EarthCube Governance Framework: A Proposal to the Community

Documentation, Research, and Recommendations

EarthCube Governance Working Group

OPEN-FILE REPORT OFR-12-39 V 1.0

September 2012

Arizona Geological Survey

www.azgs.az.gov | repository.azgs.az.gov

Arizona Geological Survey

M. Lee Allison, State Geologist and Director

Manuscript approved for publication in September 2012
Printed by the Arizona Geological Survey
All rights reserved

For an electronic copy of this publication: www.repository.azgs.az.gov
Printed copies are on sale at the Arizona Experience Store
416 W. Congress, Tucson, AZ 85701 (520.770.3500)

For information on the mission, objectives or geologic products of the Arizona Geological Survey visit www.azgs.az.gov.

This publication was prepared by an agency of the State of Arizona. The State of Arizona, or any agency thereof, or any of their employees, makes no warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed in this report. Any use of trade, product, or firm names in this publication is for descriptive purposes only and does not imply endorsement by the State of Arizona.

Recommended Citation: EarthCube Governance Working Group, 2012, EarthCube Governance Framework: A Proposal to the Community - Documentation, Research, and Recommendations. Arizona Geological Survey Open File Report, OFR-12-39 v 1.0, 199 p.

EarthCube Governance Framework:
A Proposal to the Community
Documentation, Research, and Recommendations

by

EarthCube Governance Working Group

Arizona Geological Survey Open-File Report 12-39 v 1.0

September 2012

2012

EarthCube Governance Framework: A Proposal to the Community

Documentation, Research, and Recommendations

September 2012, v. 1.0

EarthCube Governance Working Group
Presented to the National Science Foundation
9/4/2012

EARTHCUBE GOVERNANCE FRAMEWORK: A PROPOSAL TO THE COMMUNITY

EarthCube Governance Steering Committee
M. Lee Allison, Chair, Arizona Geological Survey
Tim Ahern, IRIS Data Management
David Arctur, University of Texas at Austin
Jim Bowring, College of Charleston
Gary Crane, Southeastern Universities Research Association
Cecelia DeLuca, NESII/CIRES/NOAA Earth System Research Laboratory
Geoffrey Fox, Indiana University
Carroll Hood, Raytheon
Hannes Leetaru, Illinois Geological Survey
Kerstin Lehnert, Columbia University
Chris MacDermaid, Colorado State University/CIRA
George Percivall, Open Geospatial Consortium
Mohan Ramamurthy, Unidata
Erin Robinson, Federation of Earth Science Information Partners
Ilya Zaslavsky, San Diego SuperComputer Center

Arizona Geological Survey Staff
Genevieve Pearthree
Kim Patten

Additional EarthCube Contributors
EarthCube Brokering Concept Team: Jay Pearlman
EarthCube Stakeholder Survey: Joel Cutcher-Gershenfeld

In addition, there were countless members of the EarthCube Governance Forum, EarthCube community, and other interested partners that contributed to this report.

PREFACE

The EarthCube Governance Steering Committee was tasked with creating an initial governance framework for EarthCube, based upon the Governance Roadmap, which we presented to the National Science Foundation (NSF) and the EarthCube community prior to the June 2012 NSF EarthCube charrette (community event). The purpose of the Governance Roadmap was to provide NSF and other interested parties with a cross-spectrum of ideas and concepts from the relevant Earth science stakeholder communities regarding key elements needed for an initial EarthCube governance framework.

This Governance Framework document presents initial steps to craft a governance framework for EarthCube, and is based on action items recommended in the Governance Roadmap.

It should be noted that this report is released as a draft, and continued evolution is expected to occur as we conduct community engagement activities over the next six months. A living version is available for comments:
https://docs.google.com/document/d/1UjOwPYBnI4uihlZsh57X1DK6MAknPa_wmlQK3UXMf5I/edit.

This material is based upon work supported in part by the National Science Foundation under Grants No. 1238951 and 0753154.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

OVERVIEW OF THIS DOCUMENT

Executive Summary
Provides an overview of our key findings, process, and our community building and engagement program.

Section 1.0: Developing the Initial EarthCube Governance Framework
Describes in detail the formation of the Governance Working Group, the writing of the Governance Roadmap, and the processes that led to the writing of this Governance Framework Document.

Section 2.0: Key Concepts Regarding EarthCube and Governance
Explores definitions of EarthCube and of governance, examines EarthCube's place within existing national and international cyberinfrastructures, and offers three emerging governance concepts for EarthCube brought forth by EarthCube community participants.

Section 3.0: Initial EarthCube Governance Framework
Presents our key findings including a list of initial governance functions for EarthCube, a list of guiding principles to guide the implementation of the governance functions, and a list of several general governance recommendations for EarthCube.

Section 4.0: Community Building and Engagement Program
Describes in detail the activities we will be conducting to engage EarthCube stakeholders and build the EarthCube community. This section includes a detailed Community Engagement Work Plan.

Section 5.0: Community Building and Engagement Work Plan
Includes a detailed Community Building and Engagement Work Plan divided into three main (and often iterative or simultaneous) steps.

Conclusion
Summarizes the processes, key findings, and the community building and engagement program presented in this document.

Appendices:
Includes community feedback gathered during and since the June 2012 EarthCube charrette, including summaries of the EarthCube Working Group and Concept Team roadmaps, a master list of EarthCube governance functions, and community input from the June charrette and additional meetings, and comments posted on the EarthCube Governance Ning site. An additional list of important documents and websites is also included.

Contents

EarthCube Governance Framework: A Proposal to the Community (Contributors) 1
Preface 2
Overview of This Document 3
Executive Summary 7

Introduction 7
Governance Framework Goals 8
Key Findings 9
Defining EarthCube and EarthCube's Scope 13
Defining Governance 15
Community Building and Engagement Program 17
Community Building and Engagement Work plan 18

Section 1.0: Developing the Initial Framework 19

1.1 Introduction 19
1.2 Forming the Governance Working Group 21
1.3 Writing the Governance Roadmap 21
1.4 Developing the Governance Framework 23

Section 2.0: Key Concepts Regarding EarthCube and Governance 25

2.1 Defining EarthCube 25
2.2 Defining Governance 27
2.3 Determining EarthCube's Scope 30
2.4 Emerging EarthCube Governance Concepts 35

2.4.1 Emerging Governance Concept: Cross-Domain Interoperability 35
2.4.2 Emerging Governance Concept: Three Tiers of Governance 38
2.4.3 Emerging Governance Concept: Outcomes of a Concept Award Principal Investigator Meeting 48

Section 3.0: Initial Governance Framework for EarthCube 51

3.1 EarthCube Governance Functions 51
3.2 EarthCube Governance Guiding Principles 54
3.3 EarthCube Governance Recommendations 55

Section 4.0: Community Building and Engagement Program 56

4.1 Generating Actionable Intelligence 57
4.2 Messaging 59
4.3 Engagement 60

4.3.1 Community-Specific Engagement 61
4.3.2 Community Engagement Methods 70

Section 5.0: Community Building and Engagement Work Plan 73

5.1 Generate Actionable Intelligence 73
5.2 Develop a Messaging Strategy 74
5.3 Engage EarthCube Stakeholders 76
5.4 Analysis and Synthesis 77

Conclusion 78
Appendix 1: June 2012 EarthCube Charrette Materials 83

EarthCube Governance Presentation: June 12, 2012 83

Audience Recommendations 83
Governance Roadmap Pluses 85
Governance Roadmap Challenges 86

Governance Question and Answer Session: June 13, 2012 88
Group Mixing: June 13, 2012 94

Semantics & Ontologies and Governance 94
Governance and Layered Architecture Concept Team 96

Appendix 2: EarthCube Working Group and Concept Team Roadmap Summaries 98

Data Discovery, Access and Mining Working Group 100
Semantics and Ontologies Working Group 103
Workflow Working Group 106
Brokering Concept Team 109
Cross-Domain Interoperability Concept Team 114
Earth System Models Concept Team 118
Layered Architecture Concept Team 122
Web Services Concept Team 126

Appendix 3: EarthCube Working Group and Concept Team Goals and Governance Needs 129

Introduction 129
Data Discovery, Access and Mining Working Group 130
Semantics and Ontologies Working Group 134
Workflow Working Group 136
Brokering Concept Team 140
Cross-Domain Interoperability Concept Team 143
Earth System Models Concept Team 153
Layered Architecture Concept Team 155
Web Services Concept Team 156

Appendix 4: EarthCube Governance Functions Master List 157

Governance Functions from the EarthCube Governance Roadmap 157
Governance functions From Other EarthCube Groups 159
Governance Functions from An EarthCube Concept Award Principal Investigator Meeting 163

Appendix 5: Community Input and Engagement 164

Workshops & Meetings Hosted 164

Community input To-Date: Workshops and Conferences 166
EarthCube Concept Award Principal Investigator Meeting 166
Boulder, CO, July 10, 2012 166
ESIP Federation Mid-Year Meeting, Madison, WI, July 17-20, 2012 171
Other EarthCube ESIP Workshops 179
Geosciences Australia seminar and roundtable 186
Canberra, Australia, July 30, 2012 186
34th international Geological Congress 186
Brisbane, Australia, August 5-10, 2012 186
Community Input to–Date: EarthCube Governance Ning Site 187
Is there a membership model for EarthCube? 187
Revised Classification of EarthCube Governance Functions - OPEN FOR EDITING! 189
EarthCube Governance Roadmap - Final Version 191
Additional Community Input 195
Appendix 6: Important Documents and WEbsites 196
Working Documents 196
Important Documents and Websites 196
References 198

EXECUTIVE SUMMARY

INTRODUCTION

EarthCube, established by the National Science Foundation (NSF), is envisioned to transform the conduct of research through the development of community-guided cyberinfrastructure for the Geosciences. EarthCube is intended to create a networked, knowledge management system that integrates Earth system and human dimensions data in an open, transparent, and inclusive manner. In order to achieve this vision, EarthCube requires broad community participation in concept, framework, and implementation, and must not be hindered by rigid preconceptions such as assuming one particular management or organizational structure. If successful, EarthCube may be a prototype for development of similar systems across other domain sciences as part of the NSF Cyberinfrastructure for the 21st Century (CIF21) initiative and related efforts nationally and internationally.

During the spring of 2012, NSF merged several governance-related Expressions of Interest into a single Governance Working Group, led by the Governance Steering Committee, with connections to a broad network forum of individuals and organizations known as the Governance Forum. The governance research review and community engagement efforts by this group culminated in a Governance Roadmap delivered to the NSF-sponsored June charrette (community event). The Governance Roadmap was the product of six months of research on governance theory and models, historical infrastructure case studies, and community feedback and engagement. Based on our findings from the research review and from community engagement, we the crafted following definition of governance:

> ***"Governance*** *refers to the processes, structure and organizational elements that determine, within an organization or system of organizations, how power is exercised, how stakeholders have their say, how decisions are made, and how decision makers are held accountable."*[1]

The Governance Roadmap also presented a plan with an aggressive timetable to define and implement a governance framework to enable the elements of EarthCube to become operational expeditiously. The governance framework we present in this document was produced by the initial implementation of action items 1 and 2 from the Governance Roadmap:

1. <u>August 31, 2012:</u> Determine the appropriate governance framework to meet community needs and NSF goals for successful cyberinfrastructure.
2. <u>Beginning of 2013:</u> Determine the stakeholder community and identify initial governance committee for engaging the community for input on the governance framework developed in Step 1.
3. <u>Date TBD:</u> Establish Terms of Reference for EarthCube Governance.

[1] Allison et al., "EarthCube Governance Roadmap: Documentation, Research, and Recommendations, Version 2.0," August, 2012, 5.

4. Date TBD: Implement the suggested Terms of Reference for EarthCube Governance.

GOVERNANCE FRAMEWORK GOALS

Our mission is to help ensure the timely and successful realization of EarthCube by developing and vetting a community-endorsed Governance Framework. This initial EarthCube Governance Framework is the product of eight months of community engagement activities with Earth scientists, industry, government agency and international representatives, in addition to extensive background research on governance models. The Framework, and corresponding community outreach, will maximize engagement of the broader EarthCube community, which in turn will minimize the risks that the community will not adopt EarthCube in its development and final states. The target community includes domain, information and computer scientists in academia, government, and the private-sector, and both nationally and internationally.

Our goal in developing an initial Governance framework is to provide the foundation that an eventual EarthCube governing entity(ies) will use to govern (i.e., manage and organize) EarthCube. Specifically, our aim is provide the foundation from which a governing entity can establish methods, processes, and procedures to carry out the governance functions presented in this document, according the governance guiding principles, also present in this document, with the goal of meeting the community needs and fulfilling NSF goals.

This governance framework defines a set of governance functions that need to be carried out, based on our research on governance and our community engagement to-date with EarthCube stakeholders and other interested parties. We then provide a set of guiding principles again from community discussions that should guide implementation of the governance functions, in addition to several recommendations regarding EarthCube governance in the longer term.

We chose not to recommend any specific governance model, or set of models, for EarthCube, because it is too early in the process and not necessarily appropriate to make this recommendation. Instead, we intend that the body or group responsible for carrying out the initial governance functions will have to define the governance model(s) they believe is(are) most effective way for carrying EarthCube forward. It will be up to the organizations to identify what governance models they feel would be most effective to carry out functions based on guiding principles, while also meeting the goals of the community and NSF.

This approach is consistent with how organizations have traditionally proposed running large programs to NSF. It leaves open the possibility for a variety of plausible governing models for EarthCube, several of which are described in the EarthCube Governance Roadmap.[2]

Therefore, the goal of this document is to provide the first steps in establishing EarthCube governance. This document serves as a summary of our initial Governance Framework

[2] See EarthCube Governance Roadmap, Section 5: Status.

recommendations which will need to be vetted by the community. As we engage the broader earth, computer, and information science communities, the governance functions and guiding principles will likely be modified. We plan to implement our community engagement plan throughout the second half of 2012, and present the results of our community engagement to NSF and the greater EarthCube community by early 2013. In this way, the body that implements Steps 3 and 4 of the Governance Roadmap may build on the community engagement program and Governance Framework that are initiated here.

KEY FINDINGS

Based on eight months of research we compiled and synthesized system-wide governance requirements to draft an initial set of EarthCube governance recommendations, governance functions and guiding principles. These functions, guiding principles and recommendations form the basis of an initial Governance Framework and will be revised and updated as part of an aggressive community outreach and engagement program. We plan to incorporate additional community input and release a modified version of the Governance Framework at the beginning of 2013. A summary of our initial findings is presented below.

INITIAL EARTHCUBE GOVERNANCE FUNCTIONS

We developed a master list of 143 functions (see Appendix 4: Governance Functions Master List) collected from community feedback and published research on governance. We then categorized and consolidated the 143 functions into four overarching categories:

1. Identify and Implement an EarthCube Vision, Mission, and Goals
2. Engage and Coordinate Across the EarthCube Community
3. Management: Create and Implement EarthCube By-Laws and Charter
4. Develop and Maintain a Viable Architecture and Concept of Operations that Enable the Realization of the Goals and Objectives of the EarthCube Vision

These functions identify and prioritize what we feel needs to be carried out by EarthCube governance at this time. They are non-prescriptive, however, thereby leaving room for any number of interested individuals, institutions, organizations and consortiums from both the public and private sector to propose to NSF and the EarthCube community how these functions might be carried out.

Each function has a policy, infrastructure and services component, or a combination of two, or all three components. Each of the components is ordered according to what we feel should be implemented first, on a scale of 1-3 (1 = near term or next 6 months, 2 = mid-term or next 6-24 months, and 3 = long-term or next 2-5 years). Each function is then assigned an overall prioritization, using the same 1-3 scale (Table 1).

This list is a living document, open for community editing and comment here: https://docs.google.com/spreadsheet/ccc?key=0AhE8IKrxOMWndDZfTXBUUFE1QmhOT01XT2JrZ3BQSWc#gid=0.

Initial EarthCube Governance Functions

	EarthCube Governance Functions	Overall Timing	Policy	Services	Infrastructure
	1. Identify, Implement and maintain an EarthCube vision, mission, and goals.	**1**	**1**	**2**	**2**
1	Lead the implementation of the EarthCube vision, mission and goals.	**1**	2	2	2
2	Develop metrics to evaluate progress towards EarthCube vision, mission and goals.	**1**	1	3	1
3	Establish processes to prioritize goals while building an open and transparent collaborative community for all EarthCube participants.	**2**	2	3	NA
	2. Engage and Coordinate Across the EarthCube Community	**1**	**1**	**1**	**1**
1	Establish processes to determine how EarthCube will collaborate with other organizations.	**1**	1	1	1
2	Promote education and outreach.	**1**	1	1	1
3	Establish processes to seek and prioritize pilot projects.	**2**	2	2	2
4	Establish processes to identify and engage potential EarthCube participants.	**1**	1	2	NA
5	Establish processes to recognize and respond to changing community needs.	**1**	1	1	1
6	Identify and manage EarthCube-wide services.	**2**	1	3	2
7	Identify and coordinate touchpoint services (services that link organizations, etc., to EarthCube).	**2**	1	3	2
8	Promote the integration of existing communities of practice into EarthCube, and foster the establishment of needed communities of practice, such as software development.	**2**	2	3	NA
9	Manage EarthCube-wide use cases.	**2**	2	2	2
	3. Management: Create and Implement by-laws and charter	**2**	**2**	**2**	**NA**
1	Establish decision-making processes.	**1**	1	NA	NA
2	Establish processes to define an EarthCube participation policy.	**1**	1	2	NA
3	Establish processes to identify roles and responsibilities, and determine how they may be filled.	**1**	1	NA	NA

4	Establish processes to ensure broad representation in EarthCube leadership at all levels.	**1**	2	3	NA
5	Establish processes to address EarthCube sustainability.	**2**	2	NA	3
6	Provide for the adaptive evolution of the governance framework.	**2**	1	NA	2
7	Establish processes to allocate resources and make recommendations to NSF	**2**	2	NA	NA
8	Establish processes to define and maintain support services.	**2**	1	2	2
9	Establish processes to ensure EarthCube governance is scalable.	**2**	2	NA	NA
10	Establish processes to resolve disputes, address legal issues, and manage risk.	**2**	2	3	NA
11	Establish processes that encourage best practices for sharing, curation, citation, credit, production, publication, and reuse of data models, software, and services.	**2**	2	3	2
4. Develop and maintain a viable architecture and concept of operations that enable the realization of goals and objectives of the EarthCube vision		**2**	**2**	**2**	**3**
1	Create processes to evaluate technologies, produce documentation, and recommendations for systems requirements	**2**	2	NA	2
2	Create processes to identify gaps in coverage of needed cyberinfrastructure capabilities, and determine recommendations on how to fill them	**2**	2	3	3
3	Establish processes for inventorying and assessing the current state of capabilities.	**2**	1	2	3
4	Identify optimal processes to fill cyberinfrastructure gaps, whether these processes are bottom-up, top-down, or a combination of both.	**3**	2	3	3
5	Establish processes to promote the adoption, adaptation, and only as a last resort, duplication of existing, or development of new, capabilities.	**3**	2	3	2

TABLE 1. INITIAL EARTHCUBE GOVERNANCE FUNCTIONS

EARTHCUBE GOVERNANCE GUIDING PRINCIPLES

These guiding principles were developed in order to guide the implementation of the EarthCube governance functions.

1. EarthCube governance shall serve the advancement of interdisciplinary science through collaboration among community members and with other cyberinfrastructure initiatives.

2. EarthCube governance shall rely on open, transparent processes and shall vet and inform its decisions through active community engagement.

3. EarthCube governance shall encourage environmentally sustainable processes and practices.

4. EarthCube governance shall support development that draws from best practices based on interoperability and reuse of resources.

5. EarthCube governance shall strive for the free and open sharing of data, information, software and services.

6. EarthCube governance shall evolve with changing technologies, practices and user needs while remaining robust.

This list is a living document, open for community editing and comment here: https://docs.google.com/document/d/1HMSRwpTTPqSXnrspPdDd7LSU6bAvzHa0JxGm-AZkbzU/edit.

GENERAL GOVERNANCE RECOMMENDATIONS

Based on the findings of the Governance Roadmap and the community engagement process since the June 2012 NSF EarthCube charrette, we developed a list of governance recommendations. Although these recommendations focus on the mid to long-term development of EarthCube, they should be taken into consideration during the next six months, and if necessary, improved upon as further feedback is provided. It is essential to consider the long-term implications and vision of EarthCube as we work to frame it in the present.

1. An interim prototype governance body should be established in the mid-term (starting early 2013) to carry out Steps 3 and 4 of the EarthCube Governance Roadmap (establish and implement Terms of Reference for EarthCube Governance), according to the findings presented in this Governance Framework document.

2. During the building of EarthCube, we recommend establishment of an umbrella entity to coordinate governance functions across EarthCube and facilitate the optimal interoperability of existing cyberinfrastructure into EarthCube and vice versa.

3. During the build phase of EarthCube, we recommend EarthCube governance to focus on the development, operation and maintenance of EarthCube-wide services and touchpoints (services that enable existing cyberinfrastructure components to interoperate effectively within EarthCube). The existing systems that are being interconnected would retain their local autonomy, continue to determine their own science priorities, and continue to serve their constituent communities. EarthCube would serve as an enabler of intra- and cross-disciplinary activity, identify gaps, and establish criteria for consistent with EarthCube infrastructure.

4. As part of the development of EarthCube-wide services and touch points, EarthCube governance would develop and maintain an effective enterprise-level architecture, proactively identify and manage enterprise-level risks, support solicitations for risk reduction prototypes, provide a community forum, and support solicitations for cross-domain testbeds.

5. As it matures, EarthCube will become responsible for discovering and encouraging the development of a set of cross-domain science priorities with the goal of creating new cross-disciplinary research communities within EarthCube and between EarthCube and external programs.

6. EarthCube will function principally as a community service provider, not a competitor to infrastructure produced by individuals, organizations, institutions or consortiums.

This list is a living document, open for community editing and comment here:
https://docs.google.com/document/d/1OqDMwRndYSt8Tx2noJbJWizeHDQ9ssNAdoFnILg_90k/edit

DEFINING EARTHCUBE AND EARTHCUBE'S SCOPE

Key questions related to all developments across the EarthCube community are: What is EarthCube and what is its scope? And, given these questions, what is the scope of EarthCube governance? While we do not present definitive answers to these questions, we do explore definitions of EarthCube and EarthCube's scope, focusing on EarthCube within national and international cyberinfrastructure, and we introduce several emerging governance concepts brought forth by EarthCube community members that informed this discussion.

What is EarthCube?
We recognize that NSF's definition of EarthCube has been subject to a wide variety of interpretations, including the ideas that EarthCube is:

1. An approach to respond to daunting science and cyberinfrastructure challenges
2. A knowledge management system

3. An architectural framework
4. A process
5. An outcome
6. An integrated system that builds on CI investments of the past two decades
7. A cyberinfrastructure
8. An integrated set of services
9. An architectural framework
10. Integrated set of solutions and best practices
11. Environment for adopting new approaches
12. Environment for human-centric interoperability of data and information
13. New modes of learning and training

Although EarthCube is still in the process of being defined, we believe it is possible to take concrete steps in moving EarthCube governance forward. In fact, the goal of the initial EarthCube Governance Framework presented in this document and the concurrent community engagement program is to further refine community requirements for EarthCube, thereby helping to give EarthCube more definition.

