development) and the type of interface it could have with external products. Please give URLs if easier for you
BB1-BCube	Brokering, Web crawling using Nutch and semantic technologies	Hydrology, oceanography, meteorology, permafrost research	Broker *neither early nor final development* – evolutionary development – crawler/catalog *early development* – advertising tool *early development*
BB2-CHORDS	Cloud, GeoServer, GIS databases, virtual machine and Docker images, OGC Web services	Geosciences, to include Atmospheric Science, Hydrology, Solid Earth Sciences, Oceanography, Information Technology, etc.	Initial requirements, design and specification of the CHORDS architecture – *Initial prototype* demonstration system – Demonstration of ingest of example real-time data streams
BB3-CINERGI	Comprehensive modular pipeline for enhancing metadata from across geoscience domains and information systems. A set of modules for improving metadata quality. Unified CINERGI ontology for resource tagging and search across domains. Standards-based services for metadata access, provenance management and validation; Catalog visualization interfaces based on HTML5. SOLR & Geoportal – based search UI	Basic inventory and search: across all geoscience domains – Advanced search use cases: for specific use cases, in geochemistry, stratigraphy, sedimentology, critical zone science, metagenomics, paleogeoscience, marine	Community resource viewers (*final development*; Odata, Google Spreadsheet api/JSON); Metadata curation pipeline and several metadata enhancers (*final development*; REST services, SISO 19115/19139 XML); Cross-domain ontology for discovery (*early development*; RDF, REST services); Provenance services (*early development*; REST services, W3C PROV); Validation services (*early development;* REST services, ISO 19115 XML); Searchable inventory (*final development*; CSW/ISO 19115)
BB4-Cyberconnector	Standard-based geospatial interoperability technology, chainable Web services, workflow engine, OGC Web service and sensor Web specifications, product virtualization, distributed data discovery and access	The CyberConnector is intended to be a multiple-disciplinary in nature. The project will demonstrate that CyberConnnector can support modeling and applications in atmosphere, ocean, cloud and land surface process	
BB5-Digital Crust	MongoDb, ElasticSearch, USGS ScienceBase	Near term: stratigraphy, hydrogeology, paleontology–Long term: technology is not inherently domain specific; can be adapted for many domains	Extract-Transform-Caching middleware (link document dB to Annotation service, Data Access Services, Discovery Services), *early development*, REST interact with document dB, Web interfaces with external components TBD. Information Exchange registry, *early development*, REST services for information model retrieval, registration, vocabulary usage and schema mapping

(*continued*)

Project	Question 9 – What are the key technologies being used or developed in your project?	Question 10 – What are the application areas (include science disciplines, if appropriate)?	Question 12 – Which components/products will you deliver at the end of the project? Next to each component in the list, indicate, if appropriate, its maturity (mature product, final development or early development) and the type of interface it could have with external products. Please give URLs if easier for you
BB6-DisConBB	Tethys Platform, developed by the NSF EPSCoR CI-Water project, is proving to be a key useful technology for developing and deploying cloud based applications	Primary application area is for US and global flood forecasting (DisConBB tools plus data and workflows). The core DisConBB methodology, information model and tools could be useful in many applications at the intersection of discrete and continuous data types, such as seismic studies, polar studies of ice column composition, etc.	1. nwis2ncdf – an R package for converting WaterML streams from the National Water Information System to NetCDF/CF format. Custom R scripts can be run locally using the nwis2ncdf package. We are currently providing 24-hour snapshots via THREDDs (http://nwisnfie-a.cloudapp.net/thredds/catalog/NWIS/catalog.html) and Unidata's LDM *Mature development*. 2. Tethys apps – several data conversion and visualization apps at http://tools.hydroshare.org
BB7-Earth System Bridge	See Question 5. Earth System Framework Description Language (ES-FDL) for describing modeling frameworks. Model Coupling Metadata (MCM) schema for describing models. Browser-based, graphical MCM Tool for entering model metadata via MCM. Ontology and extensions to the CSDMS Standard Names; Crosswalk between the CSDMS Standard Variable Names and CF Standard Names. Crosswalk between CSDMS Standard Variable Names and CUAHSI VariableName CV. Basic Model Interface (BMI) adapters that allow any BMI-enabled model to be used as a component in multiple modeling frameworks (i.e. CSDMS, ESMF/NUOPC, OMS, OpenMI, Pyre)	Geoscience computational models and model-coupling frameworks	Earth System Framework Description Language (ES-FDL) for describing modeling frameworks. – *middle development* stage; Model Coupling Metadata (MCM) schema for describing models. – *middle development* stage; Browser-based, graphical MCM Tool for entering model metadata via MCM. – *middle development* stage; Ontology and extensions to the CSDMS Standard Names – *middle development* stage, TTL and OWL; Crosswalk between the CSDMS Standard Variable Names and CF Standard Names. – *middle development* stage, SKOS; Crosswalk between CSDMS Standard Variable Names and CUAHSI VariableName CV. – middle development stage, SKOS; Basic Model Interface (BMI) adapters that allow any BMI-enabled model to be used as a component in multiple modeling frameworks (i.e. CSDMS, ESMF/NUOPC, OMS, OpenMI, Pyre). – *early to middle development* stage

(*continued*)

Table AIII.