What is EarthCube's Scope?

We also recognize that EarthCube is yet a nascent activity within a rapidly growing, diverse, and wide-ranging global environment that is moving forward without, in many cases, even awareness of EarthCube. EarthCube is a significant but still modest prototype for the NSF Cyberinfrastructure 21st Century initiative (CIF21), which itself is but the contributions of one research agency in the federal government. Outside of EarthCube there are an untold number of people, programs, institutions, and organizations building and running systems and networks that make up the emerging cyberinfrastructure. These range from individual data sets and software applications to national and international cyberinfrastructure systems. Many of the contributors to cyberinfrastructure do not yet recognize their roles in the larger picture.

One of the debates throughout the development of this Governance Framework document focused on the role of infrastructure in EarthCube, and whether or not EarthCube would be responsible for building and maintaining infrastructure, or concentrate on coordinating existing and emergent infrastructure instead. Participants came to the conclusion that EarthCube would primarily facilitate, support, and coordinate, not compete with existing infrastructure, while leaving open the potential for building of new infrastructure if necessary (running the EarthCube website, for example, or other services the EarthCube community decides are better carried out centrally). In a facilitating role, EarthCube might for instance, identify key elements of CI that are critical for the initiative to advance, such as the Interoperability Institute proposed by the EarthCube Cross Domain Work Group. In that case, proponents of such an effort could prepare a funding proposal based on the criteria and standards determined as contributing to EarthCube success. An endorsement by the EarthCube community (or office) of the need for such infrastructure and its consistence with EarthCube principles and functions, could be valuable or even critical to a proposal being funded. Therefore it is important to be aware of and catalogue existing cyberinfrastructures.

Emerging Governance Concepts for EarthCube

Finally, we present three emerging governance concepts, brought forward by stakeholders in the growing EarthCube community, which influenced our thought process in creating this document, and are related to determining the scope of EarthCube. These concepts are not exhaustive, however, and as a wider spectrum of individuals and organizations are engaged in EarthCube, a plethora of potential governance concepts is likely to emerge. Although it will eventually be necessary to decide on a structure for EarthCube governance, we concluded that it is premature at this time to make any final decisions, given that only a sliver of the potential EarthCube community has been engaged thus far.

1. Cross-Domain Interoperability: This perspective, brought forth in the EarthCube Cross-Domain Interoperability Concept Team Roadmap, defines the scope of EarthCube as cross-domain integration of domain systems and research sites. Although these domain systems and research sites maintain local autonomy in the way they govern themselves, they are linked to EarthCube through cross-domain integration mechanisms, such as social networking, EarthCube-wide governance policies and agreements, brokering services, and management of standards and identifiers, among many others (see Figure 5).

2. Three Tiers of Governance: Enterprise-Level, Touchpoints, Domain-Specific: This perspective was brought forth by Carroll Hood of the EarthCube Governance Steering Committee and categorizes EarthCube governance into three tiers:
    a. *Enterprise-Level:* EarthCube-wide governance functions (that often exist outside of the EarthCube community as core elements of cyberinfrastructure).
    b. *Touchpoints:* Governance functions that connect domains to EarthCube and ensure they interoperate effectively, but that are locally optimized to each entity.
    c. *Domain-Specific:* Governance functions that are unique to a particular domain and are maintained and operated autonomously within that domain (see Figure 9).

3. EarthCube Concept Award Principal Investigator Reference Architecture: This perspective was introduced by the EarthCube Brokering Concept Team and later modified during a Concept Award Principal Investigator meeting. Current interactions within science domain infrastructures use established catalogs, semantics, data access mechanisms, information models and policies. These interactions are facilitated through the use of brokers, which then mediate between community resources and users (see Figure 14).

DEFINING GOVERNANCE

Based on our initial community engagement, we've realized there are a number of definitions regarding governance. Governance of EarthCube will have to take in account all of the interpretations of governance brought forward by EarthCube stakeholders, in particular as additional individuals are brought into the EarthCube community. In order to clarify a broader interpretation of governance, we present several definitions of governance encountered during the governance research review:

1. "Governance refers to the processes, structure and organizational elements that determine, within an organization or system of organizations, how power is exercised, how stakeholders have their say, how decisions are made, and how decision makers are held accountable."[3]
2. IT governance "specifies the decision rights and accountability framework to encourage desirable behavior in using IT. IT governance is not about making specific IT decisions–management does that–but rather determines who systematically makes and contributes to those decisions."[4]
3. Governance "provides the structure for determining organizational objectives and monitoring performance to ensure that objectives are attained."[5]
4. A "governance model describes the roles that project participants can take on and the process for decision making within the project. In addition, it describes the ground rules for participation in the project and the processes for communicating and sharing within the project team and community."[6]
5. Governance "aligns an organization's practices and procedures with its goals, purposes, and values. Definitions vary, but in general governance involves overseeing, steering, and articulating organizational norms and processes (as opposed to managerial activities such as detailed planning and allocation of effort). Styles of governance range from authoritarian to communalist to anarchical, each with advantages and drawbacks."[7]

We also present several governance models currently in use to gather input and make decisions within organizations, non-profits, business enterprises, and other groups. These include:

1. **Benevolent Dictatorship:** benevolent dictator
2. **Business monarchy**: top managers
3. **IT monarchy**: IT specialists
4. **Feudal**: each unit makes independent decisions
5. **Federal**: combination of corporate center and business units with or without IT people involved
6. **IT duopoly**: IT group and one other group (ex: top management or business unit leaders)
7. **Anarchy**: isolated individual or small group decision making
8. **Meritocracy**: distributed control awarded based on contributions to the project.

A further analysis of governance and governance models is presented in the Governance Roadmap or in Section 2 of this document.

---

[3] Allison et al., "EarthCube Governance Roadmap," 5.
[4] Weill and Ross, *IT Governance*, 2.
[5] Organization for Economic Cooperation and Development, *OECD Principles of Corporate Governance* quoted in Weill and Ross, *IT Governance*, 4-5
[6] Ross Gardler and Gabriel Hanganu, "Governance Models."
[7] EarthSystem Commodity Governance Project, "Governance."

COMMUNITY BUILDING AND ENGAGEMENT PROGRAM

We've developed an aggressive community engagement plan to vet the recommendations presented in this Governance Framework document with the broader scientific and IT communities, including: Atmospheric Scientists, Geoscientists, Ocean Scientists, Computer Scientists, Software Developers, Information and Communications Technologies Standards Communities, Business and Industry, Federal and State Governments and Decision-Makers, and International Partners. As we engage the broader cross-spectrum of EarthCube stakeholders, the Governance Framework is likely to change. We plan to implement our community engagement plan throughout the second half of 2012, and present the findings to NSF and the greater EarthCube community throughout the process with a report released in 2013. In this way, the body that implements Steps 3 and 4 of the Governance Roadmap may build on the community engagement program and Governance Framework that we recommend.

The overall community engagement strategy we plan to implement is conceptualized into three iterative, often organic phases: 1) Generating actionable intelligence, 2) Messaging, and 3) Engagement. Goals of this iterative community feedback process are to:

1. Vet the initial governance functions list and guiding principles presented in this document
2. Obtain website functionality and user requirements (user-centered design) for online collaboration
3. Engage individuals and organizations not yet involved in EarthCube
4. Maintain momentum gained from the June 2012 NSF charrette
5. Identify gaps and overlaps in the existing EarthCube community
6. Provide for community engagement to minimize risk and ensure that EarthCube will be of use to the broader geoscience community.

Engagement activities will be focused on the broad spectrum of EarthCube participants including geo, atmosphere, ocean, computer and information sciences communities, software developers, standards bodies, IT experts, private industry, government agencies, and the international community.

COMMUNITY BUILDING AND ENGAGEMENT WORK PLAN

Our community engagement program will be an iterative process. Many activities will be taking place in concurrently. Each element will be incorporated into this Framework document as feedback is received. The complete work plan is presented in Section 5.0 of this document.

PART 1: GENERATE ACTIONABLE INTELLIGENCE

1. Leverage the EarthCube Stakeholder Alignment Survey to identify gaps in the current EarthCube stakeholders, if any.
2. Leverage the EarthCube Stakeholder Alignment Survey to gather community requirements on EarthCube governance
3. Create an EarthCube network map of EarthCube stakeholders
4. Develop EarthCube stakeholder profiles

PART 2: DEVELOP A MESSAGING STRATEGY

1. Determine EarthCube messaging strategies tuned to different stakeholder groups identified in Part 1.
2. Create marketing material to explain the EarthCube vision and introduce governance
3. Work with EarthCube Working Groups and Concept Teams to create short introductions to their groups, to be included in EarthCube marketing materials
4. Use online forums to market EarthCube
5. Identify community champions
6. Publish articles in scholarly journals and trade publications

PART 3: ENGAGE EARTHCUBE STAKEHOLDERS

1. Engage EarthCube stakeholders via social media
2. Engage EarthCube stakeholders via virtual meetings and workshops
3. Engage via in-person meetings and workshops
4. Gather EarthCube website requirements.

PART 4: ANALYSIS AND SYNTHESIS

1. Analyze and synthesize all input gathered from community engagement activities.
2. Incorporate this information into an updated Governance Framework document, to be released in early 2013.

SECTION 1.0: DEVELOPING THE INITIAL FRAMEWORK

1.1 INTRODUCTION

Introducing EarthCube

EarthCube, launched by the National Science Foundation (NSF) in July 2011, is envisioned to transform the conduct of research through the development of community-guided cyberinfrastructure for the Geosciences, in effect, a prototype for the NSF Cyberinfrastructure for the 21st Century (CIF21). EarthCube is both a process and an outcome, aimed to create a networked, knowledge management system of systems and infrastructure that integrates all Earth system and human dimensions data in an open, transparent, and inclusive manner. In order to achieve this vision, EarthCube requires broad community participation in concept, framework, and implementation and must not be hindered by rigid preconceptions. If successful, EarthCube may be a prototype for deployment across other domain sciences. EarthCube was highlighted in the White House Office of Science and Technology Policy rollout of the "Big Data" initiative.

During the spring of 2012, NSF merged several governance-related EarthCube Expressions of Interest into a single Governance Working Group, led by the Governance Steering Committee, with connections to a network forum of individuals and organizations known as the Governance Forum. The work of this group culminated in a Governance Roadmap delivered to the NSF-sponsored June charrette with an aggressive timetable to define and implement a governance structure to enable the elements of EarthCube to become operational expeditiously. This governance framework document represents the implementation of initial recommendations laid out in the Governance Roadmap.

Defining Governance

Based on our initial community engagement, we've realized there are a number of definitions regarding governance. Even among the 250+ June EarthCube charrette virtual and in-person participants, there were several different interpretations of governance and expectations of what the Governance Roadmap would propose. For example, some participants assumed governance referred to consensus-based decision-making processes and committees, while others thought of top-down management strategies, while others interpreted governance to refer to standards.

Additionally, each of the Working Groups and Concept Teams focused on different components of governance, as they relate to their specific topic of interest. For example, the Data Access, Discovery and Management Roadmap focuses on governance related to data, while the Workflows Roadmap focuses on workflow-specific governance. While all of these interpretations are correct, governance in general is much more comprehensive, and governance of EarthCube will have to take in account all of these interpretations, and many more, as additional individuals are brought into

the EarthCube community. For a more comprehensive review of governance, please refer Section 2.2 of this document, or to the Governance Roadmap.[8]

Writing the Governance Roadmap

A key question the Governance Working Group struggled with while writing the Governance Roadmap is whether something as complex as EarthCube can be governed by one single model or by one encompassing entity? Six months of research on historical infrastructure case studies, governance theory and models, in addition to community feedback and engagement prior to writing the Roadmap, provided a venue to begin to tackle this challenging and complex question.

Governance Roadmap, we learned that many large-scale, complex infrastructures, including the Internet, have no single, overarching, central control, administration, or management. No national infrastructure that we examined is governed by a single entity, let alone a single governance archetype. Although the Governance Roadmap does not provide answer to this debate, it did provide a set of action items to forge a viable path forward in crafting an initial governance framework for EarthCube.

In addition, the National Science Foundation has not historically defined a particular governing model in their solicitations. Instead, they list the functions and goals that need to be met and invite proposers to outline the organizational and management structures and procedures ("governance") best suited to achieve success. The effectiveness of those criteria is a key element in proposal evaluation.

Governance Framework Goals

The goal of writing and vetting the governance functions, guiding principles and recommendations (referred to as the 'Governance Framework' of 'Framework' in this document) across a broad spectrum of EarthCube stakeholders is to help ensure the timely and successful conceptualization and operation of EarthCube. The Framework, and corresponding community outreach, will maximize engagement of the broader EarthCube community, which in turn will minimize the risks that the community will not adopt EarthCube in its development and final states. The target community includes academia, government, and the private-sector, both nationally and internationally.

Based on community feedback to-date, we compiled and synthesized system-wide governance requirements to draft an initial set of EarthCube governance recommendations, governance functions and guiding principles. These recommendations, functions and guiding principles form the basis of an initial Governance Framework that will be revised and updated as part of an aggressive community outreach and engagement program. This is a living document and we plan to update the Governance Framework throughout this process.

[8] See Section 5: Status, and Appendix 1: Governance Research Review.

1.2 FORMING THE GOVERNANCE WORKING GROUP

The EarthCube Governance Working Group was formed (by merging proposes of governance-related EarthCube Expressions of Interest) in March 2012 with the purpose of engaging the geosciences community and conducting the background research needed to provide the scaffolding for a roadmap on how to implement an initial governance framework for EarthCube. The Group is composed of the Governance Steering Committee and a larger Governance Forum. Members of the Steering Committee consist of active members from the geo, atmosphere, ocean, computer and information sciences community, in addition to industry, international, and government agency partners.

The Governance Forum is composed of individuals and organizations forming a larger, broader constituency that had demonstrated their interest and participation in governance issues, and a representative from each of the EarthCube Community Groups and Concept Award Teams. The latter served as liaisons between the more technically focused groups and the Governance Working Group to help ensure unique governance issues and needs from each segment of EarthCube were identified and addressed. One of the goals of organizing the Forum was to ensure representation from each of the Geoscience and closely related domains (e.g. Earth, Atmospheres, Ocean, Environment and Ecology), IT sectors, academia, agencies, and industry, involved in determining agendas, setting goals, and having an active role in formulating the roadmap.[9]

A series of virtual meetings and workshops were hosted by the Governance Steering Committee with different working groups and stakeholders, covering a range of topics, with the aim of gathering governance feedback. These workshops began to tackle the complexity, scope, and diversity of issues regarding governance of EarthCube, and were exceptional in providing the initial community feedback and basis required to author the Governance Roadmap. These workshops, in addition to a substantial background research review on governance, formed the foundation of the Governance Roadmap.

1.3 WRITING THE GOVERNANCE ROADMAP

As we were writing the Governance Roadmap, which was released in June 2012, we noted an ongoing discussion among EarthCube participants regarding whether EarthCube governance is that of a single integrated entity, or a system of governance of separate, but collective entities. Regardless, the Roadmap argues that EarthCube governance is a process that is responsible for developing infrastructure that becomes ubiquitous and effectively invisible to users by building on existing and emerging cyberinfrastructure components. Thus, research on past infrastructure development was a key component of the Roadmap, because past infrastructure development offers several insights and case studies into analogies of where EarthCube has been, where it is going, and what governance can do immediately to move the process along.

[9] For a complete list of EarthCube Governance Forum Members, see Appendix 3: Process (Expanded) of "EarthCube Governance Roadmap: Documentation, Research, and Recommendations, Version 2.0," August, 2012

The Roadmap also explored the various governance models, or archetypes, varying from centralized ("benevolent dictatorship"), to more decentralized (federalist system), to most decentralized ("crowd-sourcing" or "anarchy"), each with its own pros and cons depending on meeting different governing functions. The Roadmap then employed case studies to evaluate the governance framework of existing domain science, IT, federated, and large-scale organizations.

The Roadmap did not recommend any particular governance model or framework for EarthCube, but it did outline four action items with target dates to carry out the community engagement and requirements-gathering processes necessary to implement an initial governance framework for EarthCube:

1. August 31, 2012: Determine the appropriate Governance Framework to meet community needs and NSF goals for successful cyberinfrastructure.
2. Beginning of 2013: Determine the stakeholder community and identify initial governance committee for engaging the community for input on the governance framework developed in Step 1.
3. Date TBD: Establish Terms of Reference for EarthCube Governance.
4. Date TBD: Implement the suggested Terms of Reference for EarthCube Governance.

Discussion of the Working Group and Concept Team roadmaps at the June 2012 NSF EarthCube charrette demonstrated that within the EarthCube community, there are many differences in understanding about what governance is, and its potential role in EarthCube. While the Governance Group came to the charrette asking what other groups needed in terms of governance, many of the other groups assumed a governance framework, or even a specific governance model, had already been chosen. Additionally, most groups envisioned governance to primarily consist of committees and consensus-based decision-making processes.

Community input and background research indicate governance is much more comprehensive, highlighting the need for community-wide dialogue about what governance is and what it may mean for EarthCube. These differences in understanding of governance make clear the importance of conducting comprehensive stakeholder engagement to validate and enhance the recommendations we make.

Additionally, numerous interactions with EarthCube Working Groups and Concept Teams indicate they are waiting on guidance from an enterprise-level governance body. When questioned about what governance means and what requirements are needed from such a body, however, participants are largely unsure. This indicates that additional outreach, discussion, and education to the key stakeholder communities of EarthCube are necessary, while concurrently engaging the long-tail and domain scientists to further gauge the community needs and wants regarding EarthCube governance.

1.4 DEVELOPING THE GOVERNANCE FRAMEWORK

The Governance Roadmap anticipated that a successor group to the ad hoc Governance Steering Committee would be empowered by the EarthCube community and NSF to continue on after the charrette. That did not explicitly occur; rather there was a broadly accepted presumption that the existing group had achieved its goals and had the momentum, willingness, and vision to take the plan forward. Therefore, the Governance Steering Committee is moving forward with implementing the Governance Roadmap. In an effort to be inclusive, we have recently identified, and confirmed, additional Steering Committee members who encompass key constituents in EarthCube: industry, government, and academia.

We are in the process of carrying out Tasks 1 and 2 of the EarthCube Governance Roadmap: 1) Determine the appropriate Governance Framework to meet community needs and NSF goals for successful cyberinfrastructure; and 2) Determine the stakeholder community and identify initial governance committee for engaging the community for input on the governance framework developed in Step 1. Upon closer consideration, and based on community feedback, we believe the Terms of Reference and implementation of EarthCube Governance (Tasks 3 and 4) should be performed over an extended period of time to lower the risk of acting prematurely, or without adequate community commitment. Tasks 3 and 4 will likely be implemented by a NSF initial governance award, if NSF chooses to make such an award in the future.

An initial draft of the Governance Roadmap[10] with these four recommendations was posted to the EarthCube Governance Ning site for public comment and review prior to the June EarthCube charrette Although we polished and finalized the Governance Roadmap by releasing versions 1.1[11] and 2.0,[12], instead of making substantive changes to the Roadmap, we created additional documents to provide the foundation for our Governance Framework recommendations.

We started by reviewing the draft roadmaps produced by the other EarthCube Working Groups and Concept Teams, with a focus on their goals, governance needs and wants, and their proposed internal governance framework (governance within their specific area of interest). Some of these governance wants and needs were stated explicitly; others were implicit. For more information about the other EarthCube Roadmaps, please see Appendix 2: EarthCube Working Group and Concept Team Roadmap Summaries, and Appendix 3: EarthCube Working Group and Concept Team Goals and Governance Needs.

Our goal was to gather governance use cases to evaluate potential governance models for EarthCube (see "Defining EarthCube below for more information on governance models). Upon

[10] EarthCube Governance Roadmap Version 1.0 is accessible online at:
http://earthcube.ning.com/group/governance/forum/topics/roadmap-draft-june-7th
[11] EarthCube Governance Roadmap Version 1.1 is accessible online at:
http://earthcube.ning.com/group/governance/forum/topics/earthcube-governance-roadmap-version-1-1
[12] EarthCube Governance Roadmap Version 2.0 is accessible online at:
http://earthcube.ning.com/group/governance/forum/topics/earthcube-governance-roadmap-final-version

attempting to map different governance models to governance use cases during the July 17th EarthCube Governance Workshop at ESIP, we realized this approach would not give us the answers we needed in order to develop an initial governance framework for EarthCube. We thought that the governance use cases would help us identify the most appropriate governance model, or series of governance models, to carry out all the different components that will likely fall under EarthCube governance.

We came to the conclusion, however, that recommending an initial governance model, or series of models, for EarthCube was too prescriptive at this point because EarthCube and EarthCube stakeholders are still being defined. Without a definition of what EarthCube will be, and without a broad picture of what stakeholders need it to do, it makes no sense to recommend a specific governance model.

In addition, we recognized that major projects funded by NSF typically have defined their management, organization, and procedures in their funding proposals, rather than NSF mandating them in the funding solicitations. Solicitations usually outline the functions and goals required and invite proposals to offer a governing model best suited to achieve the desired results.

Therefore we decided to take a different approach in developing a governance framework for EarthCube, focused on identifying the initial functions EarthCube governance needs to carry out, along with principles to guide the implementation of these functions, and a series of general governance recommendations for EarthCube.

First, however, we needed to consider several important questions: 1)What is EarthCube? 2) What is governance? And 3)What is EarthCube's potential scope?

SECTION 2.0: KEY CONCEPTS REGARDING EARTHCUBE AND GOVERNANCE

Much discussion has taken place regarding the definition of EarthCube itself, the definition of governance, and the scope of EarthCube. These are key questions that must be addressed throughout the EarthCube development process. This section presents various definitions of EarthCube and governance, explores EarthCube's place in national and international cyberinfrastructure, and offers three emerging concepts of EarthCube governance, brought forth by participants in the emerging EarthCube community.

2.1 DEFINING EARTHCUBE

NSF provided a definition of EarthCube very early into the process. However, as the discussions prior to and since the June charrette demonstrated, NSF's definition of EarthCube has been subject to a wide variety of interpretations. For example if one were to ask ten different people about their vision for EarthCube, there would likely be at least 15 different answers. A dilemma, then is how can we determine the optimum Governance framework for EarthCube when we are really not sure what the framework is supposed to be governing?

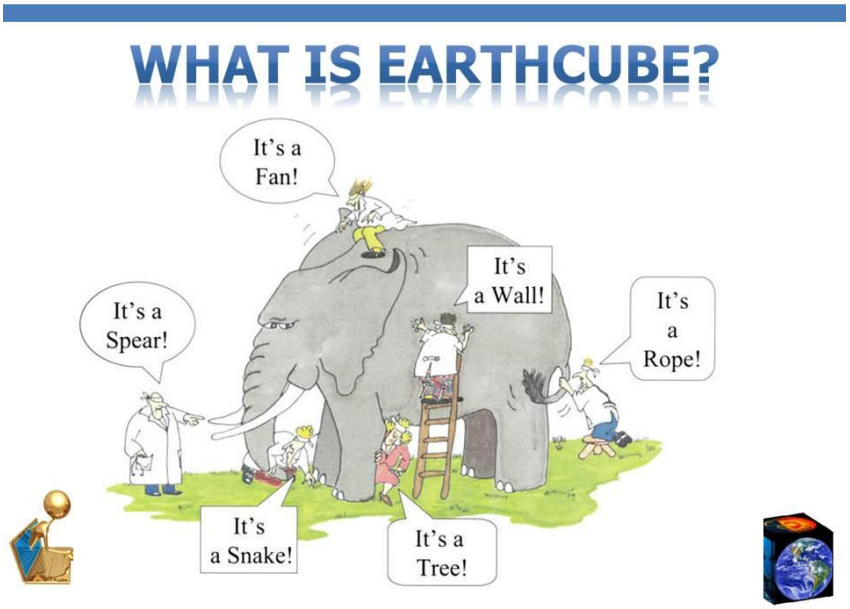

FIGURE 1. EARTHCUBE HAS MANY DIFFERENT DEFINITIONS, DEPENDING ON WHO YOU TALK TO.

Part of the reason for the plethora of different EarthCube definitions, is NSF's original EarthCube definition allows for many unique interpretations:

> "The goal of EarthCube is to create a knowledge management system and infrastructure that integrates all geosciences data in an open, transparent and inclusive manner. No integrated framework currently exists to meet the information management challenges implicit in the GEO Vision call to action. EarthCube is this goal, as well as a process that will require years of development, investment, and community engagement. The decade-long vision for EarthCube is the convergence towards an integrated system to access, analyze and share information that is used by the entire geosciences community. This convergence can only occur through community dialogue and collaboration. NSF will facilitate this convergence through currently planned and future activities.

> Realizing EarthCube requires building a community derived and governed cyberinfrastructure that can be easily adopted by geosciences researchers and educators. It will require the introduction of new approaches and technologies and/or combining productive tools and solutions in different ways. EarthCube will promote integration, flexibility, inclusiveness, and easy adoption by connecting the several layers of data and information management, from the resource layer with access to data and information, to the data curation and management layer, and to the overarching interface layer that exposes data and information to knowledge creation through data-enabled science. Inter-workability of data and information will be a characteristic of EarthCube, which implies interoperability and sharing, but with a human-centric focus that allows researchers and educators to easily work with multiply layers of technology and information. EarthCube will create new modes of learning and training that result in a more informed public and policy-makers while simultaneously broadening participation in the creation of a sustainable Earth system.
>
> EarthCube will be supported by the substantial cyberinfrastructure investments, including databases, software services and community facilities that have been created by the geosciences community over the past two decades. The community must build on this existing foundation of cyberinfrastructure and community knowledge to create an integrated set of services serving the entire geosciences community.
>
> The success of EarthCube will depend in part on identification of commonalities of solutions and best practices that reside within our current infrastructure and strategic adoption of new technologies and approaches external to the community's infrastructure. All geosciences cyberinfrastructure will evolve over time to accommodate changing user needs and emerging technologies and services. Results from the community-driven EarthCube effort will provide an architectural framework to guide the integration of efforts as well as the evolution of existing cyberinfrastructure."[13]

Based on this definition, one can conclude that EarthCube is

1. An approach to respond to daunting science and cyberinfrastructure challenges
2. A knowledge management system
3. An architectural framework
4. A process
5. An outcome
6. An integrated system that builds on CI investments of the past two decades
7. A cyberinfrastructure
8. An integrated set of services
9. An architectural framework
10. Integrated set of solutions and best practices
11. Environment for adopting new approaches
12. Environment for human-centric interoperability of data and information
13. New modes of learning and training

[13] National Science Foundation, "EarthCube—Building the Cyberinfrastructure to Better Understand Our Complex and Changing Planet," 2011.

Clearly, these concepts are all interrelated, but it is easy to see why so many interpretations of the EarthCube vision exist simultaneously. One could argue that EarthCube incorporates all ten of the definitions above. Although EarthCube is still in the process of being defined, it is still possible to take concrete steps in moving EarthCube governance forward. In fact, the goal of the initial EarthCube Governance Framework presented in this document and the concurrent community engagement program is to further refine community requirements for EarthCube, thereby helping to give EarthCube slightly more definition.

2.2 DEFINING GOVERNANCE

Based on our initial community engagement, we've realized there are a number of definitions regarding governance. For example, among the 250+ June EarthCube charrette virtual and in-person participants, there were several different interpretations of governance. Some participants assumed governance referred to consensus-based decision-making processes and committees, while others thought of top-down management strategies, while others interpreted governance to refer to standards. Additionally, each of the Working Groups and Concept Teams focused on different components of governance, as they relate to their specific topic of interest. For example, the Data Access, Discovery and Management Roadmap focused on governance related to data, while the Workflows Roadmap focused on workflow-specific governance.

While all of these interpretations are correct, governance as we address it, is much more comprehensive, and governance of EarthCube will have to take in account all of these interpretations, and many more, as additional individuals are brought into the EarthCube community. In order to clarify a broader interpretation of governance, we present the following definitions:

1. "Governance refers to the processes, structure and organizational elements that determine, within an organization or system of organizations, how power is exercised, how stakeholders have their say, how decisions are made, and how decision makers are held accountable."[14]
2. IT governance "specifies the decision rights and accountability framework to encourage desirable behavior in using IT. IT governance is not about making specific IT decisions–management does that–but rather determines who systematically makes and contributes to those decisions."[15]
3. Governance "provides the structure for determining organizational objectives and monitoring performance to ensure that objectives are attained."[16]
4. A "governance model describes the roles that project participants can take on and the process for decision making within the project. In addition, it describes the ground rules for

[14] Allison et al., "EarthCube Governance Roadmap," 5.
[15] Weill and Ross, *IT Governance*, 2.
[16] Organization for Economic Cooperation and Development, *OECD Principles of Corporate Governance* quoted in Weill and Ross, *IT Governance*, 4-5

participation in the project and the processes for communicating and sharing within the project team and community."[17]

5. Governance "aligns an organization's practices and procedures with its goals, purposes, and values. Definitions vary, but in general governance involves overseeing, steering, and articulating organizational norms and processes (as opposed to managerial activities such as detailed planning and allocation of effort). Styles of governance range from authoritarian to communalist to anarchical, each with advantages and drawbacks."[18]

The styles of governance mentioned in governance definition # 5 refer to the various governance models that determine who makes decisions and who has input within an organization. For more complex organizations, more than one governance model may be used to involve different people in making different types of decisions. For example, within any given organization or enterprise, IT-related decisions might be made in a different way and by different people than decisions regarding strategic direction. The most appropriate governance model, or series of model, for any given organization depends on that organization's overall goals. The governance models below describe different means of making decisions (for example, in the Benevolent Dictatorship model, a lone figure makes decisions much in the same way as a benevolent dictator would do within a dictatorship):

1. **Benevolent Dictatorship:** benevolent dictator
2. **Business monarchy**: top managers
3. **IT monarchy**: IT specialists
4. **Feudal**: each unit makes independent decisions
5. **Federal**: combination of corporate center and business units with or without IT people involved
6. **IT duopoly**: IT group and one other group (ex: top management or business unit leaders)
7. **Anarchy**: isolated individual or small group decision making
8. **Meritocracy**: distributed control awarded based on contributions to the project.

The governance models mentioned above determine what decisions need to be made, who makes them, and how they will be made and monitored.[19] The Governance Ontology from the Earth System Commodity Governance Project breaks down these types of decisions:[20]

1. How do you manage what you do? Do you have any management bodies (e.g., committees, director, etc.)?
    a. Strategic level
        i. What is their purpose?

[17] Ross Gardler and Gabriel Hanganu, "Governance Models."
[18] EarthSystem Commodity Governance Project, "Governance."
[19] Weill and Ross, *IT Governance*, 2-3.
[20] These questions are adapted from Earth System Commodity Governance, http://earthsystemcog.org/projects/cog/governance_object.

            1. Resource allocation, strategic direction, outreach, priorities, fund raising, etc.?
        ii. How many participants?
        iii. How often do they meet?
    b. Operational level
        i. What is their purpose?
            1. Design, review, task prioritization, testing, outreach, etc.?
        ii. How many participants?
        iii. How often do they meet?
2. Leadership roles
    a. Strategic level
        i. Name and description of roles
        ii. How do you decide how leadership roles are filled?
    b. Operational level
        i. Name and description of roles
        ii. Do you decide how leadership roles are filled?
3. Governance Processes
    a. How do you prioritize tasks?
        i. Who is engaged in prioritization?
        ii. How frequently does it happen?
    b. How do you communicate?
        i. What type?
            1. Telecom, mailing list, wikis, etc.
        ii. Purpose?
        iii. Frequency?
        iv. Is it open?
        v. Mechanisms?
    c. What process do you use to identify requirements?
    d. Dispute adjudication
        i. How do you resolve disputes?
    e. Policies?
        i. Links to policies
    f. Do you have a charter/by-laws/charter?
        i. If so, can it be attached?
4. Are there other aspects of management and organization that should be included that need to be considered for a governance model?

These governance models are explored in great detail in the Governance Roadmap.[21] While it is important to be familiar with the variety of governance models available to make decisions within EarthCube, determining what governance models are most appropriate for EarthCube is not the purpose of this Framework document.

[21] See Section 5: Status, and Appendix 1: Governance Research Review.

Instead, we chose not to recommend any specific governance model, or set of models, for EarthCube, because it is too early in the process and not necessarily appropriate to make this recommendation. Instead, whichever body or group responsible for carrying out the initial governance functions will have to define the governance model(s) they believe is (are) most effective way for carrying EarthCube forward. It will be up to the organizations to identify what governance models they feel would be most effective to carry out functions based on guiding principles, while also meeting the goals of the community and NSF. This approach is consistent with how organizations have traditionally proposed running large organizations to NSF. It leaves open the possibility for a variety of plausible governing models for EarthCube, such as those described above.

2.3 DETERMINING EARTHCUBE'S SCOPE

EarthCube is just one piece within global cyberinfrastructure. It is a significant but still modest prototype for the NSF Cyberinfrastructure for the 21st Century initiative (CIF21) which itself is but the contributions of one research agency in the federal government. Outside of EarthCube there are an untold number of people, programs, institutions, and organizations building and running things that make up the emerging cyberinfrastructure. These range from individual data sets and software applications to national and international cyberinfrastructure systems. Many of the contributors to cyberinfrastructure do not yet recognize their roles in the larger picture. Although the goal of EarthCube, as stated by NSF, is to "transform the conduct of research by supporting the development of community-guided cyberinfrastructure to integrate data and information for knowledge management across the Geosciences,"[22] we recognize that EarthCube is yet a nascent activity within a rapidly growing, diverse, and wide-ranging global environment that is moving forward without in many cases even awareness of EarthCube.

One of the debates throughout the development of this Governance Framework document focused on the role of infrastructure in EarthCube, and whether or not EarthCube would be responsible for building and maintaining infrastructure, or concentrate on coordinating existing and emergent infrastructure instead. Participants came to the conclusion that EarthCube would primarily facilitate, support, and coordinate, not compete with existing infrastructure, while leaving open the potential for building of new infrastructure if necessary (running the EarthCube website, for example, or other services the EarthCube community decides are better carried out centrally). In a facilitating role, EarthCube might for instance, identify key elements of CI that are critical for the initiative to advance, such as the Interoperability Institute proposed by the EarthCube Cross Domain Work Group. In that case, proponents of such an effort could prepare a funding proposal based on the criteria and standards determined as contributing to EarthCube success. An endorsement by the EarthCube community (or office) of the need for such infrastructure and it's compliance with EarthCube principles and functions, could be valuable or even critical to a

[22] National Science Foundation, EarthCube, http://earthcube.ning.com

proposal being funded. Therefore it is important to be aware of and catalogue existing cyberinfrastructures.

National Cyberinfrastructure

EarthCube is one of many cyberinfrastructure initiatives across the United States, spearheaded by Federal, state and local governments, private industry, and academia. Even within the Federal government, EarthCube is among a handful of cyberinfrastructure initiatives. In addition to NSF-sponsored cyberinfrastructure initiatives, such as DataONE, DataNet Federation, iPlant Collaborative, TeraGrid, and major research programs with significant cyberinfrastructure components such as Ocean Observing Initiative (OOI) and the National Ecological Observatory Network (NEON), the Executive Branch of the Federal Government has also launched a number of cyberinfrastructure initiatives within the past few years, including the Digital Government Initiative[23], the National Geospatial Platform Initiative,[24] and Data.gov.[25]

The U.S. Geological Survey launched the agency-wide Community on Data Integration (CDI) initiative in 2009 to promulgate a community of practice in cyberinfrastructure, that has now been incorporated in the 10-year strategic plan for core science systems.[8] One of the primary goals "enhances and develops new strengths in computer and information science to make it easier for USGS scientists to discover data and models, share and publish results, and discover connections between scientific information and knowledge."[26] The USGS Digital Earth concept envisions a robust internal cyberinfrastructure that meet that agency's needs, but draws on and contributes to a larger global system (Figure 2).

[23] See www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html
[24] See www.geoplatform.gov/home
[25] See www.data.gov
[8] U.S. Geological Survey, "Science Strategy for Core Science Systems in the U.S. Geological Survey, 2013–2023," Public Review Release, USGS Open-File Report 2012-1093, 2012.
[26] Ibid.

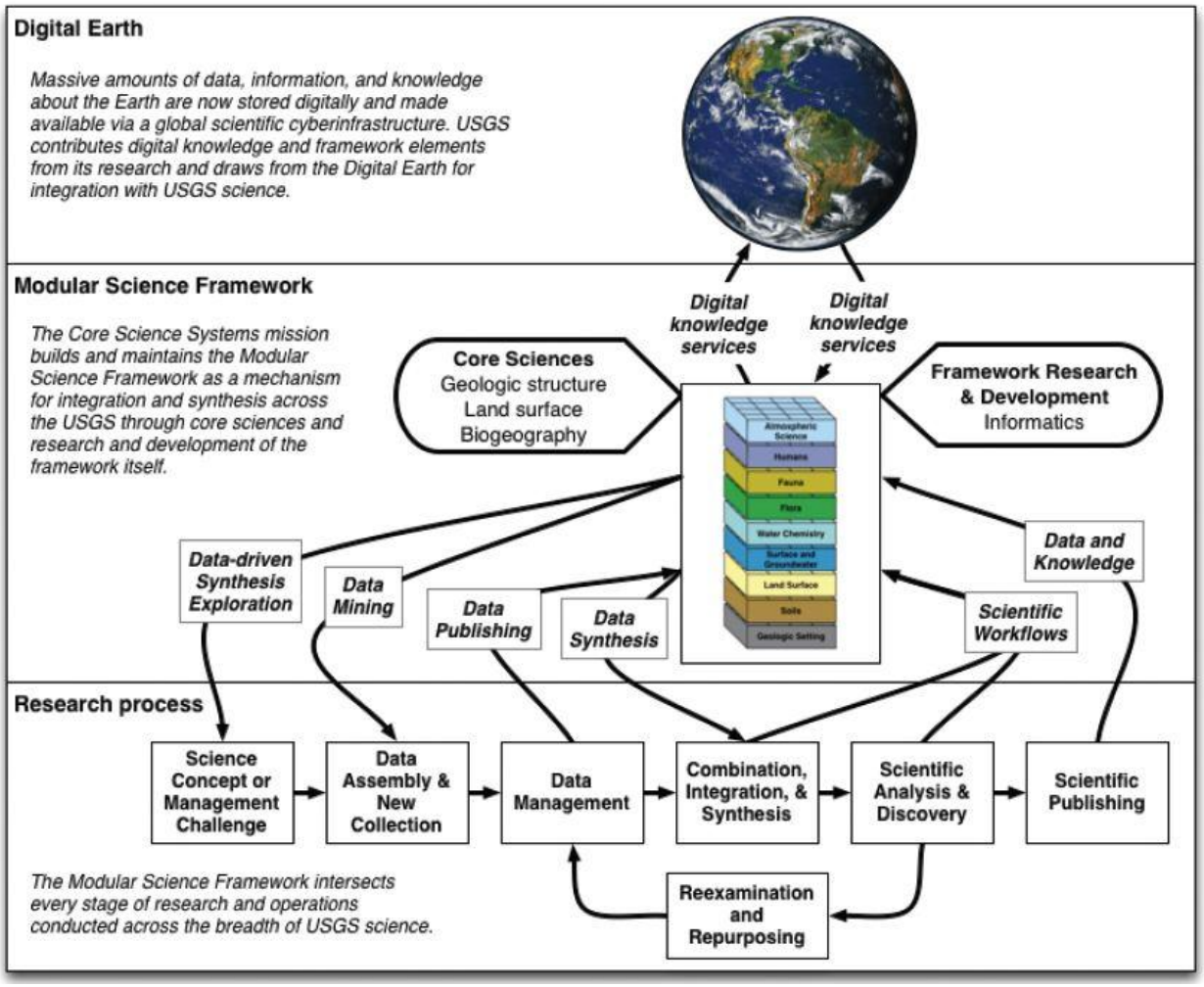

FIGURE 2. USGS DIGITAL EARTH

The Digital Earth concept illustrates the large scientific and worldwide community to which USGS contributes digital knowledge and from which USGS extracts data, information, and knowledge resources developed in the broader science community.[27]

[27] U.S. Geological Survey, "Science Strategy for Core Science Systems in the U.S. Geological Survey."

As Figure 3 demonstrates below, EarthCube is also just one component within NSF, which received only a small portion of the total congressional budget allocated to scientific research in 2009. NSF Federal funding for scientific research is mostly allocated to defense and health and human services, followed by NASA, and then NSF (Figure 3).

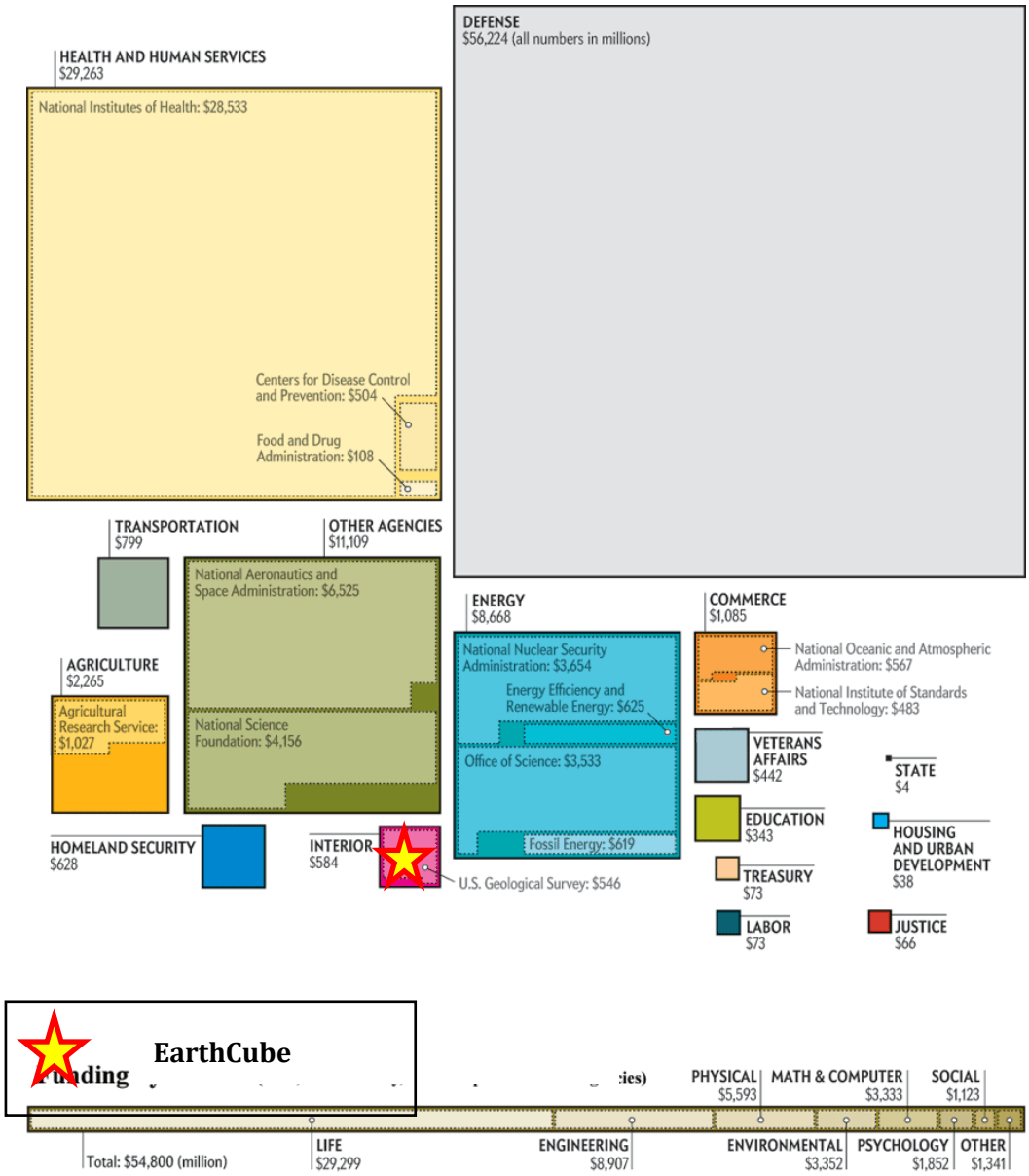

FIGURE 3. FUNDING SCIENCE IN THE 2009 U.S. FEDERAL BUDGET. [28]

[28] Fischetti, Mark, "Money for Science: U.S. Funding over the Years: Federal R & D spending shows how government priorities stack up," Scientific American, January 2011, http://www.scientificamerican.com/article.cfm?id=money-for-science.

International Cyberinfrastructure

In addition to national cyberinfrastructure, there are a multitude of cyberinfrastructure systems outside of the United States (Figure 4). Some of these initiatives are region or country-specific, such as the Infrastructure for Spatial Information in the European Community (INSPIRE) and the Australian National Data Service (ANDS) and AuScope (Australian natural resources cyberinfrastructure). Other cyberinfrastructure initiatives are global, such as OneGeology, which has 117 participating countries, and the International Union of Geological Sciences Commission for the Management and Application of Geoscience Information (IUGS-CGI).