Project	Question 9 – What are the key technologies being used or developed in your project?	Question 10 – What are the application areas (include science disciplines, if appropriate)?	Question 12 – Which components/products will you deliver at the end of the project? Next to each component in the list, indicate, if appropriate, its maturity (mature product, final development or early development) and the type of interface it could have with external products. Please give URLs if easier for you
BB8-Enabling Scientific Collaboration	The VIVO software suite, http://vivoweb.org/	Arctic sciences and geodesy	Our project is focusing on two case studies, data and information related to the Bering Sea Project (supported by NCAR's Earth Observing Laboratory) and data and information supported by UNAVCO. The components/products include: VIVO instance for the Bering Sea Project – *early development* type of interface = SPARQL endpoint VIVO instance for UNAVCO – *early development,* type of interface = SPARQL endpoint Service that allows exchange of linked data across VIVO instances – *early development*, type of interface = *early development*
BB9-Geo Semantic Framework	Play framework as a Web application framework to integrate the components and APIs related to the Geosemantics framework. Graph database for serializing information about resources (data, models and ontologies)	Knowledge discovery. Semantic Interoperability between resources (data and models)	Knowledge base – *early development*; GeoSemantics Wiki – *early development*; Data Network Service – *final development*; Data Alignment service – *final development*; Semantic tagging service – *final development;* Ontology mapping – *early development*; Logic Ingestion – *early development*; Semantic processor – *final development*
BB10-Geodataspace	Reproducibility tools (PTU), Python, PROV standard, Annotation Standards, Globus services	Hydrology, Space Science, Seismology	A hosted service for managing, curating, sharing and preserving your geounits; A client for automatically creating geounits; A distributed search facility; Demonstration of 4 reproducibility challenges solved by using geounits; A community agreement on how reproducibility must be enabled

(*continued*)

Table AIII.

Project	Question 9 – What are the key technologies being used or developed in your project?	Question 10 – What are the application areas (include science disciplines, if appropriate)?	Question 12 – Which components/products will you deliver at the end of the project? Next to each component in the list, indicate, if appropriate, its maturity (mature product, final development or early development) and the type of interface it could have with external products. Please give URLs if easier for you
BB11-GeoDeepDive	Developing a complete pipeline for accessing, pre-processing, subsetting and delivering content from the scientific literature for the purposes of text and data mining. In addition, we will be developing cyberinfrastructure to support domain-specific TDM applications that identify and extract data and relationships among them from the scientific literature	Building structured databases from unstructured or heterogeneously structured sources (i.e. the scientific literature)	The biggest deliverable will be a cyberinfrastructure to support the secure access and storage of privileged content (e.g. from commercial publishers), the large-scale pre-processing of this content for standard TDM tool kits (e.g. NLP/OCR), the on-demand processing of content for specialized applications (e.g. image retrieval and analysis). The second biggest deliverable will be a new type of TDM-ready digital library in which content is ready for large-scale (and legal) processing and analysis
BB12-GeoLink	Semantic Web technologies include Resource Description Framework (RDF), Ontology Web Language (OWL) and RDF Query Language (SPARQL), via HTTP	GeoLink integrates content from programs, expeditions, datasets, persons, funding awards, etc., across the academic geosciences	Outcomes will include a set of reusable ontology design patterns (ODPs) that describe core geoscience concepts (published at schema.geolink.org), a network of Linked Data published by participating repositories using those ODPs (results merged and integrated at data.geolink.org), and tools to facilitate discovery of related content in multiple repositories
BB13-Geosoft	Apache Tika, Apache Solr, Apache Tomcat, Apache CXF, Apache Nutch, Apache Jena	The main application area of our target demonstration is the Critical Zone Observatories, which include a range of research in geosciences. We are also working with early career scientists in hydrology, hydroecology, marine biology, volcanology, Earth surface dynamics and climate modeling to understand their requirements and help their communities share software and adopt best practices	TurboSoft portal to publish and browse software-*early development*; GeoSoft portal for community software sharing -ed 3) SoftCamp educational materials for scientists to learn about best practices in software sharing–*early development*
BB14-GeoWS	Restful Web services Swagger	Restful Web services Swagger	The Web services at the six funded partners with documentation. RAMADDA expanded to support GeoWS style services

(*continued*)

Table AIII.