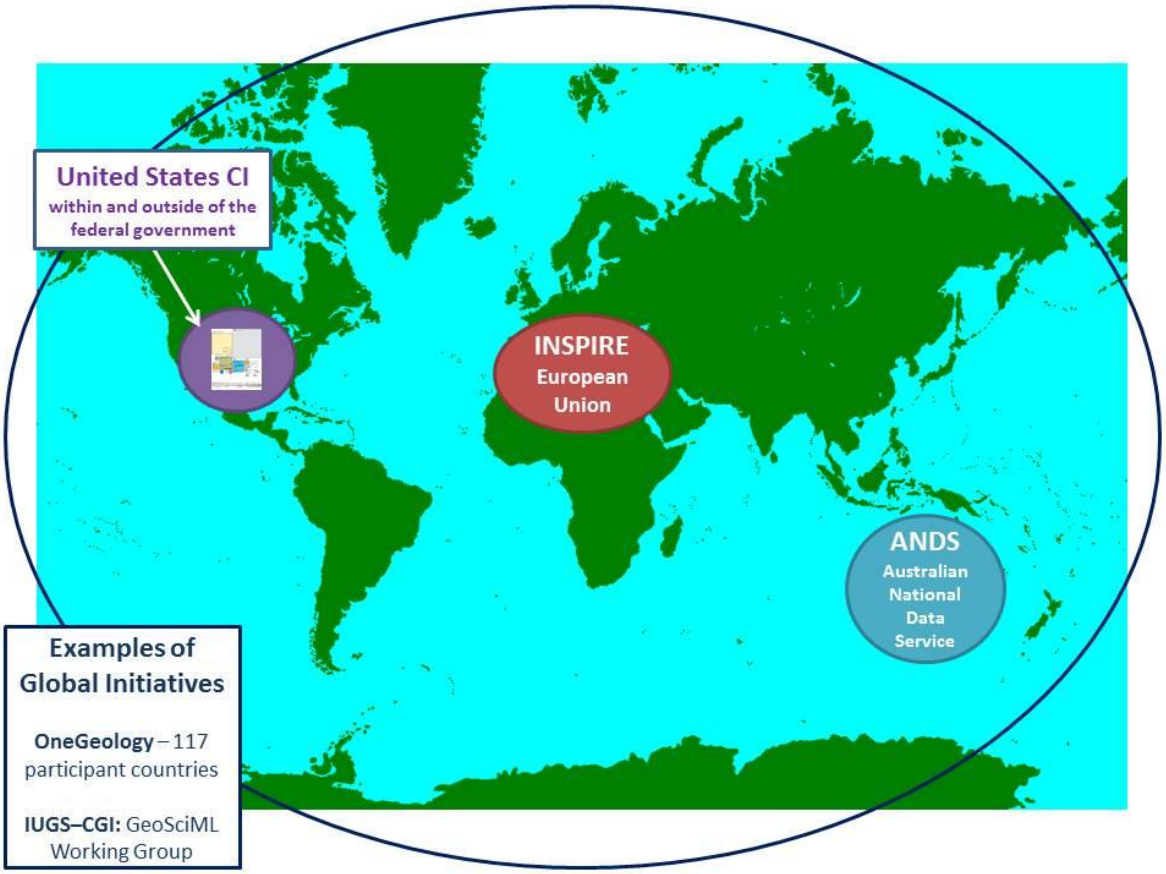

FIGURE 4. CYBERINFRASTRUCTURE IN THE UNITED STATES FITS INTO CYBERINFRASTRUCTURE WORLD-WIDE.[29]

The examples presented here are just a few of the possibly thousands of cyberinfrastructure initiatives worldwide. One component of building EarthCube could be cataloging and identifying potential areas of coordination between existing cyberinfrastructure systems, starting in the United States and expanding globally as EarthCube grows.

[29] Base diagram from http://en.wikipedia.org/wiki/File:World_Map_flat_Mercator.png.

2.4 EMERGING EARTHCUBE GOVERNANCE CONCEPTS

We reviewed several governance concepts as part of our background research review on governance and our current community engagement efforts. The three emerging concepts presented here have been brought forth by EarthCube participants and have influenced our thought process in creating this document. Concepts for EarthCube governance are not limited to the ideas presented here and this list should not be considered exhaustive, only a brief overview of the emerging concepts that have influenced this paper. As a wider spectrum of individuals and organizations are engaged in EarthCube, a plethora of potential governance concepts is likely to emerge. Although it will eventually be necessary to decide on a structure for EarthCube governance, it is premature at this time to make any final decisions, given that only a sliver of the potential EarthCube community has been engaged thus far.

2.4.1 EMERGING GOVERNANCE CONCEPT: CROSS-DOMAIN INTEROPERABILITY

This perspective, brought forth in the EarthCube Cross-Domain Interoperability Concept Team Roadmap, defines the scope of EarthCube as cross-domain integration of domain systems and research sites. Although these domain systems and research sites maintain local autonomy in the way they govern themselves, they are linked to EarthCube through cross-domain integration mechanisms. These mechanisms include, but are not limited to:

1. management of standards and identifiers
2. long term preservation
3. social networking
4. provenance, trust
5. research workflows
6. governance policies and agreements
7. brokering services
8. compute storage and resources
9. collaborative code development
10. search over federated catalogs
11. synthetic curated datasets
12. vocabulary cross-walks

The Cross-Domain Interoperability Roadmap Version 1.1 describes a general vision of how EarthCube could be organized:

> From the perspective of cross-domain interoperability, we organize the presentation by key types of CI resources that enable discovery, interpretation, data access, data integration and processing across geoscience domains.
>
> Our experience in several disciplinary data system, and the collection of domain architectures we have been assembling, suggests that these functions are typically represented in the following basic infrastructure components which need to be present to enable cross-domain interoperability in the geosciences: metadata catalogs, at the appropriate community defined granularity, that provide standard discovery services over datasets, data access services and other resources of the domain; vocabularies that support unambiguous interpretation of domain resources and metadata; services used to access

data repositories and other resources including models, visualizations and workflows, and support data processing, modeling and visualization; and formal information models that define structure and semantics of the information returned on service requests.

"A general vision of EarthCube logical organization…is of an integrated information system (or a "system of systems") that includes research observatories generating large volumes of observations and analytical/simulation results, domain systems that publish the information according to community conventions about data models, vocabularies and protocols, and a cross-domain knowledge layer that includes federated catalogs, normalized and curated datasets integrating data from domain systems and observatories, cross-linked vocabularies, service brokers, as well as social networking, governance and compute infrastructure. This conceptual diagram is consistent with the consensus EarthCube diagram presented earlier[30]…and emphasizes the central role of the cross-domain interoperability layer enabling discovery, interpretation, data access and integration across domain infrastructures: the component named "EC Infrastructure[31]… has similar content and functionality to the upper layer in Figure [5], while "domain clouds"[32]…emphasize the same functions of cross-domain systems as the vertical "domain boxes" in Figure [5]."[33]

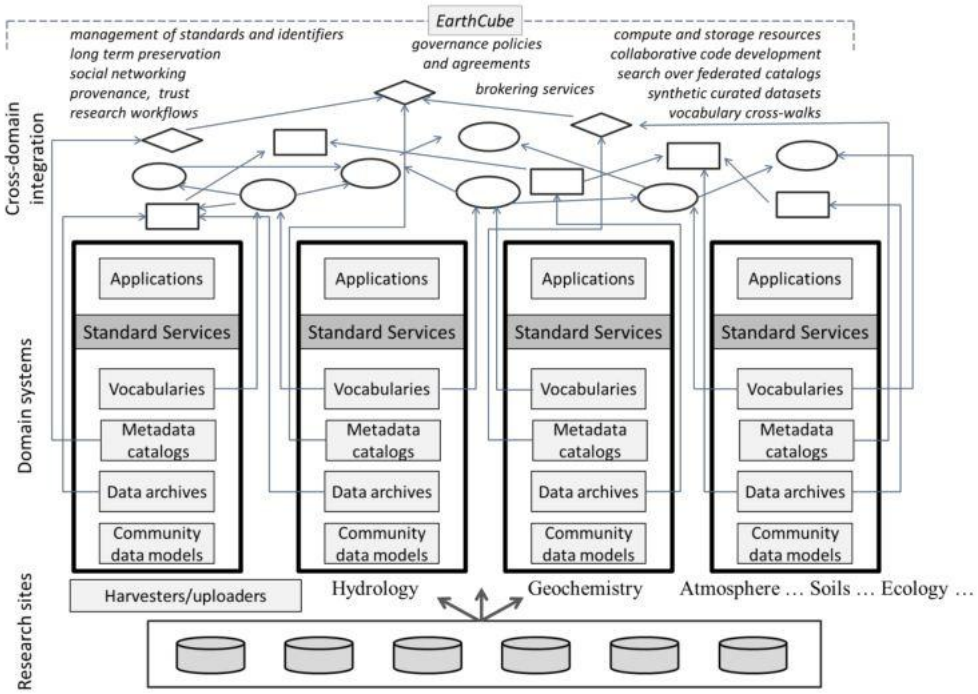

FIGURE 5. CROSS-DOMAIN EARTHCUBE GOVERNANCE[34]

[30] See Figure 13.
[31] Ibid.
[32] Ibid.
[33] Zaslavsky, Ilya, et al., "Cross-Domain Interoperability Roadmap, Version 1.1," EarthCube Cross-Domain Interoperability Concept Award, Released August 16, 2012, 42.
[34] Zaslavsky, Ilya, et al., "Cross-Domain Interoperability Roadmap, Version 1.1," 43.

"Domain infrastructures have been the focus of NSF investment in geoscience cyberinfrastructure over many years, and need to be leveraged within EarthCube. Presently, the interoperability of the domain infrastructures is limited. Several social and technical challenges contributing to this limitation have been reviewed in earlier sections [of the Cross-Domain Interoperability Roadmap]. One of the central issues is lack of a separately governed cross-domain interoperability layer. Another central issue is lack of standardization of the infrastructure components listed above. General standards for these components have been proposed, e.g. Open Geospatial Consortium's (OGC) Catalog Services for the Web (CSW) for interoperable catalogs, Simple Knowledge Organization System (SKOS) for vocabularies, OGC Sensor Observation Service (SOS) for requesting observational data, and OGC Observations and Measurements as a general information model and encoding schema, with emerging domain specifications such as the OGC WaterML 2.0, GeoSciML or CSML.

By utilizing these or similar standards, EarthCube-enabled research designs can take advantage of data discovery across disciplines using the commonality in key data characteristics related to shared models of spatial features, time measurements, and observations. Data can be discovered via federated catalogs and linked nomenclatures from neighboring domains, while standard data services can be used to transparently compile composite data products.

Both standardization of key interfaces supporting discovery, interpretation, access and integration of domain resources, and development of cross-domain mappings and brokering solutions, are critical components of EarthCube CI. A mature cross-domain CI needs a combination of the both approaches. There is a strong trend towards development of standards-compliant software components, as support for community standards is added to mainstream software and a growing number of data and other resources are becoming available via standard service interfaces. For government data providers, this trend has been recently highlighted in the Digital Government Strategy (Executive Office of the President, 2012). At the same time, new scientific feature types are being constantly introduced by research practice and may initially lack standardization: for these cases information mapping and brokering solutions are required. The two approaches are complementary and strongly interrelated."[35]

[35] Cross-Domain Interoperability Roadmap, Version 1.1, 2012, 42.

2.4.2 EMERGING GOVERNANCE CONCEPT: THREE TIERS OF GOVERNANCE

This model of EarthCube Governance is based on the scope and extent of services and capabilities that enable and facilitate cross-disciplinary research. (This notion of EarthCube Governance was first presented in Carroll Hood's white paper to the Governance Framework Steering Committee "Developing an Optimal Governance Framework for EarthCube.") In general, this governance concept presents three categories of these services and capabilities

- **Enterprise-Level:** Those that are common to all or nearly all domains and can be managed and operated effectively in a consolidated fashion to achieve economies of scale.
- **Touchpoints:** Those that are common to all of most domains and require cross-domain coordination yet should be optimized locally within each domain.
- **Domain-Specific:** Those that are unique to a particular domain and are maintained and operated autonomously within that domain.

THE DEVELOPMENT OF THIS GOVERANCE MODEL

This decomposition of services and capabilities in this fashion has specific implications on EarthCube Governance. The first step in describing those implications is to define the scope and extent of the EarthCube enterprise. A simple yet elegant view of the EarthCube enterprise was provided by NSF during the June 2012 Charrette (Figure 6).

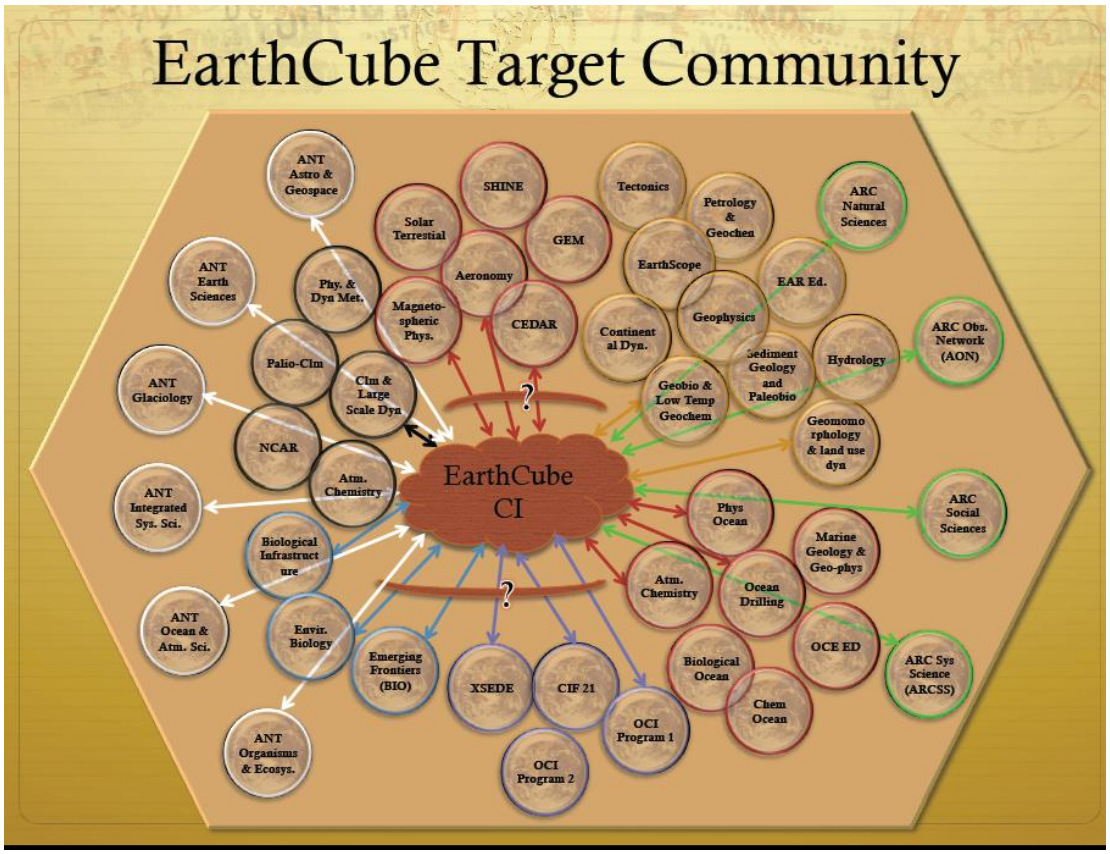

FIGURE 6. SELECTED EARTHCUBE TARGET COMMUNITY WITHIN THE NATIONAL SCIENCE FOUNDATION COMMUNITIES[36]

This image depicts EarthCube as a cyberinfrastructure (CI) that is the centerpiece of and is connected to a number of other NSF-funded or sponsored initiatives ("the balls"). Two important items to note from this diagram:

1) The nature of the connections between EarthCube and the other initiatives is not specified; and
2) Little if any connectivity exists between the individual balls.

In order to gain some insight into what those connections might entail (and what the implication of the nature of the connections is to governance), it is useful to consider an idealized model of a collaboration initiative (CI). Figure 7 is an idealized model of the internal structure of any single ball that that is connected with EarthCube, as depicted in Figure 6. This idealized depiction of a CI is not intended to be comprehensive (some initiatives may have additional or different capabilities, some may have less) but only a notional representation of a cyberinfrastructure or collaborative initiative.

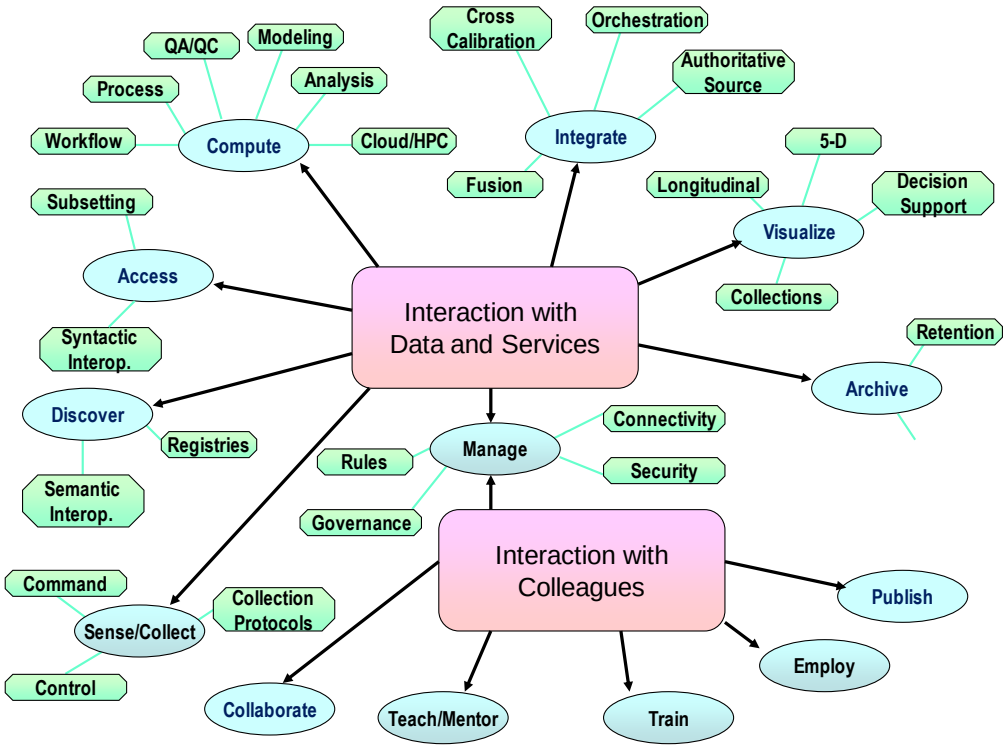

FIGURE 7. NOTIONAL FUNCTIONS WITHIN A GIVEN NSF COMMUNITY.[37]

In general, Figure 7 suggests that the primary purposes of a CI or collaboration initiative (e.g., one of the balls) is to facilitate interaction with both data and services and between colleagues. The blue

[36] National Science Foundation, EarthCube Target Communities Graphic, 2012.
[37] Hood, Carroll, "Developing an Optimal Governance Framework," 2012, 4.

ovals represent primary functional capabilities. The green blocks represent elements or distinct aspects of those functions. In some sense, each of these initiatives (each of the balls in the EarthCube Target Communities diagram) is autonomous, self-governed and has a defined community of interest. We will define that combination of characteristics as being "cyberinfrastructurally sentient."

This method of categorizing services enables us to speculate the nature of the connections between any one initiative and EarthCube (depicted in Figure 6) and between initiatives (relationships between initiatives are NOT depicted Figure 6). We can do this by placing each of the blue ovals into one of three categories mentioned above.

1) **Enterprise-Level Governance: common functions/services:** These are functions that are common across the various initiatives. NSF can realize cost savings through economies of scale by managing them at an enterprise level. Examples include:
    - Compute: e.g., single, scalable cloud augmented by appropriate level of High-performance Computing (HPC)
    - Visualize: e.g., common core capability for visualizing, 1-D datasets, 2-D datasets , raster images, etc.
    - Publish: defined by best practices (including methods for data citations)
    - Create and implement an EarthCube vision, mission and goals
    - Management: create and implement by-laws and charter
    - Engage and coordinate across the EarthCube community

2) **Touch Points:** These are functions that share a common architecture, can be logically connected but will likely be tailored within each domain. Example include:
    - Discover: Distributed but logically connection registries
    - Archive: Distributed but logically connected repositories
    - Access/integrate: multiple solutions that depend on the degree to which semantic and syntactic barrier to interoperability are normalized

3) **Domain-specific:** These are functions that are unique and provided/managed within a particular initiative or domain. Examples include:
    - Sense/collect: unique data type and collection protocols
    - Manage: different local governance archetypes for different communities
    - Governance framework within each domain.

With this categorization, the nominal depiction now looks like this, which has implications for the type and nature of connections (Figure 8).

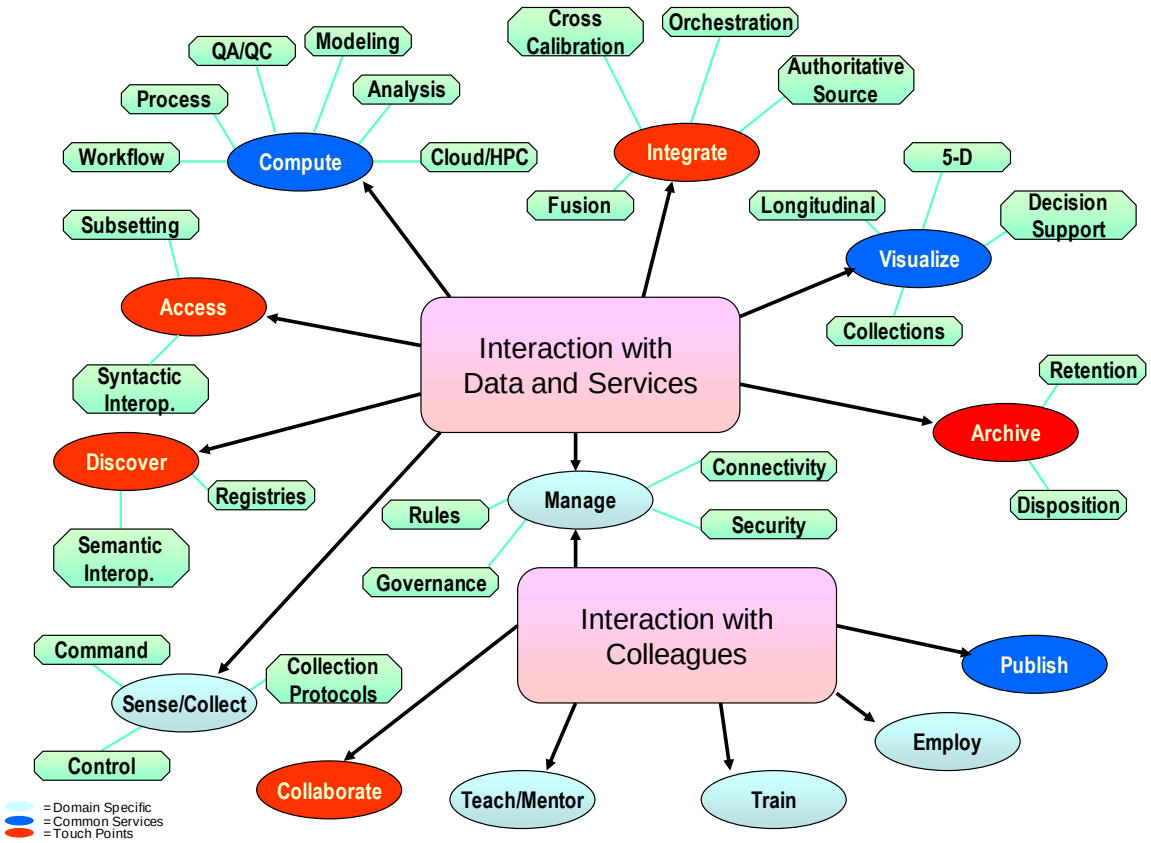

FIGURE 8. CATEGORIZATION OF GENERAL FUNCTIONS WITHIN A GIVEN NSF COMMUNITY.[38]

Light blue ovals represent domain-specific functions; Dark blue ovals represent enterprise-level common functions and services present across all of EarthCube; and Red ovals represent touchpoints that share a common architecture but are locally optimized within each domain,. This model is not necessarily relegated to only NSF communities, but could apply to any community of practice, organization, institution, and consortium that participates in EarthCube. Figure 9 below provides a visual depiction of how common services and touchpoints provide a set of "connecting tissue" between the various balls in Figure 6 (although, in this case, a "ball" is now represented by a single plane.)

[38] Hood, Carroll, "Developing an Optimal Governance Framework," 2012, 5.

Thus, the connection between any series of balls looks like this:

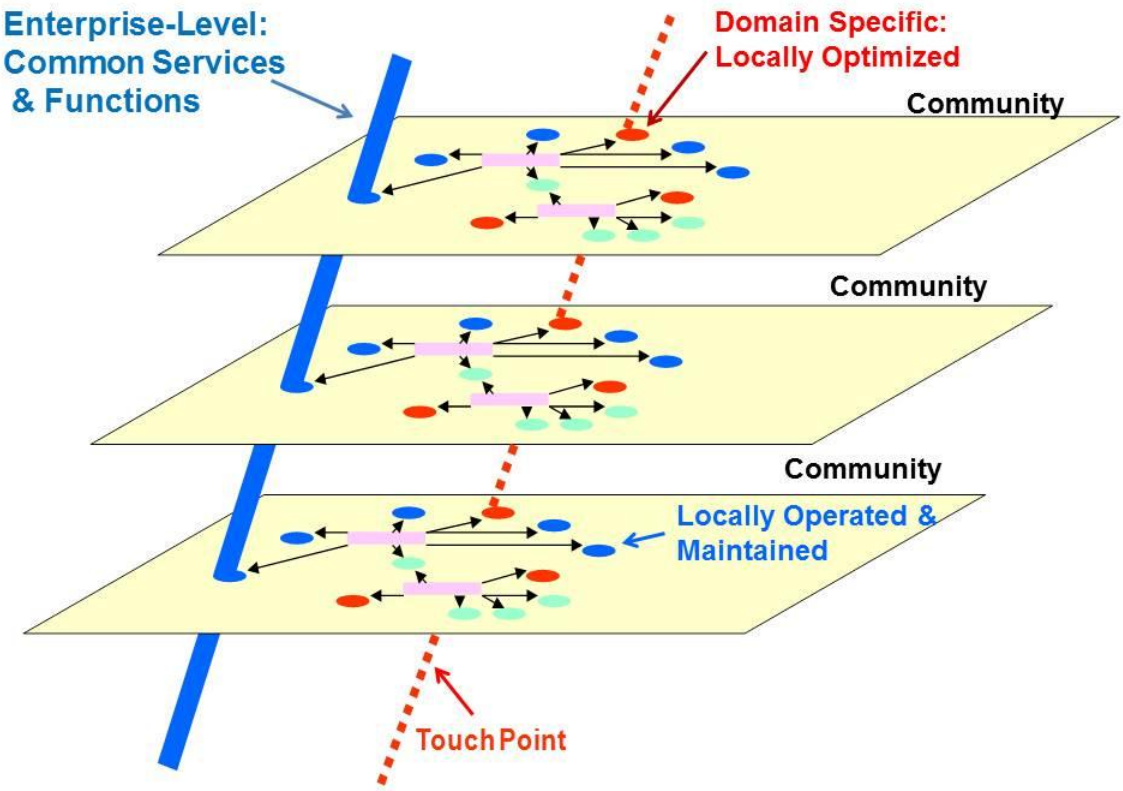

FIGURE 9. DIAGRAM HOW COMMUNITIES, ORGANIZATIONS, INSTITUTIONS, AND CONSORTIUMS PARTICIPATING IN EARTHCUBE MIGHT BE LINKED TO ENTERPRISE-LEVEL EARTHCUBE GOVERNANCE AND TO EACH OTHER.[39]

The dark blue line represents an enterprise level service that is managed at an enterprise level outside of the individual communities. Many of the enterprise level services extant today lie outside of EarthCube and outside of EarthCube governance per se. Rather, an EarthCube function will be to engage with externalities for coordination and leveraging of resources.

The dotted red line is service that is coordinated at an enterprise level, is logically connected, but is locally optimized within each individual community. The light blue ovals continue to be operated and maintained locally. Note that each community maintains its local science autonomy; each merely allows these common and logical touchpoint services to be integrated into their fabric.

[39] Hood, Carroll, "Developing an Optimal Governance Framework," 2012, 6.

If we agree that Figure 9 provides a genetic model of how the balls in Figure 6 could be connected, then the implications for EarthCube and EarthCube governance are now straightforward:

- At this stage in it maturity (e.g., the first few years) , EarthCube is not "cyberinfrastructurally sentient." It is merely the connecting tissue and glue that enable the initiatives that are "cyberinfrastructurally sentient" to interoperate at some prescribe Service Level Agreement (SLA).

- EarthCube Governance, at this stage in its life cycle, has the following responsibilities:
  - Institute effective outreach methods to ensure effective community input and feedback so that all IT decisions are science-driven
  - Develop and manage the architecture and design for the connecting tissue/glue
  - Conduct the engineering trades that support the development, implementation, and management of the following activities:
  - Develop/implement and manage the common services; support the evolution of existing initiatives into the common service paradigm
  - Develop/implement and manage the touch point services; ensure that the touch point services are logically integrated correctly; support the evolution of existing initiatives into the touch point paradigm
  - Ensure that all new initiatives fold common services and touch point services into their design early on
  - Identify key functional gaps, manage and mitigate them
  - Identify and fund pilot projects/demonstrations to support targeted risk reduction and gap closure
  - Ability to adjudicate IT issues

Thus, there are three-tiers to this governance model:

1) Enterprise level Governance
2) Coordinated Governance between Enterprise-level and Local-level
3) Local level Governance (e.g., , within each ball)

THE EVOLUTION OF THIS GOVERNANCE MODEL

Over time, as common services become more pervasive and as the distinction between individual domain initiatives begins to blur (as interdisciplinary research becomes more and more the norm), EarthCube will evolve to become more "cyberinfrastructurally sentient" (i.e., it will become more autonomous, it will begin to self-govern its own cross-domain specific services and will has its own distinct science community.) The small circle on the right side of Figure 10 is intended to represent (in Figures 11-13) the complete richness of an individual CI community connected to EarthCube (the left side of Figure 10).

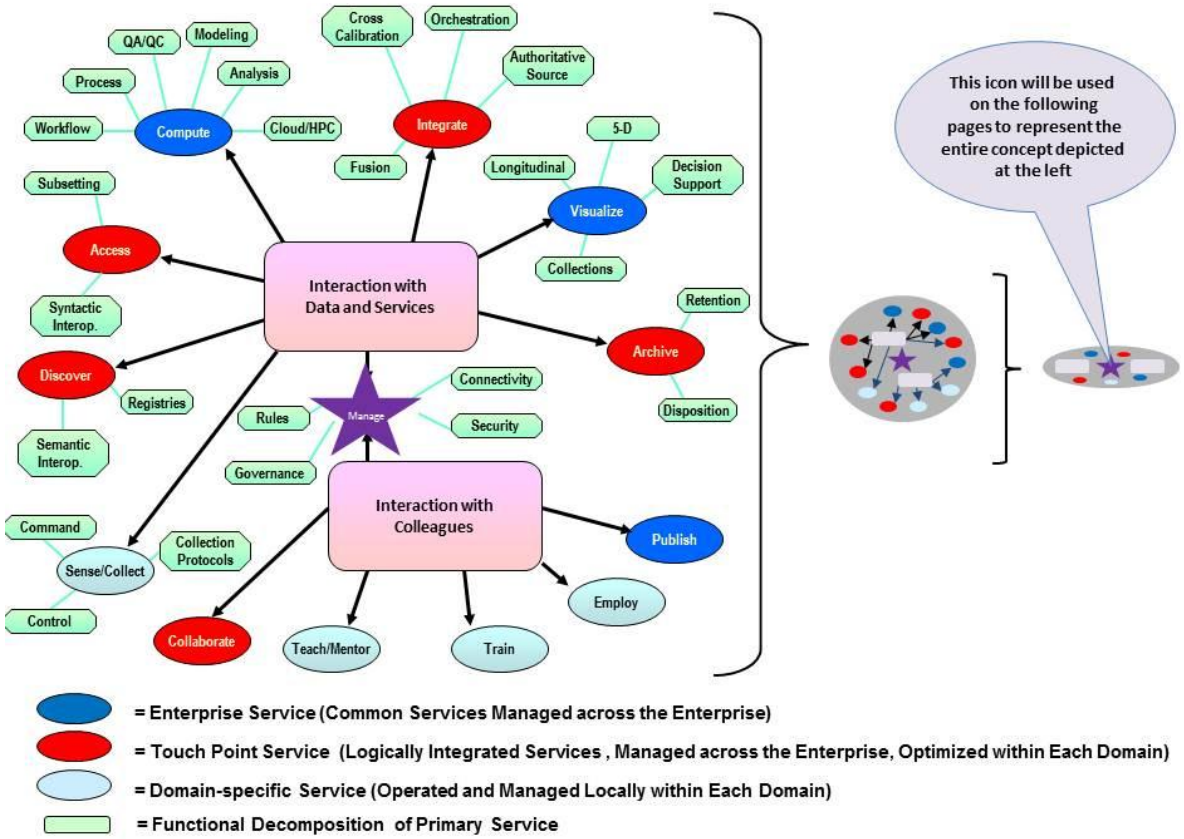

FIGURE 10. CYBERINFRASTRUCTURE,OR OTHER COLLABORATION INITIATIVES WITHIN EARTHCUBE WILL BE REPRESENTED AS THE SMALL OVAL ON THE RIGHT.[40]

[40] Modified from Hood, Carroll, "Abstractions on EarthCube Governance," virtual presentation to EarthCube Governance Steering Committee, July 27, 2012.

The present state of pre-EarthCube NSF governance is depicted below in Figure 11. In this state, NSF's sphere of direct influence governs the individual principal investigators (PI's) of each Working Group and Concept Team, in addition to PIs from other NSF-funded projects, and communities of interest from the various science directorates within NSF, such as GEO and ATM (geosciences and atmospheric sciences). Each PI or community of interest makes local governance decisions regarding planning and resource allocation, operational decisions or trade-offs, resource adjudication, and external collaboration or coordination, as needed.

Standards organizations, external systems, other funding agencies, and external organizations are outside the NSF sphere of direct influence, but many of expressed interested in becoming involved in EarthCube. Their role at this time is to coordinate, collaborate, exploit, utilize and participate in EarthCube activities taking place at this time.

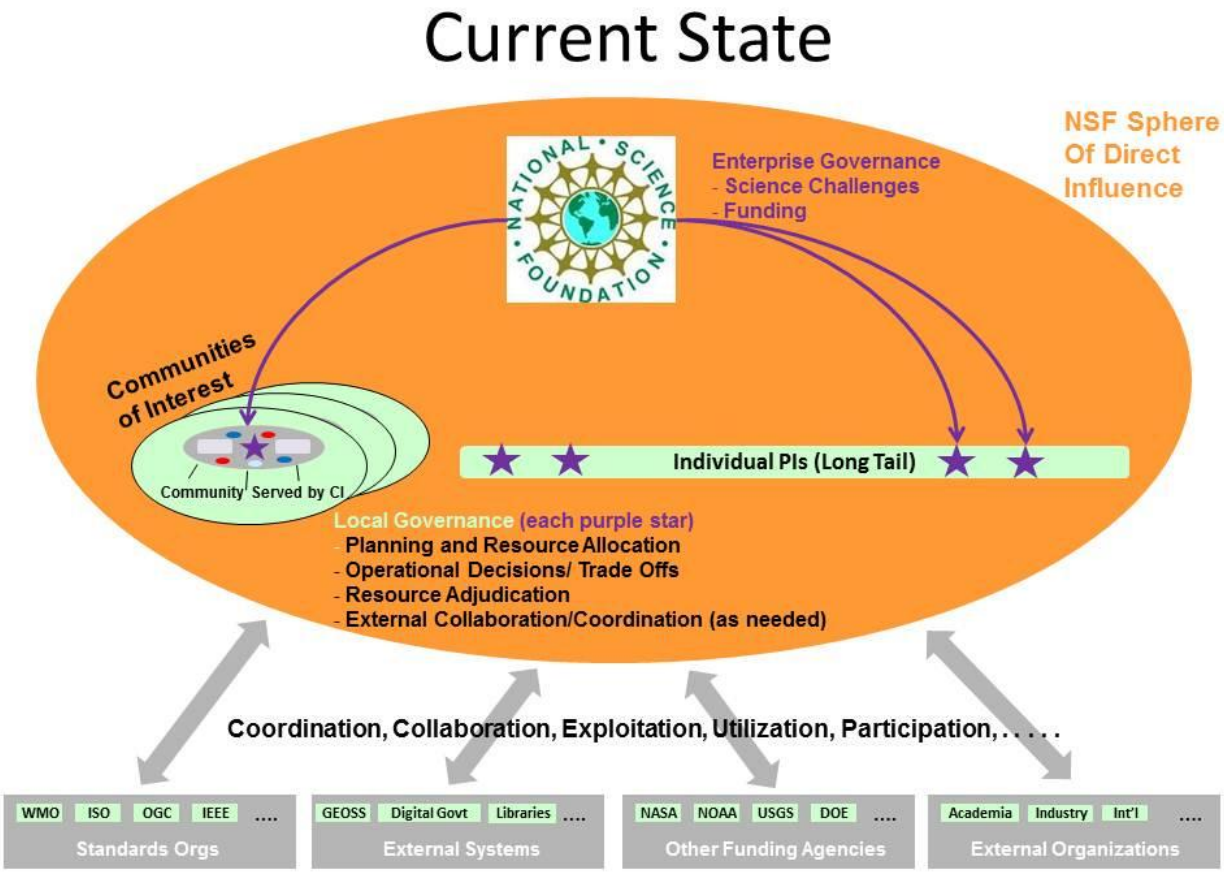

FIGURE 10: CURRENT STATE OF EARTHCUBE GOVERNANCE.[41]

[41] Modified from Hood, Carroll, "Abstractions on EarthCube Governance," virtual presentation to EarthCube Governance Steering Committee, July 27, 2012.

The next step in the evolution of EarthCube governance is the implementation of an initial governance framework (likely to occur in early to mid-2013). In this way, initial EarthCube governance implements common enterprise-level functions, coordinates touchpoint services between EarthCube communities and supports interfaces with external elements (including standards organizations, external systems, other funding agencies, and external organizations), and enables long-tail scientists to participate in EarthCube. EarthCube governance does not impact local governance of each community (Figure 12).

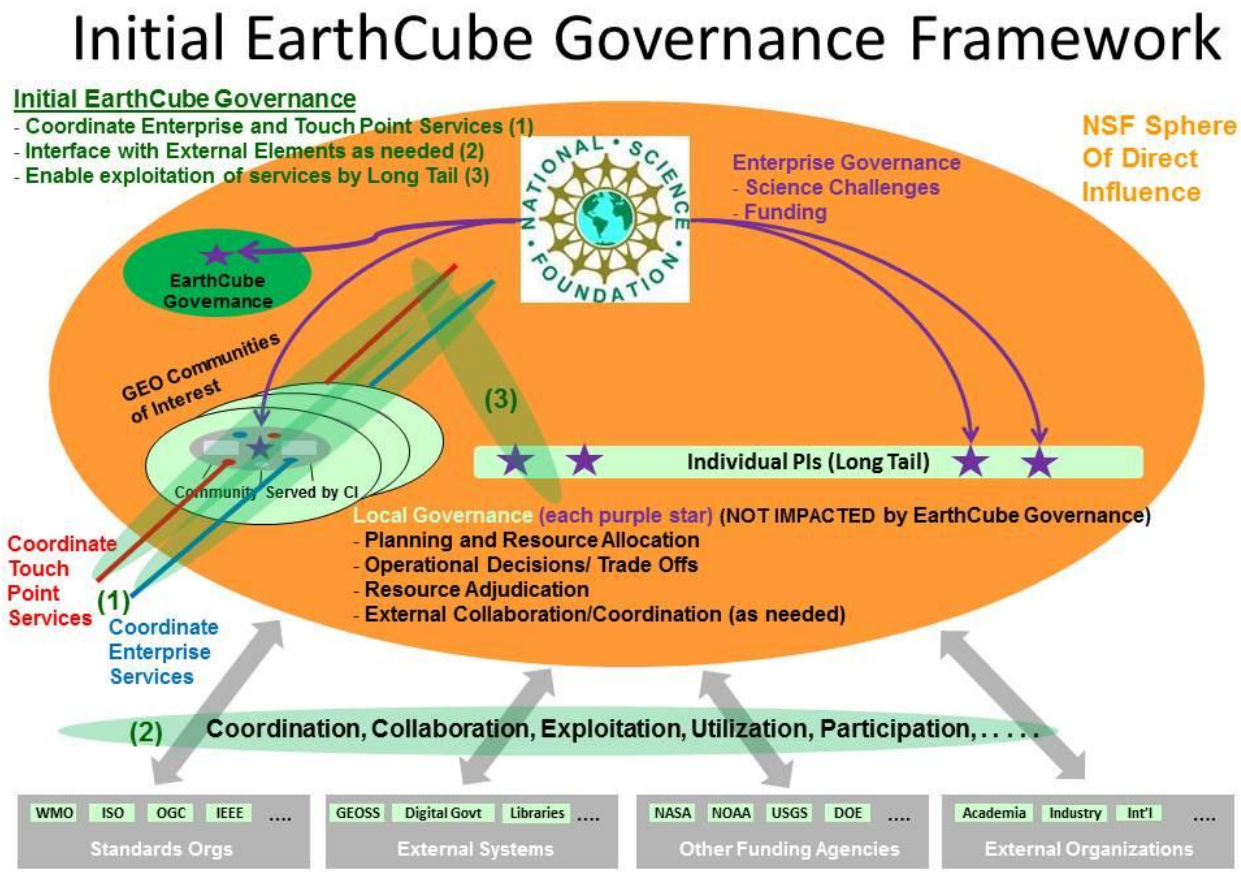

FIGURE 12. INITIAL EARTHCUBE GOVERNANCE FRAMEWORK.[42]

[42] Modified from Hood, Carroll, "Abstractions on EarthCube Governance," 2012.

A possible evolved EarthCube governance framework is depicted in Figure 13. Instead of just being the IT-based enabler for cross-domain research, EarthCube may now begin to operate as its own science initiative (no longer just the connecting tissue and glue). It will now include the infrastructure needed to create and maintain science priorities internally (this power would necessarily be shared with the individual initiatives). Although the need to coordinate the IT infrastructure will always remain, EarthCube may facilitate a shift toward cross-domain research, and the Governance will be modified accordingly. Since this phase of the EarthCube lifecycle is several years away and undoubtedly will be evolutionary, it is premature to speculate what the optimal governance framework will be at that time, as innovations in technology and social interaction are likely to change the research enterprise significantly.

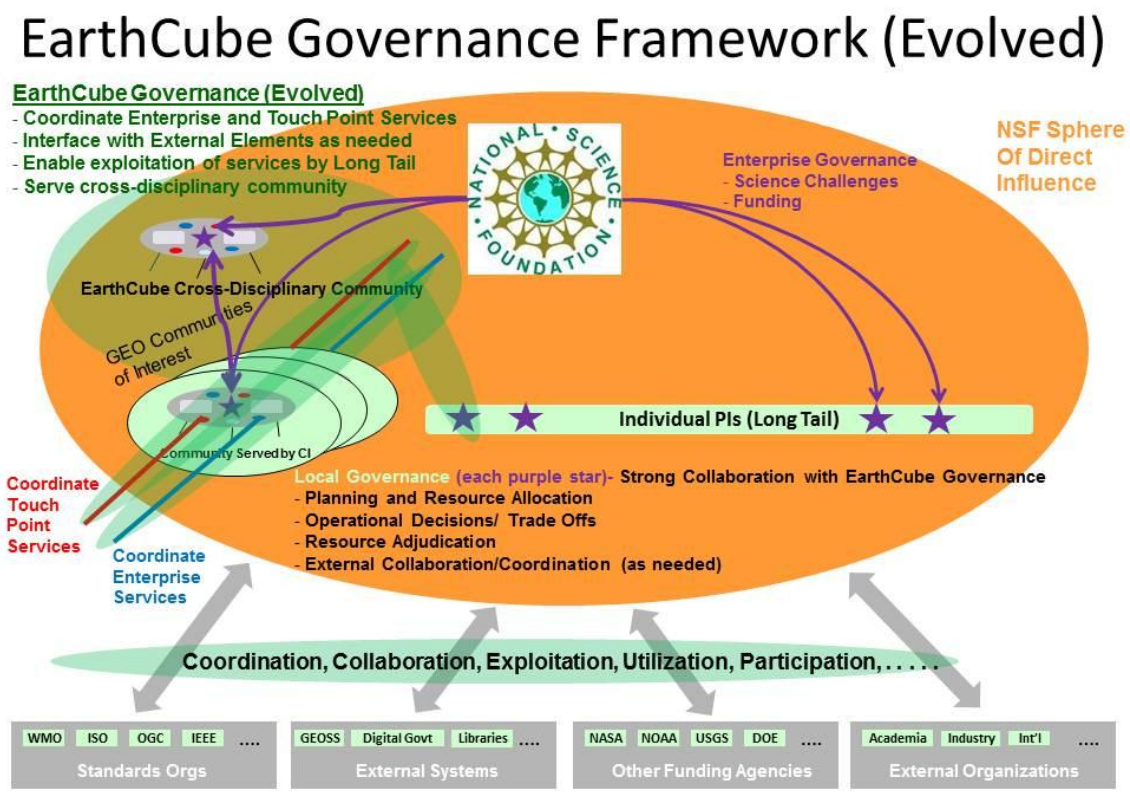

FIGURE 13. EVOLVED EARTHCUBE GOVERNANCE FRAMEWORK.[43]

[43] Modified from Hood, Carroll, "Abstractions on EarthCube Governance," 2012.

2.4.3 EMERGING GOVERNANCE CONCEPT: OUTCOMES OF A CONCEPT AWARD PRINCIPAL INVESTIGATOR MEETING

This emerging governance concept presents the reference architecture first proposed by the Brokering Concept Team, and later modified and adopted as a recommendation for a possible model for an EarthCube reference architecture by the EarthCube Principal Investigator (PI) Concept Award Meeting in Boulder, CO on July 10, 2012. The PI meeting group came to consensus on a high-level architecture for EarthCube. This model is presented in Figure 14 below.