Project	Question 9 – What are the key technologies being used or developed in your project?	Question 10 – What are the application areas (include science disciplines, if appropriate)?	Question 12 – Which components/products will you deliver at the end of the project? Next to each component in the list, indicate, if appropriate, its maturity (mature product, final development or early development) and the type of interface it could have with external products. Please give URLs if easier for you
BB15-ODSIP	Some widely used, Web-based, data-access infrastructure (OpenDAP) will be extended to offer new server-side functionality. Conceptually, this is evolving data-access-as-a-service toward data-processing-software-as-a-service. Because the DAP protocols are accompanied by production-quality software libraries in multiple languages	In general, we consider the DAP infrastructure to be discipline-neutral and hence applicable across all domains (where mathematical functions and structured data pertain). Initially, however, the ODSIP extensions will target specific geoscience applications: climate-model downscaling, storm-surge prediction and frontal analysis of sea-surface temperatures	A draft specification for an Open Data Services Invocation Protocol, an *early-development* extension to a mature-product spec. A prototype implementation of ODSIP in open-source software/libraries, an *early-development* extension to mature-product software. Prototype server-side codes for: Manipulating variables (i.e. coverages) organized on non-rectangular meshes. Calculating statistics and un-modeled parameters from climate- and weather-model outputs. Retrieval, from coverages, of features based on range-value constraints

Table AIII.

Appendix 4

Table AIV. Conceptual design (CD) planned outcomes, use, benefits, key technologies and application areas

Project	Question 4 – Please list up to five planned outcomes of your project including technical, science and/or community outcomes.	Question 5 – Which outcomes can be used directly by the scientific community, for example, specific tools?	Question 8 – Please give specific example(s) of the benefit to the scientific community.	Question 9 – What are the key technologies being used or developed in your project?	Question 10 – What are the application areas (include science disciplines, if appropriate)?
CD1-Enterprise Architecture for Transformative Research and Collaboration Across the Geosciences	Summary of user needs and system requirements. Information model for EarthCube. Conceptual design document. Contributions to EarthCube governance and other developments. Feedback on the design from geoscientists and cyber scientists	Conceptual design document	The goal of Conceptual Design is to improve understanding of EarthCube development pathways, foster convergence to a common vision, and help geoscientists find their place in the EarthCube process, in particular: how to influence EarthCube direction, use EarthCube resources and contribute resources	Enterprise architecture frameworks, RM-ODP, information modeling with UML	Entire geosciences

(*continued*)

Table AIV.

Project	Question 4 – Please list up to five planned outcomes of your project including technical, science and/or community outcomes	Question 5 – Which outcomes can be used directly by the scientific community, for example, specific tools?	Question 8 – Please give specific example(s) of the benefit to the scientific community	Question 9 – What are the key technologies being used or developed in your project?	Question 10 – What are the application areas (include science disciplines, if appropriate)?
CD2-Developing a Data-Oriented Human-Centric Enterprise Architecture for EarthCube	The final design will be in a four-volume report containing different views which describe the EarthCube conceptual architecture from different perspective. *Volume I* is an overview document and comprises an executive summary of the EarthCube conceptual architecture, serving as an overview in the initial phases of conceptual architecture development. *Volume II* is the major body of the design product. It outlines all the conceptual architectural design components or		Resource discovery is one of the most common activity of geoscientists. Figure 1 illustrates the workflow process of resource discovery activity. The responsibilities of end-users and different types of supporting roles are suggested. Cyberinfrastructure of EarthCube will be provide to support the activity of resource discovery. Through resource discovery, users can search resources, such as data, services (data service, processing service and computing service) and	Null	Null

(*continued*)

Project	Question 4 – Please list up to five planned outcomes of your project including technical, science and/or community outcomes.	Question 5 – Which outcomes can be used directly by the scientific community, for example, specific tools?	Question 8 – Please give specific example(s) of the benefit to the scientific community	Question 9 – What are the key technologies being used or developed in your project?	Question 10 – What are the application areas (include science disciplines, if appropriate)?
	viewpoints (refer to section 4 for details). *Volume III* provides taxonomy of the EarthCube enterprise augmented with semantics relations. *Volume IV* describes an example of utilization of this conceptual architecture for a geoscience project		tools/applications that are registered in EarthCube		
CD3-A Scalable Community Drive Architecture	Conceptual architecture model, CD document, Progress reports, Engagement/OutReach presentations, A set of findings and recommendations	Conceptual architecture model; CD document	The architecture concept can serve as a roadmap for building an extensible and sustainable EarthCube system	IEEE 1471, TOGAF, Zackman, etc architectural specifications for our framework	Architecture is relevant to EC at large

Table AIV.