The PI Meeting Report describes the architecture model for EarthCube in this way:

> "Sample architectures or architectural elements were offered in a number of the roadmaps, both from CA and Community Groups (CG) teams. The architectures offered generally had similar elements, but on the whole varied in appearance and functionality. A discussion on reference architecture for EC was organized as part of the meeting to look for convergence to a common construct that could serve the broad EC community and users. This was done in three phases: outlining the benefits of a conceptual-level architecture framework, review of current recommendations and then the formulation of a common model. At the end of the discussion, the participants formulated a list of recommendations that came from these discussions.
>
> Reference architecture is important for EC roadmap development from several viewpoints:
>
> 1. As a platform on which to build flexible applications
> 2. As a basis for assessing "EC readiness" and gap analysis
> 3. As a way to scope EC governance
> 4. As a way to define interfaces between components
> 5. As a way to specify coordination between existing components and define missing components, and prioritize development
> 6. Continuously updated as EC progresses
> 7. An enabling basis for technology evolution and innovation
>
> Discussions of the EC architecture frameworks from the various EC roadmaps led to a common vision for a high level EC architecture construct given in the figure below. The reference architecture can be characterized as an environment enabling effective collaboration across multiple science domains and domain data systems to address challenges in the geosciences through efficient discovery and re-use of data, information and knowledge. Interactions within existing domain infrastructures use established catalogs, semantics, data access mechanisms, information models and policies. These components, when exposed via community standards, facilitate interactions between different domain infrastructures. In particular, brokering services and related cross-domain components (semantic cross-walks, cross-domain registries, information model mappings) are employed to mediate between community resources and users/researcher clients that follow different domain specific models. In this manner, the EC infrastructure facilitates cross-domain science through creation of mechanisms for discovery, access, processing and

semantic mediation; workflows services are provided as well for distributed process chaining.

EarthCube should enable reproducible science, collaboration across the "long-tail" of researchers, and preservation of research results. A strong goal for the EC is the ability for a research group to collaborate on generation of new research results based on current repositories and then form a new community resource to enable future research to build upon their results. Collaboration between domains/disciplines will be supported within a collaborative environment, which would support social networking, consensus building and community seeding capabilities." [44]

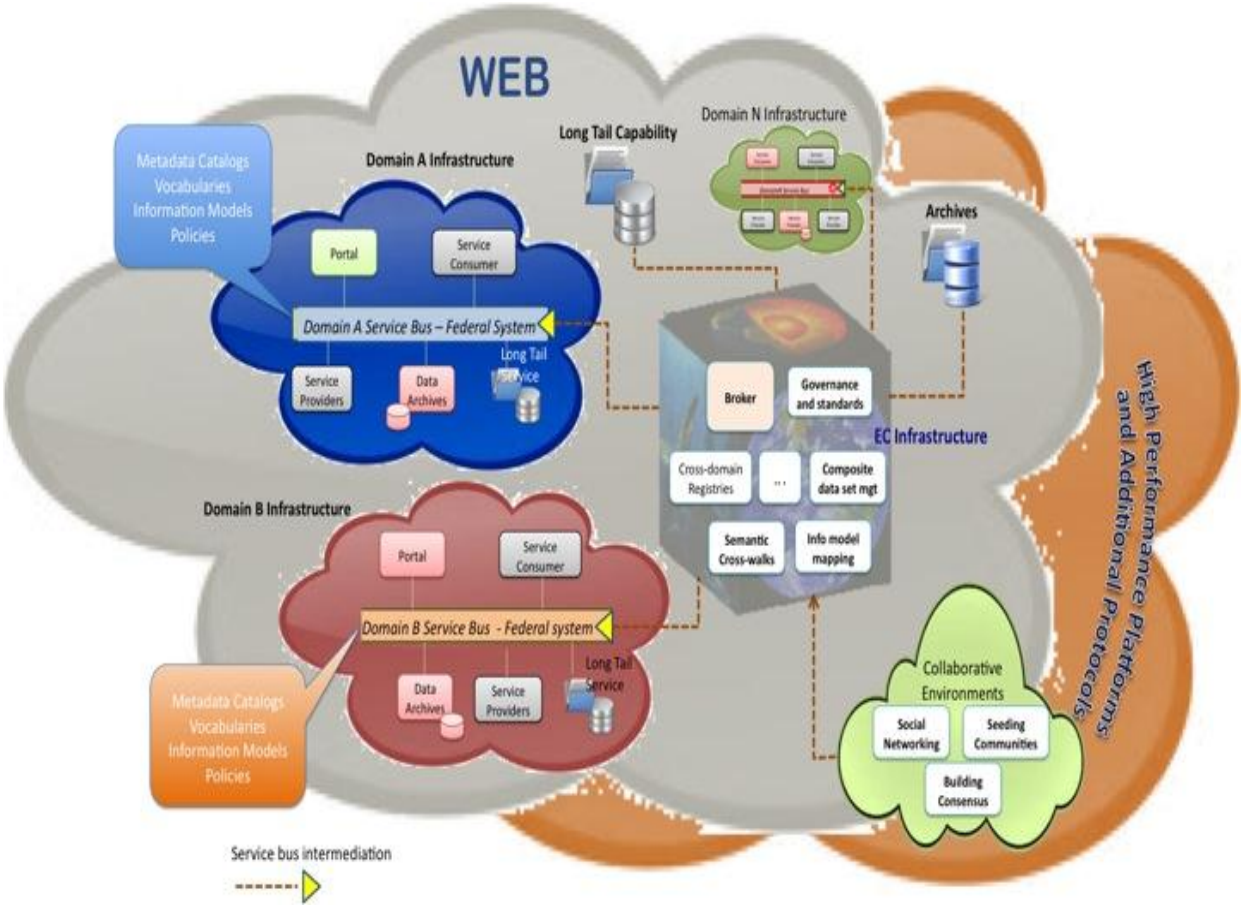

FIGURE 14. EARTHCUBE ARCHITECTURE RECOMMENDATION RECOMMENDED AT THE EARTHCUBE CONCEPT AWARD PRINCIPAL INVESTIGATOR MEETING IN BOULDER, CO, JULY 10, 2012.[45]

"The figure shows various disciplines and domains with their own community infrastructure. Both the "long-tail" and smaller archives are addressed. The system includes

[44] "Summary and Recommendations from EarthCube CA PI Meeting," 2-3.
[45] Ibid., 4.

both web-based and high performance platform and other protocol support. The fundamental attributes of the EC Infrastructure element allow for services that link multiple domains without imposing additional burdens on the participants, be they archives or users or others. The EC Infrastructure addresses not only current services such as discovery and access, but desired capabilities such as quality and provenance and should be designed to evolve as EC progresses.

These attributes are derived from a common vision starting from:

1. Existing domain systems, with EC managing interoperability resources (catalog federations, vocabulary cross-walks, service brokers, information model profiles, standards, etc.)
2. Common cross-domain infrastructure needs (collaboration, archiving, grids, workflows)
3. A concept of a DataSpace environment common across "big-head" and long-tail" of science, where different types of data can be shared, explored and annotated
4. An understanding that interoperability is not only a technical issue, but also a social, cultural and legal issue
5. The need to minimize the burden on participants/users to facilitate acceptance and broad use.

Recommendations from the discussions are:

1. Review the common architecture framework figure with the broader EC community for feedback and buy-in
2. Understand architecture viewpoints and gaps: missing components, interfaces and processes; recommend best architecture practices and convergence
3. Assess how close existing Cyber Infrastructures (CI) are to implementing the common architecture or architectural elements
4. From the assessment, identify and develop missing components, interfaces and processes in the context of use cases
5. Test applicability of the architecture for different EC constituents: large domain systems; large cross-domain projects; small "long-tail" projects; different pairs of domains
6. Examine how evolvable it is, given science and computing trends, addressing extensibility mechanisms and coordinating within a governance framework
7. Enable tight non-web coupling for high performance computing
8. Define Scope of EC and its governance and understand the impacts on architecture and priorities
9. Provide an open environment which is easy to join and easy to use
10. Demonstrate Architecture capabilities through use cases" [46]

[46] "Summary and Recommendations from EarthCube CA PI Meeting," 4-5.

SECTION 3.0: INITIAL GOVERNANCE FRAMEWORK FOR EARTHCUBE

Based on eight months of research on governance theory and models, historical infrastructure case studies, and initial community engagement activities with Earth scientists, industry, government agency and international representatives, we compiled and synthesized system-wide governance requirements to draft an initial set of EarthCube governance recommendations, governance functions and guiding principles. These functions, guiding principles and recommendations form the basis of an initial Governance Framework that will be revised and updated as part of an aggressive community outreach and engagement program. We plan to incorporate additional community input and release a modified version of the Governance Framework at the beginning of 2013. A summary of our initial findings is presented below.

3.1 EARTHCUBE GOVERNANCE FUNCTIONS

We developed a master list of 143 functions (see Appendix 4: Governance Functions Master List) collected from research on governance, the EarthCube Working Group and Concept Team draft roadmaps released prior to the June 2012 EarthCube charrette, community input during Governance Steering Committee virtual webinars, and the June charrette, a July 2012 EarthCube Principle Investigator workshop, and the EarthCube sessions at the ESIP Federation Mid-Year Meeting in Madison, WI, July 17-20.

We then categorized and consolidated the 143 functions into four categories:

1. Identify and Implement an EarthCube Vision, Mission, and Goals.
2. Engage and Coordinate Across the EarthCube Community.
3. Management: Create and Implement EarthCube By-Laws and Charter.
4. Develop and Maintain a Viable Architecture and Concept of Operations that Enable the Realization of the Goals and Objectives of the EarthCube Vision.

These functions identify and prioritize what needs to be carried out by EarthCube governance at this time. They are non-prescriptive, however, thereby leaving room for any number of interested individuals, institutions, organizations and consortiums to propose to NSF and the EarthCube community how these functions might be carried out. Each function has a policy, infrastructure and services component, or a combination of two, or all three components. Each of the components is ordered according to what we feel should be implemented first, on a scale of 1-3 (1 = near term or next 6 months, 2 = mid-term or next 6-24 months, and 3 = long-term or next 2-5 years). Each function is then assigned an overall prioritization, using the same 1-3 scale (Table 2).

This list is a living document, open for community editing and comment here: https://docs.google.com/spreadsheet/ccc?key=0AhE8IKrxOMWndDZfTXBUUFE1QmhOT01XT2JrZ3BQSWc#gid=0.

Initial EarthCube Governance Functions

	EarthCube Governance Functions	Overall Timing	Policy	Services	Infrastructure
	1. Identify, Implement and maintain an EarthCube vision, mission, and goals.	**1**	**1**	**2**	**2**
1	Lead the implementation of the EarthCube vision, mission and goals.	**1**	2	2	2
2	Develop metrics to evaluate progress towards EarthCube vision, mission and goals.	**1**	1	3	1
3	Establish processes to prioritize goals while building an open and transparent collaborative community for all EarthCube participants.	**2**	2	3	NA
	2. Engage and Coordinate Across the EarthCube Community	**1**	**1**	**1**	**1**
1	Establish processes to determine how EarthCube will collaborate with other organizations.	**1**	1	1	1
2	Promote education and outreach.	**1**	1	1	1
3	Establish processes to seek and prioritize pilot projects.	**2**	2	2	2
4	Establish processes to identify and engage potential EarthCube participants.	**1**	1	2	NA
5	Establish processes to recognize and respond to changing community needs.	**1**	1	1	1
6	Identify and manage EarthCube-wide services.	**2**	1	3	2
7	Identify and coordinate touchpoint services (services that link organizations, etc., to EarthCube).	**2**	1	3	2
8	Promote the integration of existing communities of practice into EarthCube, and foster the establishment of needed communities of practice, such as software development.	**2**	2	3	NA
9	Manage EarthCube-wide use cases.	**2**	2	2	2
	3. Management: Create and Implement by-laws and charter	**2**	**2**	**2**	**NA**
1	Establish decision-making processes.	**1**	1	NA	NA
2	Establish processes to define an EarthCube participation policy.	**1**	1	2	NA
3	Establish processes to identify roles and responsibilities, and determine how they may be filled.	**1**	1	NA	NA

4	Establish processes to ensure broad representation in EarthCube leadership at all levels.	**1**	2	3	NA
5	Establish processes to address EarthCube sustainability.	**2**	2	NA	3
6	Provide for the adaptive evolution of the governance framework.	**2**	1	NA	2
7	Establish processes to allocate resources and make recommendations to NSF	**2**	2	NA	NA
8	Establish processes to define and maintain support services.	**2**	1	2	2
9	Establish processes to ensure EarthCube governance is scalable.	**2**	2	NA	NA
10	Establish processes to resolve disputes, address legal issues, and manage risk.	**2**	2	3	NA
11	Establish processes that encourage best practices for sharing, curation, citation, credit, production, publication, and reuse of data models, software, and services.	**2**	2	3	2
4. Develop and maintain a viable architecture and concept of operations that enable the realization of goals and objectives of the EarthCube vision		**2**	**2**	**2**	**3**
1	Create processes to evaluate technologies, produce documentation, and recommendations for systems requirements	**2**	2	NA	2
2	Create processes to identify gaps in coverage of needed cyberinfrastructure capabilities, and determine recommendations on how to fill them	**2**	2	3	3
3	Establish processes for inventorying and assessing the current state of capabilities.	**2**	1	2	3
4	Identify optimal processes to fill cyberinfrastructure gaps, whether these processes are bottom-up, top-down, or a combination of both.	**3**	2	3	3
5	Establish processes to promote the adoption, adaptation, and only as a last resort, duplication of existing, or development of new, capabilities.	**3**	2	3	2

TABLE 2. INITIAL EARTHCUBE GOVERNANCE FUNCTIONS

3.2 EARTHCUBE GOVERNANCE GUIDING PRINCIPLES

These guiding principles were developed in order to guide the implementation of the EarthCube governance functions.

1. EarthCube governance shall serve the advancement of interdisciplinary science through collaboration among community members and with other cyberinfrastructure initiatives.

2. EarthCube governance shall rely on open, transparent processes and shall vet and inform its decisions through active community engagement

3. EarthCube governance shall encourage environmentally sustainable processes and practices.

4. EarthCube governance shall support development that draws from best practices based on interoperability and reuse of resources.

5. EarthCube governance shall strive for the free and open sharing of data, information, software and services.

6. EarthCube governance shall evolve with changing technologies, practices and user needs while remaining robust.

This list is likely to evolve as the Governance Framework is vetted with EarthCube stakeholders throughout the second half of 2012. This list is a living document, open for community editing and comment here: https://docs.google.com/document/d/1HMSRwpTTPqSXnrspPdDd7LSU6bAvzHa0JxGm-AZkbzU/edit.

3.3 EARTHCUBE GOVERNANCE RECOMMENDATIONS

Based on the findings of the Governance Roadmap and the community engagement process since the June 2012 NSF EarthCube charrette, we developed a list of governance recommendations. Although these recommendations focus on the mid to long-term development of EarthCube, they should be taken into consideration during the next six months, and if necessary, improved upon as further feedback is provided. It is essential to consider the long-term implications and vision of EarthCube as we work to frame it in the present. This list is likely to evolve as the Governance Framework is vetted by EarthCube stakeholders. This list is a living document, open for community editing and comment here:
https://docs.google.com/document/d/1OqDMwRndYSt8Tx2noJbJWizeHDQ9ssNAdoFnILg_90k/edit

We recommend:

1. An interim prototype governance body should be established in the mid-term (starting early 2013) to carry out Steps 3 and 4 of the EarthCube Governance Roadmap (establish and implement Terms of Reference for EarthCube Governance), according to the findings presented in this Governance Framework document.

2. During the building of EarthCube, we recommend establishment of an umbrella entity to coordinate governance functions across EarthCube and facilitate the optimal interoperability of existing cyberinfrastructure into EarthCube and vice versa.

3. During the build phase of EarthCube, we recommend EarthCube governance to focus on the development, operation and maintenance of EarthCube-wide services and touchpoints (services that enable existing cyberinfrastructure components to interoperate effectively within EarthCube). The existing systems that are being interconnected would retain their local autonomy, continue to determine their own science priorities, and continue to serve their constituent communities. EarthCube would serve as an enabler of intra- and cross-disciplinary activity, identify gaps, and establish criteria for consistent with EarthCube infrastructure.

4. As part of the development of EarthCube-wide services and touch points, EarthCube governance would develop and maintain an effective enterprise-level architecture, proactively identify and manage enterprise-level risks, support solicitations for risk reduction prototypes, provide a community forum, and support solicitations for cross-domain testbeds.

5. As it matures, EarthCube will become responsible for discovering and encouraging the development of a set of cross-domain science priorities with the goal of creating new cross-disciplinary research communities within EarthCube and between EarthCube and external programs.

6. EarthCube will function principally as a community service provider, not a competitor to infrastructure produced by individuals, organizations, institutions or consortiums.

SECTION 4.0: COMMUNITY BUILDING AND ENGAGEMENT PROGRAM

In order to achieve its full potential, EarthCube must be connected to and be guided by the Earth science community actively engaged in research, which often involves collecting, using, and sharing data; simulating processes; and visualizing complex interacting systems. An issue that has been repeatedly identified is the engagement of both the informatics community that builds and operates cyberinfrastructure (e.g. computer scientists, technical standards experts, middleware implementers, etc.) and target EarthCube scientific end-users, the domain scientists, who are the presumptive benefactors of such cyberinfrastructure. The Governance workgroup is dedicated to ensuring that community contributions remain an essential part of EarthCube. We are carrying out an aggressive community engagement program to solicit ideas for EarthCube governance among Earth, atmosphere, ocean, computer, information and social scientists, and other interested parties and vet the draft governance framework presented in this document.

The overall community engagement strategy may be conceptualized into three iterative, often organic phases:

- **Generating Actionable Intelligence.** This describes the implementation of a capability that enables EarthCube to collect, analyze, and utilize actionable intelligence about the scientific and informatics community (see community identification, below). Activities to generate actionable intelligence will include leveraging and coordinating with the EarthCube Stakeholder Alignment Survey and creating a network map of EarthCube stakeholders.
- **Messaging.** Designing the appropriate messaging from the intelligence gathered from the earlier phase, which may involve identifying community champions, and addressing concerns (e.g. competition for research budgets and commensurate reward structures are often top concerns), and determining the most appropriate means to disseminate information to diverse EarthCube stakeholders.
- **Engagement.** Enacting engagement tools (websites, workshops, briefings, virtual and in-person meetings) for the informatics and scientific end-user communities. Included in this phase is the engagement with professional societies and non-profit special interest groups, whose early involvement is useful for issues that eventually relate to resource coordination between different constituents.

This iterative community feedback process will:

1. Vet the initial governance functions list and guiding principles presented in this document
2. Obtain website functionality and user requirements for online collaboration
3. Engage individuals and organizations not yet involved in EarthCube
4. Maintain momentum gained from the June 2012 NSF charrette
5. Identify gaps and overlaps in the existing EarthCube community

6. Provide for community engagement to minimize risk and ensure that EarthCube will be of use to the broader geoscience community.

4.1 GENERATING ACTIONABLE INTELLIGENCE

The first step in community engagement is community identification and stakeholder alignment.

Stakeholder Alignment Survey

Progress in the development of systems for governance for EarthCube will be aided by work on a parallel NSF project focused on advancing social science theory, methods, and tools for "Stakeholder Alignment" though applications with EarthCube, which is based at the University of Illinois, Champaign-Urbana. The challenges to stakeholder alignment in the Geosciences and Cyberinfrastructure are substantial, given the many different cultures, structures, goals, and motivations that need to be bridged. Stakeholder alignment is defined in this work as:

"The extent to which interdependent stakeholders orient and connect with one another to advance their separate and shared interests."

A central challenge in the governance of EarthCube involves enabling diverse stakeholders to orient and connect in new ways that deliver tangible results individually and collectively. The Stakeholder Alignment Survey Team has specified the stakeholders based on fields and disciplines. In the Geosciences this includes: Atmospheric or Space Weather scientist; Oceanographer; Geologist; Geophysicist; Hydrologist; Critical zone scientist; Climate scientist; Biologist or Ecosystems scientist; Geographers; Computer or Cyberinfrastructure scientist; and Social scientist (Anthropologist, Economist, Psychologist, Sociologist, etc.). For the Cyberinfrastructure and other categories, this includes: Data manager; High performance computing expert; Software engineer; IT user support personnel; K-12 educator; Designer/developer of geoscience instrumentation; and Environmental resource manager (e.g. local, state, or federal). This list is expected to grow as additional stakeholders relevant to the future of EarthCube are identified.

Among the ways that the stakeholder alignment project will complement the governance efforts are:

1. Use of current data from the recent stakeholder survey (with approximately 700 responses) addressing issues on access and use of data, preferences for leadership to be member elected versus NSF selected, issues of trust and transparency, and related matters.
2. Introduction of governance-related questions in future planned stakeholder surveys to provide additional inputs into the development and operation of governance systems.
3. Utilization of network maps as baseline data to better understand communities of scholars and practitioners who are or are not effectively engaged in EarthCube.
4. Utilization of planned applications of the stakeholder alignment tools and methods to mobile devices, allowing for distributed inputs on key issues from across the EarthCube community.

Additionally, the overall conceptual framework for stakeholder alignment represents the core dimensions to be addressed through governance processes. The conceptual framework involves three dimensions on which there might be stakeholder alignment (or misalignment), which are featured in the Figure 15:

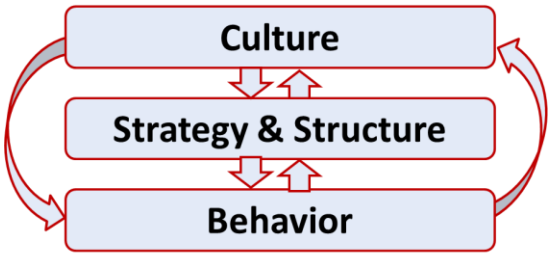

FIGURE 15. STAKEHOLDER ALIGNMENT MODEL[47]

Thus, systems of governance need to identify and advance certain behaviors, such as sharing information or seeking partners for collaboration. As well, there are structures (incentives, charters, etc.) that will be explicitly addressed through EarthCube governance and other structures, such as the rules governing tenure, that are beyond the reach of EarthCube, but relevant to take into account. The various public and private parties engaged in EarthCube will have different strategies and contrasting underlying cultures that could be more or less aligned.

Ultimately, the research on stakeholder alignment promises to help advance governance in a case where there is not an established institutional model and the capacity to govern and ultimately transform aspects of the Geosciences and Cyberinfrastructure will depend on these new, still emerging models for collaboration across stakeholders.

Network Map of EarthCube

The Governance workgroup will leverage the recent stakeholder alignment survey and additional, upcoming surveys, to put together a network map diagram of EarthCube and continue to evolve this picture as the group grows. The benefits of creating the network map include visualizing connections that are already occurring, providing one metric for community success as well as visualizing gaps in topics or key components. Further by providing the community a map, they can better place themselves in the EarthCube process, making the experience beneficial to their own work. This map is also useful for the EarthCube governance body to identify focal communities of concern, and identifying targets of opportunity, and is the first step in aligning stakeholders, as pictured in Figure 16 (below).

The network map of EarthCube derived from the surveys would be complemented and triangulated against other measures of interdisciplinary and cross-domain connections, such as a map of cross-

[47] Joel Cutcher-Gershenfeld, "Stakeholder Alignment," 2012.

domain data flows generated from inventories of geoscience models (e.g. http://maxim.ucsd.edu/crossdomain/ developed by the Cross-Domain Interoperability Concept Group.)

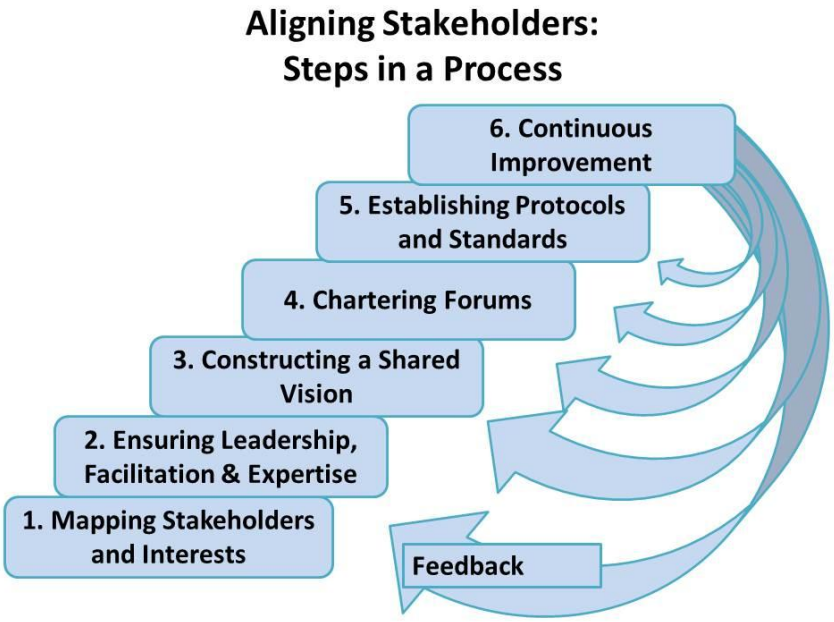

FIGURE 16. ALIGNING STAKEHOLDERS: STEPS IN A PROCESS[48]

This network map exercise foster development of key community stakeholder profiles. Stakeholder profiles are composite personas created from a collection of real stakeholder needs. These profiles will be used to identify the services that EarthCube needs to provide for the variety of different types of participants. The persona requirements will also begin to serve as baseline for future EarthCube evaluation.

4.2 MESSAGING

The second step in community engaging is messaging, focusing on the most appropriate means to disseminate information to EarthCube's diverse stakeholders. One way to formulate a messaging strategy is to encapsulate in a compelling vision the key drivers of motivations and outcomes for participation in collaborative activities: Resource allocation (i.e. R&D funding) and reward structures (WIIFM: "What's in it for me"). That messaging needs to be simple, capable of being communicated visually using an illustrative graphic, and accompanied by clear, concise language. There may be a small suite of messages (targeted at different communities or sub-communities), but all traceable back to some overarching messaging strategy that is informed by actionable intelligence (viewed along different axes, hence the data driven, map-oriented mechanism

[48] Joel Cutcher-Gershenfeld and Betty Barrett, "Lateral Alignment in Complex Systems: Valuing the Commons," PowerPoint Presentation, date unknown.

suggested above). EarthCube governance needs to be involved in the creation and evolution of the suite of messages.

Community Champions
Part of the messaging strategy is the identification of community champions, both in the domain arenas and IT fields, who would be willing to take the message to their respective communities. Such champions are often in heavy demand because of their acknowledged (often implicit) standing in those communities. For this reason, the messaging needs to be translated into materials (one-pagers, short PowerPoint slides, illustrative image resources in various resolutions for use in print [e.g. handouts, posters, etc.] and online media that are amenable for incorporation by the community champions in a media of their choosing.

Print Media
Our messaging strategy will review and target a number of peer reviewed journals in an effort to further disseminate our work on the EarthCube Governance Framework. Members of the Governance Group have already submitted an article to the American Geophysical Union (AGU)'s international Earth and space science newspaper *Eos,* summarizing the discussions and outcomes of the June 2012 EarthCube charrette from the perspective of the Governance Group. We will continue to seek appropriate opportunities for scholarly publication.

Social Media: Twitter and Linked-In
The governance group will also take an intentional approach and use social media where it amplifies the EarthCube message and connects stakeholders.

The group will maintain a twitter account where we share upcoming events, key resources and community member activities. The group will tweet from the community meetings that we attend over the next six months with highlights and observations. The twitter account will also maintain a twitter group of EarthCube-relevant tweeters, and promote the use of the #EarthCube tag.

As a way to build the professional network of scientist in EarthCube, the group will support the maintenance of the EarthCube LinkedIn group with a forum more focused on career challenges. The LinkedIn group will be one way of connecting young scientists with the EarthCube effort.

4.3 ENGAGEMENT

We recognize that each of the geoscience communities (e.g., earth science, ocean sciences, atmospheric sciences) and IT communities have their own histories, cultures, social norms, communication channels, and paradigms that help define them and offer challenges and opportunities for engagement. In generalized terms that may border on cliché, the earth science community has a large number of widely scattered individual researchers many of whom comprise the long-tail of scientists. In contrast, the atmospheric and ocean sciences communities have larger, centralized facilities due to the requirements to support expensive and limited satellite resources or research ships. The Governance Framework Steering Committee Members represent

these diverse communities, and are thus a key resource in building and engaging EarthCube stakeholders from these communities over the next few months.

As a scientific endeavor, EarthCube needs to embrace domain and computer scientists, and software engineers and to engage them in all development efforts. Target groups of our community engagement program include: Geoscientists, Atmospheric Scientists, Ocean Scientists, Computer Scientists, Software Developers, Information and Communications Technologies Communities, private industry, Federal and state governments, and the international community

Community engagement methods will vary slightly from one community to another, but will consist of some common concepts and events related to marketing EarthCube to stakeholders through virtual and in-person meetings and workshops, working with professional organizations, and gathering requirements for the EarthCube.org website. Communities that we will focus on engaging include the geosciences, atmosphere, and ocean sciences communities, IT and geoinformatics experts, software developers, the international community, early career scientists, and other domain scientists. A list of targeted community engagement events and a description of their community of interest is included at the end of the following section. The Community-Specific Engagement section, below, describes specific instances employing the engagement activities mentioned above, while tailoring these activities to target specific domain and computer science groups.

Most of the enterprise level services described earlier in this report are being built largely outside of the geosciences. The breadth of groups and organizations involved in cyberinfrastructure around the world is huge and growing quickly. It includes not only other domain sciences, but a tremendous amount of activity in the business-industry sector, and key developments internationally. While we propose focusing on engaging the core EarthCube stakeholder initially, we have to be aware that we are operating in a dynamic, robust, and innovative environment and we cannot work in isolation. Therefore, the engagement plan we outline here should be seen as carrying out two primary functions, one is to engage the primary EarthCube stakeholders, and two is to develop materials and processes that will foster EarthCube in the broader global cyberinfrastructure.

4.3.1 COMMUNITY-SPECIFIC ENGAGEMENT

Engaging the Atmospheric Sciences Community

We plan to employ all avenues in our ongoing engagement of the atmospheric sciences community. A few specific examples of past engagement activities are provided below. We will continue to engage the atmospheric sciences community through similar mechanisms, evaluating the results and breadth of engagement after each event.

The atmospheric science community in academia was informed about EarthCube during the 2012 Unidata Users Workshop, July 9-13, 2012. The theme of this workshop, Navigating Earth System Science Data, was highly pertinent to EarthCube, so EarthCube topics were integrated into the agenda. Presentations on EarthCube and related topics were made by Cliff Jacobs (NSF), Michael

Morgan (Atmospheric and Geospace Sciences Director), and many others. Twenty-six presenters from the Unidata community shared their insights on doing science in an environment of expanding data availability with the nearly 100 workshop attendees. The talks ranged from high-level descriptions of big initiatives like EarthCube, Global modeling at NOAA, and the joint NOAA-NASA GOES-R satellite program to hands-on demonstrations of data analysis tools including Python, GrADS, Unidata's Integrated Data Viewer, and the still-in-development AWIPS-II analysis and display system.

An NSF-sponsored meeting, focused on how the real-time modeling community in the atmospheric sciences can help shape and benefit from EarthCube, is currently in the planning stages. These plans call for inviting about 30 scientists to this workshop that will be held this fall, most likely in Boulder, CO.

In addition to the above workshops, the upcoming American Meteorological Society (AMS) Annual Meeting, which will be held 6-10 January, 2013 in Austin, TX, provides another opportunity to engage the Atmospheric Sciences community as over 3,000 atmospheric scientists and other professionals in related fields attend this meeting each year. Sessions on data stewardship and related topics are planned. The Townhall meetings, which are traditionally held at the AMS Annual Meeting, also offer an excellent forum for community engagement and gathering feedback on EarthCube because the American Meteorological Society "promotes the development and dissemination of information and education on the atmospheric and related oceanic and hydrologic sciences and the advancement of their professional applications."[49]

Engagement of specific atmospheric agencies, such as the National Oceanic and Atmospheric Administration (NOAA), should be explored. NOAA engagement groups, such as the Environmental Data Committee[50] the GIS committee[51] and the Data Management Integration Team[52] might be a good place to begin. In addition, multiple members of our Steering Committee are NOAA staff and will be able to provide further insight in engaging the federal agency. Additional information on engaging federal agencies is included in a following section.

Engaging the Geosciences Community

There are a few large national professional meetings in the earth sciences that provide excellent opportunities for presenting technical and conceptual progress, and where much of the 'business' of the community is conducted (e.g., committee meetings, forums, organizational business meetings, etc.).

The American Geophysical Union has become a center for informatics activities in the broader geosciences, with the Earth and Space Science Informatics Focus Group (equal to a section) being

[49] The American Meteorological Society, "About the AMS," http://ametsoc.org/aboutams/index.html, last modified May 2012.
[50] See https://www.nosc.noaa.gov/EDMC/
[51] See http://www.cio.noaa.gov/IT_Groups/noaa_cio_GIS_Committee.html
[52] See https://www.nosc.noaa.gov/EDMC/swg.php

the fastest growing group in the 55,000 member organization. The annual fall meeting of AGU is now hosting two dozen or more informatics technical sessions. The weekly newspaper, *Eos*, is an excellent source of informatics news and articles. Many of the NSF-supported informatics projects have booths in the Exhibit Hall as part of "NSF Alley." This is one of the more important venues for sharing technical results, networking, and negotiating collaborations. The Fall Meeting is attracting over 18,000 attendees offering a tremendous platform for community engagement.

Similarly, the European Geophysical Union annual meeting has grown to over 15,000 attendees with strong global participation. It too has a strong informatics component and serves as a prime venue for European Union INSPIRE results to be presented.

The Geological Society of America established a Geoinformatics Section early on and continues as a strong forum for community building activities. The umbrella organization, American Geosciences Institute (AGI) typically holds meetings with its 46 member societies in conjunction with GSA annual meetings, offering opportunities to engage them in policy level discussions and seek ways to inform their members of EarthCube developments and opportunities.
Many earth scientists work in the private sector. Historically they have worked in the petroleum and mining industries, more now in the environmental and water fields. These sectors are not traditionally part of the NSF core constituency. Confidentiality of data is a major factor for much of these sub-communities as well. Also, much of their involvement with cyberinfrastructure is more likely with proprietary commercial products and services. These factors require addressing different issues that we elaborate on in the section on Business and Industry below.

Engaging the Ocean Sciences Communities
Our goals for interaction will be twofold; first to inform the leadership and members of these organizations and research collaborations about the general activities of EarthCube and about the specific development goals of the EarthCube Governance community; secondly to engage these communities (through open dialogue and surveys instruments) in providing their input into the evolution of an EarthCube Governance framework.

Coastal Oceanographers will be engaged through interactions with the following professional organizations; the American Geophysical Union (AGU), the American Meteorological Society (AMS), The Oceanography Society (TOS) and the Consortia for Ocean Leadership. In addition to these large professional organizations we will target a handful of focused research collaborations such as the SURA Coastal Ocean Modeling Testbed.

Engagement of the Blue Water Ocean Sciences communities will be facilitated through interaction with three ocean-related initiatives:

- **The Integrated Ocean Observing System (IOOS).** IOOS is the US Contribution to the Global Ocean Observing System, (GOOS) and the Oceans elements of the US Integrated Earth Observation Systems (IEOS). Many aspects of EarthCube align with the Data Management and Communications (DMAC) subsystem of IOOS.

- **Ocean Observatory Initiative (OOI) –** OOI is an NSF-sponsored program to transform ocean research and is the NSF contribution to IOOS. OOI has its own CI which will become part of EarthCube.
- **Centers for Ocean Science Education Excellence (COSEE).** COSEE is an NSF-sponsored initiative to improve the ocean literacy of students and teachers nationwide. It specializes in connecting teachers with active research scientists and on supporting the development of ocean-related science curriculum requirements. COSEE is an important element of addressing the broader impact and engaging the next generation of scientists within the oceans community.

Engaging Computer Science and Software Developers

To engage the software community, connections with groups such as the Federation of Earth Science Information Partners (ESIP), the Boulder Earth and Space Science Informatics Group, Unidata, the Open Geospatial Consortium (OGC), the Association for Computing Machinery Special Interest Group on Software Engineering, the Institute of Electrical and Electronics Engineers (IEEE) will be pursued. Conferences, including the American Meteorology Society Annual Meeting and the American Geophysical Union Meeting and the International Conference on Software Engineering have sessions focusing on information systems, and can also be leveraged to engage workshop participants in EarthCube and expand the EarthCube community.

The possibility of a workshop in collaboration with either the winter or summer 2013 ESIP meetings will also be explored. This workshop could include a hack-a-thon similar to the ones held by the Brokering Concept Award or something similar to the NASA competition to develop space applications to create enthusiasm in the community. Additionally, we could model the workshop after Google's Summer of Code program to engage computer science students and promote awareness of the EarthCube project while encouraging the development of EarthCube applications.

Another outreach technique currently being tested is the seeding of a Software Engineering Community of Practice (CoP) that supports and facilitates the achievement of EarthCube science goals. If successfully initiated, this CoP will be tasked with assisting EarthCube with the orchestration and deployment of software resources, including experts, and software and domain-specific best-practice guidance for EarthCube projects during their life-cycles. The EarthCube Special Interest Group focusing on this CoP realizes that success depends on being able to demonstrate in a practical way the benefits of this approach and hopes to do so through various pilot projects developed under the guidance of EarthCube governance. The EarthCube Governance Group will leverage the Software Engineering CoP to reach out to software developers and engage them in the development of EarthCube.

There are groups like Data Management (DAMA) International which has explicit goals to:

- Help practitioners become more knowledgeable and skilled in the information and data management profession
- Influence practices, education and certification in the information and data management profession

- Form alliances with other organizations with similar principles to strengthen the profession

These goals are compliant and supportive of EarthCube and augur well for collaboration.

Engaging the Information and Communications Technologies (ICT) Standards Communities
EarthCube's goal of "supporting the development of community-guided cyberinfrastructure to integrate data and information for knowledge management across the Geosciences" will require coordinating with a complex set of organizations responsible for creating and evolving standards for Information and Communications Technologies (ICT).

The communities involved in ICT standards are segmented. Some are based on professional and technical disciplines (e.g., meteorology, hydrology, climate, biodiversity, neuroinformatics, geospatial information, and many more), such as the World Meteorological Organization (WMO) and Open Geospatial Consortium (OGC). Underlying these are a tier of standards development organizations (SDO), such as Institute for Electrical and Electronics Engineers (IEEE), and the World Wide Web Consortium (W3C). Within each SDO there are typically multiple distinct working groups (WG), sometimes called Technical Committees (TC), to address specific technologies.

EarthCube will not be a standards development organization. However, it should identify and promote use of ICT standards that would contribute to its goal of data and model integration for knowledge management across the Geosciences. Furthermore, EarthCube should seek to inform and influence the standards on which this goal is dependent.

Engagement with these ICT communities will require identifying and working with "champions" within each SDO/WG. Many Geosciences researchers are already active in various SDOs, and these roles should be recognized and coordinated. The OGC has a number of WGs dedicated for Geosciences standards development and outreach. ESIP, while not an SDO, organizes "clusters" of discussion topics which provide fertile ground for interdisciplinary coordination and outreach on standards and best practices.

Much could be done to improve ICT standards community engagement in EarthCube. ESIP holds two meetings each year (summer and winter), which are already being used for various EarthCube-related discussions, and could be expanded. OGC holds quarterly meetings around the world, which could also be used for EarthCube discussions. In OGC, the ESS DWG and the University DWG both serve as points of information sharing across the Geosciences and other academic disciplines represented in OGC, and would encourage EarthCube engagement. Through OGC's relationships with many other SDOs [26], it can also provide communication channels and facilitation for EarthCube's interests in various other standards bodies.

Besides coordination between EarthCube and SDOs, considerable outreach will be needed to engage the broad EarthCube user base, to encourage use of ICT standards, and recruit new talent for maintaining and evolving the standards needed. EarthCube's portal should make it easy to discover and learn about relevant standards, and NSF grants could eventually encourage this

discovery by requiring attention to EarthCube-designated standards as they now do with data management plans. One other tier of standards-related bodies that would be important for engagement with EarthCube include the various committees that have formed within NASA, NOAA, GEOSS and others, to identify and promote relevant standards across that organization.

Engaging Business and Industry

A study by McKinsey Global Institute found that 15 of 17 sectors in the United States have more data stored per company than the US Library of Congress. They conclude that big data (which we view as a surrogate for cyberinfrastructure) "have swept into every industry and business function are now an important factor of production." [53] The McKinsey study predicts that "use of big data will become a key basis of competition and growth for individual firms," [54] although challenges exist in data security, intellectual property rights, disruption of established organizations.

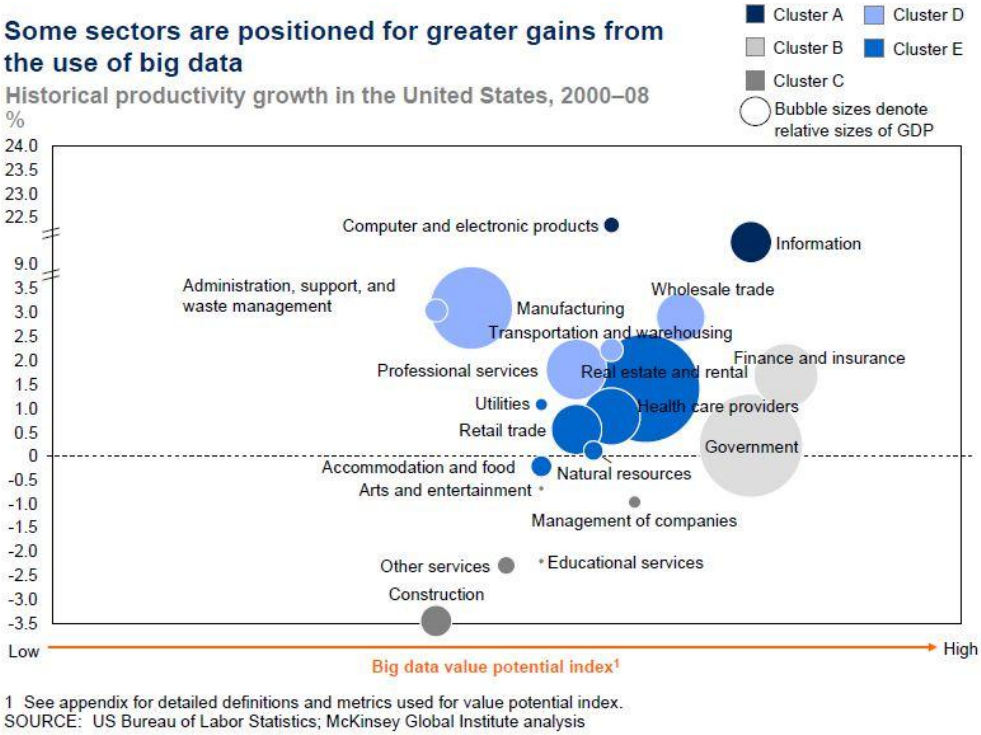

FIGURE 17. HISTORICAL PRODUCTIVITY GROWTH IN THE UNITED STATES, 2000-2008[55]

These conclusions suggest that the private sector overall should be strongly supportive of the EarthCube initiative as enabling technology. It's also evident that the business-industry activities in geoscience cyberinfrastructure, as seen in productivity growth by sector, comprise only a small part of the overall field. While not addressing the largest segments of the overall

[53] James Manyika et al., *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute, 2011, 156p.
[54] Ibid.
[55] Manyika et al., *Big Data*, 9.

cyberinfrastructure, EarthCube offers the advantage of potentially serving as the prototype for an integrated and cutting-edge approach. This should attract attention and participation from a broader swath of industry.

Engagement of the Private Sector will be implemented in multiple ways:

1. The private sector should be encouraged to participate in all EarthCube related solicitations. Thus, there should be no explicit barriers to private sector participation be included in any solicitations (e.g., RFP, Dear Colleague Letters, etc.).
2. EarthCube will reach out to organizations such as the Alliance for Earth Observations, the Alliance for Coastal Technologies, Standards Leadership Council, etc. to keep them abreast of EarthCube milestones and potential opportunities.
3. Individuals in the private sector can be engaged via their participation in professional societies and activities.
4. Promoting ways the private sector can use EarthCube for their benefit (and thus helping to institutionalize it) and create products and services of value to EarthCube contributors and end users.

> "The effective use of big data has the potential to transform economies, delivering a new wave of productivity growth and consumer surplus. Using big data will become a key basis of competition for existing companies, and will create new competitors who are able to attract employees that have the critical skills for a big data world."
>
> James Manyika et al., *Big Data: The Next Frontier for Innovation, Competition, and Productivity, 13.*

The rationale for academic involvement and private sector involvement are predicted on two different end-states and objectives. This difference must be recognized and explicitly addressed if the benefit of public/private partnerships is to become a reality.
The private sector is among the biggest beneficiaries of the Internet and World Wide Web as infrastructure and we see them realizing the opportunities and resources rising out of EarthCube and other cyberinfrastructure developments. Entrepreneurial start-up companies as well as major established corporations should envision EarthCube as a business enabler.

Engaging Government

<u>Engaging Government – Federal</u>

Engagement of Federal agencies will be critical, since many of these agencies will be both a supplier and consumer of EarthCube data. Since there are many different organizations, engagement will need to be done through venues where several of these organizations are involved. Several recent federal government initiatives, such as the Big Data and Digital Government initiatives, the National Geospatial Platform Initiative and NOAA's Geospatial Platform Prototype are similar to EarthCube and it would be useful to engage those involved with these initiatives perhaps through the White House Office of Science and Technology Policy and National Science and Technology Council.

The U.S. Geological Survey, as described above, is making cyberinfrastructure development and deployment a central focus for the entire agency. They have been engaged in EarthCube from the

beginning as observers and contributors. The new 10-year strategy for Core Science Systems appears to be fully compatible with and supportive of the EarthCube vision.