Appendix 5

Table AV. Research coordination networks (RCN) planned outcomes, benefits and functional areas of support

Project	Question 4 – Please list up to five planned outcomes of your project including technical, science and/or community outcomes	Question 8 – Please give specific example(s) of the benefit to the scientific community	Question 13 – What functional area(s) of the EarthCube will your components support? This is not a complete list and you may add additional functions as "Other"
RCN1-C4P	Catalog and provide a clearinghouse for existing community CI resources; stimulate and facilitate communication and collaboration between domain scientists and cyber/computer scientists through webinars and workshops. Conduct outreach events to broaden participation. Coordinate with related CI developments in the framework of EarthCube	New collaborations formed through the C4P events and discussion; Paleobiological Data Consortium, USGS-PBDB-IGSN collaboration; Bob Hazen's Deeptime Data Infrastructure & C4P; Catalog of resources available to anyone with interest	Data access, data discovery, data integration, data management
RCN2-CRESCNT	Development of visualization products in collaboration with educators and resource managers, connections among place-based collaborative networks, defining data standards for the coral reef community, broad transfer of disciplinary expertise among geoscientists, cybertechnologists, medical scientists and graphic artists (media), user-friendly Web-based visualization tools accessible to the public that serve to showcase the value of coral reef science and the innovative approaches employed by the EarthCube program	Our project breaks down silo's within our community and improves capacity for cross talk and sharing of data within the geosciences. The data standards will be key to effective data sharing	Data access, data discovery, data integration, data management, modeling
RCN3-ECOGEO	Resource viewer, incl. data sets, databases, tools for analysis -Identified cloud computing resource -Use cases -Training webinars	The goals of ECOGEO directly reflect the needs in the "omics" research community. We have a critical need for an open access data portal, as well as long term storage and curation of data sets/databases, federated standards and analysis tools, including statistics and visualization. Otherwise, our collaborative science will not be able to move forward as a community effort to archive, access, share and analyze the "big data" in environmental omics that are already in existence, currently being collected, and will undoubtedly be generated in the future	Data access, data discovery, data integration, data management
RCN4-EC3	Run field trips that foster collaborations between the field-based geoscience community and the cyberinfrastructure community; Document the effectiveness of bringing these two communities together in a field-based setting; Catalog the use of existing digital field tools; Make a list of recommendations for the development of future digital tools and associated metadata standards; Maintain close connections between existing RCN projects and other EarthCube activities	Facilitating the community convergence on field data standards by having these dialogues both in the field and at various town hall meetings. Once our project is complete, a path forward for the development of a field based application with the appropriate metadata standards should be made clear and then researchers can actively seek out funding for this development	NA

(*continued*)

Project	Question 4 – Please list up to five planned outcomes of your project including technical, science and/or community outcomes	Question 8 – Please give specific example(s) of the benefit to the scientific community	Question 13 – What functional area(s) of the EarthCube will your components support? This is not a complete list and you may add additional functions as "Other"
RCN5-iSamples	iSamples Stakeholder survey; iSamples Knowledge Hub (semantic wiki); inventory of sample repositories in CINERGI; Educational modules for early career scientists by early career scientists; Working groups	Enhanced access to samples and sample collections; guidance on practices and policies for sample sharing and citation of samples; consistent protocols for identification and documentation of samples; roadmap toward a shared cyberinfrastructure for sample collections	Data access, data discovery, data integration, data management
RCN6-SEN	Major outcomes from SEN will be creation of a Knowledge Base (SEN-KB), coordination of Experimental Collaboratories (SEN-EC) and integration of Educational efforts and Data standards development (SEN-ED) with tools for propagating new technology and methods. SEN-KB will be a collection of online resources for management and discovery of experimental data, metadata, analysis tools, methodologies and other user-driven needs. SEN-EC will pilot infrastructure to foster multi-laboratory collaborations on experiments addressing broad and interdisciplinary grand challenges that are difficult to solve in a single laboratory: extrapolation of experiments to natural systems and theory, comparability of experimental results from disparate facilities and decoupling of external versus intrinsic processes observed in experiments. SEN-ED will provide training for data management through workshops and outreach for collecting and sharing experimental data. The project will facilitate access and use of new and existing data for a wide range of users in the experimentalist community and beyond toward modelers and field geologists	The project provides guidelines for experimental design and data collection, i.e. overall workflow of sediment experiment. More communication in research groups and individuals about new updates on new experiments and technologies is enhanced through the project	Data access, data discovery, data management

Table AV.