The need for cyberinfrastructure is evident throughout the federal government and across agencies, as evidenced by a plethora of programs aimed at achieving the same list of capabilities that have come forward from the EarthCube community roadmaps. One of EarthCube's goals must to be to inform these disparate entities of the advantages of collaborating and leveraging resources and converging on standards, protocols, and communities of practice.

Engaging Decision-Makers

Engagement with decision-makers needs to be carefully designed to be purely informative. The ultimate objective would be the recognition of concepts like "well documented data", "discoverable data and information", "attribution of data publications", etc. In an effort to educate decision-makers, we will work with professional societies and non-profit groups to organize briefings for staffers to learn about data and resource management. The message should be how interoperability, enabled through something like EarthCube, enables decision-makers to craft informed decisions using relevant data and information.

> "The promise of big data and the potential economic growth opportunities are significant. To actually reach that potential, big data initiatives must be governed and managed like any other data and information asset through formal data management within state government enterprise architecture."
>
> Jack Doane, co-chair for the NASCIO Enterprise Architecture and Governance Committee, "Is Big Data a Big Deal for State Government?" August 23, 2012

A report by McKinsey Global Institute[56] addressed implications of "big data" on policy makers and is applicable to the broader realm of cyberinfrastructure. They identified five key areas that we paraphrase for EarthCube:

1. Build human capital
2. Align incentives to promote cyberinfrastructure for the greater good
3. Develop policies that balance the interests of companies wanting to create value from data and citizens wanting to protect their privacy and security
4. Establish effective intellectual property frameworks to ensure innovation
5. Address technology barriers and accelerate R&D in targeted areas

Another possible strategy is to focus on how interoperability, enabled through something like EarthCube, allows responders and decision-makers to craft informed decisions using relevant data and information. The recently released National Academy of Sciences report on "Disaster

[56] James Maniyaka et. al, *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute, 2011, 156p.

Resilience" identifies as one of four major recommendations, the need for data acquisition, discovery, access, and interoperability across myriad federal agencies, universities, and others, to "Collect, analyze, and communicate data, forecasts, and models about risk, hazards, and disasters."[57] The biggest challenge to this is seen as Insufficient or dispersed data sets; understanding how to share scientific information with broad audiences."

Engaging State Governments

Some of the key stakeholders at the state government level participate in national associations that are already engaged in cyberinfrastructure activities or provide enhanced channels for communications.

The Association of American State Geologists is partnered with the U.S. Geological Survey in developing the US Geoscience Information Network, focusing on creating an operational national federated interoperable data network. This framework was adopted by the US Department of Energy for the National Geothermal Data System.

The National Association of State Chief Information Officers (NASCIO) is "the premier network and resource for state CIOs and a leading advocate for technology policy at all levels of government."[58]

The National States Geographic Information Council (NSGIC) deals with adoption of geospatial information technologies at the state and national level. "NSGIC membership includes nationally and internationally recognized experts in geospatial information technologies, data creation and management as well as information technology policy." NASCIO's latest issue brief on analytics, "Is Big Data a Big Deal?" is available at www.nascio.org/publications.

Similar associations exist for state environmental protection, water resource, natural resource, and technology officials.

Engaging the International Community

We have already begun engaging elements of the international community with positive and encouraging responses. However, there is a long tail of capabilities in the international realm that will be as challenging if not more so, to involve as the domestic stakeholders.

As with other components, there are scores of groups and activities that are making significant contributions to global cyberinfrastructure. We offer a few examples of groups we are currently working with, to provide a sense of how EarthCube is viewed around the world and the potential for further engagement. Our initial assessment is that opportunities are huge to the point of overwhelming.

---

[57] Committee on increasing National Resilience to Hazards and Disasters; Committee on Science, Engineering, and Public Policy; the National Academies, *Disaster Resilience: A National Imperative*, 2012.

[58] Vaughn, Shawn, "NASCIO Welcomes the Appointment of FirstNet Board Members," http://www.nascio.org/newsroom/pressrelease.cfm?id=131, last modified August 20, 2012.

The European Union's INSPIRE (Infrastructure for Spatial Information in Europe) directive is a continent-wide effort on a scale of EarthCube. INSPIRE aims to create a European Union (EU) spatial data infrastructure, to "enable the sharing of environmental spatial information among public sector organisations and better facilitate public access to spatial information across Europe." The intent is that a European Spatial Data Infrastructure will assist in policy-making across boundaries. Therefore the spatial information considered under the directive is extensive and includes a great variety of topical and technical themes. OneGeology-Europe is a consortium of 22 nations and additional partners in the geosciences developing interoperability across the continent under the INSPIRE directive as the flagship geoscience effort. EarthCube Governance principals have kept the OneGeology-Europe project members briefed on EarthCube concepts and developments. The European geological surveys are carrying out a number of more specialized projects and were briefed and engaged in discussion at the International Geological Congress in Brisbane, Australia in August 2012.

The international OneGeology Operational Management Group also reviewed EarthCube at their annual meeting also at IGC. This resulted in two abstracts submitted on cyberinfrastructure governance at the American Geophysical Union Fall Meeting, organized by EarthCube governance members.

Presentations and discussions with Australian government officials and attendees at the International Geological Congress in Brisbane, in August 2012 demonstrated tremendous interest in EarthCube by the international community and desire to develop more formal collaborations.

One upcoming venue that offers intriguing potential is the National Data Repository (NDR11) meeting in Kuala Lumpur, Malaysia. More than 30 nations with national repositories of data and samples mostly covering energy and mineral resources, are gathering to promote global standards and capabilities. The US is one of the few major natural resource countries that do not maintain centralized facilities. Most data from resource development in the US is handled at the state level. Energistics, the non-profit consortium for standards in the upstream petroleum industry, organizes the NDR meetings. Energistics is following EarthCube closely. They recently helped coordinate the agreement among 8 standards organizations to form the Standards Leadership Council to identify areas of intersection to avoid creating duplicate or conflicting standards and opportunities for synergy. This will be an excellent opportunity for feedback from industry and other nations on the EarthCube vision and strategies.

4.3.2 COMMUNITY ENGAGEMENT METHODS

Marketing EarthCube to Stakeholders

We will initially develop some EarthCube marketing material to explain the EarthCube vision. The material will be tailored to the personas and delivered over the next six months at a preliminary list of meetings, conferences, and workshops for presenting the EarthCube Governance Framework draft to ensure broad based community feedback.

The group will work with NSF to write a summary of EarthCube and create a brief slide deck introduction. For the website the slide deck will be turned into a brief screencast on EarthCube that can be shared. The material will be easily modifiable for a given community. The resource pool that evolves will be stored first on the EarthCube Ning site, and eventually on the earthcube.org site once it is developed. The Governance Committee could work with the other Concept Teams and Working Groups to create short introductions to their groups. This will be needed for the website and should be updated regularly with the current projects/activities, or new groups, and ways to get involved.

Virtual and In-Person Meetings and Workshops

The Governance team will conduct virtual meetings by purchasing a WebEx license with 10 seats to support the current EarthCube Working Groups and Concept Teams. The WebEx will be branded as EarthCube. The WebEx recordings will be available after the meetings for others to listen and learn from on the existing EarthCube Vimeo site. WebEx will be made available for all routine telecons as well as for adding remote participation to face-to-face events. The Governance Group will work to train the other groups on how to use WebEx to ensure the barrier of usage is very low.

The Governance Group will also maintain a Google calendar of key events. The calendar will be editable by anyone in the community and will allow the tracking of events of interest. WebEx info for EarthCube events will be posted to the calendar.

Several in-person workshops and conferences have also been identified as ideal venues to engage the domain science, information science, and IT communities about EarthCube. One of the key outcomes from these meetings of opportunities will be to highlight bright spots, where communities are already leveraging cyberinfrastructure extensively (e.g. OOI, DataOne). Most of these events are events of opportunity where one or more Steering Committee members are attending and have either submitted abstracts for presentations or have been invited to present. This is not an exhaustive list, and we are currently working on identifying additional conferences and workshops to engage a broader cross-section of domain, IT, information scientists, and others.

1. Federation of Earth Science Information Partners Mid-Year Meeting, Madison, WI, July 17-20; *Target Community: Geoscientists and IT Experts*
2. 34th International Geological Congress, Brisbane, Australia, August 5-10, 2012 *Target Community: Geoscientists and the International Community*
3. Spatial Computing Visioning Workshop, Washington, DC, September 10-11, 2012; *Target Community: Software and Cyberinfrastructure Experts*
4. American Geological Institute Leadership Forum, Washington, DC, September 11, 2012 panel discussion on Big Data; *Target Community: leaders of 46 geoscience professional societies*
5. Early Career Scientist Workshop at the Carnegie Institution, proposed October 16-17, 2012 in Washington, DC in collaboration with Joel Cutcher-Gershenfeld, Robert Hazen, and Steve Diggs; *Target Community: Early Career Scientists Across Domains*

6. National Ecological Observatory Network (NEON) Community Workshop, October 17-18, 2012 in Washington, DC; *Target Community*: *Ecologists*
7. EC Governance Steering Committee Meeting held in conjunction with the October 16-18 workshops: *Goal:* review current Framework feedback and adjust the framework as necessary.
8. National Data Repositories (NDR 11), Kuala Lumpur, Malaysia, October 21-24; *Target Community: Data and Sample Repository Managers from 30+ nations*
9. ASIS&T Annual Meeting 2012 in Baltimore, MD, October 26-30; *Target Community: Information Scientists*
10. Unidata/NCAR/UCAR Workshop for Atmospheric Science, Boulder, CO, Date TBD; *Target Community: Atmospheric Scientists*
11. Geological Society of America in Charlotte, NC, November 4-7; *Target Community: Geoscientists*
12. American Geophysical Union in San Francisco, CA, December 3-7 (Steering Committee members convening Union session on Convergence in Cyberinfrastructure, and technical session on Governance in Cyberinfrastructure); *Target Community*: *Geoscientists*
13. EarthCube Governance Steering Committee Meeting held after AGU. *Goal:* Finalize any revisions based on the previous four months of aggressive community engagement and outreach prior to submitting a revised Governance Framework to NSF.
14. ESIP Winter Meeting in Washington, DC, January 5-8, 2013; *Target Community: Earth and Space Scientists*
15. 93rd Annual American Meteorological Society Meeting, Austin, Texas, January 6-10, 2013, *Target Community: Atmospheric Scientists*
16. Open Geospatial Consortium Meeting at Esri in Redlands, CA, January 14-18, 2013; *Target Community: Geospatial Standards Developers*
17. iConference 2013 in Fort Worth, TX, February 12-15, 2013; *Target Community: Information Scientists*

Gathering Website Requirements

The website is a crucial part of the EarthCube community engagement and outreach. The site will provide organizational memory and transparency through documenting the activities as they happen. As described above, one key activity in community engagement is identifying personas since the earthcube.org site will support a spectrum of users from novice to expert, domain scientist to computer scientist.

First the group will review the requirements that have already been gathered by the NSF team and assess the work done in the first round of funding on the EarthCube Ning site.

From those requirements the group will identify an initial set of core functions needed for the earthcube.org site and set-up a rapid prototype, where possible leveraging the earthcube.ning.com site and extending as additional functionality requires. The Governance team will use these tools themselves and share this at the community meetings. Based on feedback the prototype and the requirements gathered, the Governance team will map out a suggested plan for migration and integration of the earthcube.ning.com site to the permanent earthcube.org site.

SECTION 5.0: COMMUNITY BUILDING AND ENGAGEMENT WORK PLAN

Our community building and engagement program will be an iterative process. Many activities will be taking place in concurrently. Each element will be incorporated into this Framework document as feedback is received.

5.1 GENERATE ACTIONABLE INTELLIGENCE

1. **Leverage the EarthCube Stakeholder Alignment Survey to identify gaps in the current EarthCube stakeholders, if any.**
    a. Identify individuals to act as liaisons between Dr. Joel Cutcher-Gershenfeld (Stakeholder Alignment Survey Principle Investigator) and the Governance Working Group
    b. Work with Dr. Cutcher-Gershenfeld as new survey results extend the list of stakeholders identified thus far:
        i. Domain scientists
            1. Atmospheric or Space Weather scientist
            2. Oceanographer
            3. Geologist
            4. Geophysicist
            5. Hydrologist
            6. Critical zone scientist
            7. Climate scientist
            8. Biologist or Ecosystems scientist
            9. Geographers
        ii. Social scientist
            1. Anthropologist
            2. Economist
            3. Psychologist
            4. Sociologist
        iii. Computer or Cyberinfrastructure scientists
            1. Data manager
            2. High performance computing expert
            3. Software engineer
            4. IT user support personnel
        iv. Other Groups
            1. K-12 educator
            2. Designer/developer of geoscience instrumentation
            3. Environmental resource manager
            4. Policy makers

2. **Leverage the EarthCube Stakeholder Alignment Survey to gather community requirements on EarthCube governance**
    a. Identify individuals to act as liaisons between Dr. Joel Cutcher-Gershenfeld (Stakeholder Alignment Survey Principle Investigator) and the Governance Working Group.
    b. Use current data from the recent stakeholder survey (with approximately 700 responses) addressing issues on access and use of data, preferences for leadership to be community elected versus NSF selected, issues of trust and transparency, and related matters.
    c. Introduce governance-related questions in future planned stakeholder surveys to provide additional inputs into the development and operation of governance systems.
    d. Utilize planned applications of the stakeholder alignment tools and methods to mobile devices, allowing for distributed inputs on key issues from across the EarthCube community.

3. **Create an EarthCube network map of EarthCube stakeholders**
    a. Identify Governance Working Group individuals responsible for creating a network map.
    b. Create network map to visualize connections that are already occurring among EarthCube stakeholders.
    c. Upload EarthCube network map to the EarthCube website.
    d. Utilize network maps as baseline data to better understand communities of scholars and practitioners who are or are not effectively engaged in EarthCube.

4. **Develop EarthCube stakeholder profiles**
    a. Identify Governance Working Group individuals responsible for creating a stakeholder profiles.
    b. Create composite personas from a collection of real stakeholder needs
    c. Use these profiles to identify services that EarthCube needs to provide for the variety of different types of participants.
    d. Upload persona profiles to the EarthCube website.
    e. Use these persona requirements as a baseline for future EarthCube evaluation.

5.2 DEVELOP A MESSAGING STRATEGY

1. **Determine EarthCube messaging strategies tuned to different stakeholder groups identified in Part 1.**
    a. Encapsulate in a compelling vision the key drivers of motivations and outcomes for participation in collaborative activities:
        i. Resource allocation (i.e. R&D funding)

        ii. Reward structures (WIIFM: "What's in it for me")
    b. Develop messages that are be simple, capable of being communicated visually using an illustrative graphic, and accompanied by clear, concise language.
        i. Develop a small suite of messages (targeted at different communities or sub-communities), but all traceable back to some overarching messaging strategy
        ii. Make sure messaging strategy is informed by the Actionable Intelligence gathered from Part 1 above

2. **Create marketing material to explain the EarthCube vision and introduce governance**
    a. Identify individuals responsible for marketing EarthCube and developing marketing materials
    b. Develop general templates for the broader community
    c. Tailor these templates to different target communities
        i. Develop written materials
        ii. Develop slide decks
            1. Turn the slide decks into screencasts that can be shared
    d. Identify appropriate web forums to upload EarthCube marketing materials
    e. Create repositories to store these marketing materials on the EarthCube website

3. **Work with EarthCube Working Groups and Concept Teams to create short introductions to their groups, to be included in EarthCube marketing materials**
    a. Build upon the existing Governance Working Group liaisons with other EarthCube Working Groups and Concept teams to:
        i. Gather information to be included in the introductions
        ii. Write introductions and include them in EarthCube marketing materials
    b. Gather EarthCube Working Group and Concept Team feedback

4. **Use online forums to market EarthCube**
    a. Identify individuals from the Governance Working Group to identify, create (if necessary) and maintain online forums appropriate for marketing EarthCube, such as wikis.
    b. Create and maintain online forums
    c. Gather community feedback

5. **Identify community champions**
    a. Identify community champions in the domain, computer, information, and social sciences communities
    b. Identify individuals from the Governance Working Group to liaise with community champions to support them in engaging their respective communities
    c. Provide community champions with EarthCube marketing materials described above

d. Gather feedback from community champions and from their respective communities

6. **Publish articles in scholarly journals and trade publications**
    a. Identify appropriate scholarly journals and trade publications to submit EarthCube articles, target three scholarly articles
    b. Assign the appropriate author(s) from the Governance Steering Committee
    c. Write the article, gathering feedback from the Governance Steering Committee and Governance Forum
    d. Submit for publication

5.3 ENGAGE EARTHCUBE STAKEHOLDERS

1. **Engage EarthCube stakeholders via social media**
    a. Identify individuals responsible for conducting community engagement via social media
    b. Identify appropriate social media outlets to target different EarthCube stakeholders
    c. Develop and implement engagement strategies tailored to each social media outlet that amplify the EarthCube message and connect to stakeholders.
        i. Create and maintain an EarthCube LinkedIn group targeted to connecting young scientists with EarthCube
    d. Gather community feedback

2. **Engage EarthCube stakeholders via virtual meetings and workshops**
    a. Identify target groups that are not covered by in—person meetings and workshops
    b. Develop virtual webinars, incorporating the EarthCube marketing materials detailed above
        i. Determine the purpose of each webinar
            1. What are the goals of the webinar?
            2. What is the target audience?
            3. What information do we want to obtain?
    c. Schedule webinars
        i. In addition to stand-alone webinars, we propose to schedule a series of webinars to immediately follow, and thereby complement, workshops and presentations held at in-person conferences
    d. Invite participants
    e. Hold webinars
    f. Gather community feedback

3. **Engage via in-person meetings and workshops**
    a. Identify meetings of opportunity and their respective target communities

b. Identify Governance Working Group members who are able to attend
c. Submit EarthCube abstracts
d. Attend workshops
e. Gather community feedback

4. **Gather EarthCube website requirements.**
    a. Identify individual(s) from the Governance Working Group to oversee the website requirements-gathering process
    b. Develop a survey to determine website requirements
    c. Make survey available to EarthCube community and interested stakeholders
    d. Compile and analyze results
    e. Gather additional community feedback from social medial, virtual and in-person meetings and workshops on website requirements

5.4 ANALYSIS AND SYNTHESIS

1. **Analyze and synthesize all input gathered from community engagement activities.**

2. **Incorporate this information into an updated Governance Framework document, to be released in early 2013.**

CONCLUSION

This document serves as a summary of our initial Governance Framework recommendations to be vetted by EarthCube stakeholders. The Framework, and corresponding community outreach, will maximize engagement of the broader EarthCube community, which in turn will minimize the risks that the community will not adopt EarthCube in its development and final states. The target community includes domain, information and computer scientists in academia, government, and the private-sector, both nationally and internationally.

The Governance Framework presented in this document was produced via the initial implementation of action items 1 and 2 from the Governance Roadmap, presented to NSF and the EarthCube community in June 2012. Our goal in developing an initial governance framework is to provide the foundation from which a governing entity can establish methods, processes, and procedures to carry out the governance functions and guiding principles, keeping in mind the governance recommendations, presented in this document.

This Governance Framework document did not recommend a specific governance model for EarthCube. Instead, it defined a set of governance functions we feel need to be carried out, based on our research on governance and our community engagement to-date with EarthCube stakeholders and other interested parties, and it identified a set of guiding principles and governance recommendations regarding EarthCube governance in the longer term.

We specifically chose not to recommend any individual governance model, or set of models, for EarthCube, because we feel that it is too early in the process to make this recommendation. Instead, we feel that the group responsible for carrying out the initial governance functions will have to define the governance model they believe is the most effective way for carrying EarthCube forward. It will be up to the organizations selected for the prototype EarthCube governance to identify what governance models they feel would be most effective to carry out functions based on guiding principles, while also meeting the goals of the community and NSF.

Key Findings

Based on eight months of research on governance theory and models, historical infrastructure case studies, and initial community engagement activities with Earth scientists, industry, government agency and international representatives, we compiled and synthesized system-wide governance requirements to draft an initial set of EarthCube governance recommendations, governance functions and guiding principles. These functions, guiding principles and recommendations form the basis of an initial Governance Framework that will be revised and updated as part of an aggressive community outreach and engagement program. We plan to incorporate additional community input and release a modified version of the Governance Framework at the beginning of 2013. A summary of our initial findings is presented below and a complete table of the functions can be found in Section 3.1.

Initial EarthCube Governance Functions

1. Identify and Implement an EarthCube Vision, Mission, and Goals.
2. Engage and Coordinate Across the EarthCube Community.
3. Management: Create and Implement EarthCube By-Laws and Charter.
4. Develop and Maintain a Viable Architecture and Concept of Operations that Enable the Realization of the Goals and Objectives of the EarthCube Vision.

EarthCube Governance Guiding Principles

1. EarthCube governance shall serve the advancement of interdisciplinary science through collaboration among community members and with other cyberinfrastructure initiatives.
2. EarthCube governance shall rely on open, transparent processes and shall vet and inform its decisions through active community engagement
3. EarthCube governance shall encourage environmentally sustainable processes and practices.
4. EarthCube governance shall support development that draws from best practices based on interoperability and reuse of resources.
5. EarthCube governance shall strive for the free and open sharing of data, information, software and services.
6. EarthCube governance shall evolve with changing technologies and user needs while remaining robust.

General Governance Recommendations

1. An interim prototype governance committee should be established in the mid-term (starting early 2013) to carry out Steps 3 and 4 of the EarthCube Governance Roadmap (establish and implement Terms of Reference for EarthCube Governance), according to the findings presented in this Governance Framework document.
2. During the building of EarthCube, we recommend an umbrella entity to coordinate governance functions across EarthCube and facilitate the optimal interoperability of existing cyberinfrastructure into EarthCube.
3. During the build phase of EarthCube, we recommend EarthCube governance to focus on the development, operation and maintenance of EarthCube-wide services and touchpoints (services that enable existing cyberinfrastructure components to interoperate effectively within EarthCube). The existing systems that are being interconnected would retain their local autonomy, continue to determine their own science priorities, and continue to serve their constituent communities. EarthCube would serve as an enabler of cross-disciplinary activity.
4. As part of the development of EarthCube-wide services and touch points, EarthCube governance would develop and maintain an effective enterprise-level architecture, proactively identify and manage enterprise-level risks, support solicitations for risk reduction prototypes, and support solicitations for cross-domain testbeds.

5. As it matures, EarthCube will become responsible for discovering and encouraging the development of a set of cross-domain science priorities with the goal of creating new cross-disciplinary research communities within EarthCube.
6. EarthCube will function as a community service provider, not a competitor to infrastructure, communities of practice, organizations, institutions and consortiums.

Defining EarthCube

Although EarthCube is still in the process of being defined, we believe it is possible to take concrete steps in moving EarthCube governance forward. In fact, the goal of the initial EarthCube Governance Framework presented in this document and the concurrent community engagement program is to further refine community requirements for EarthCube, thereby helping to give EarthCube more definition.

Defining Governance

Based on our initial community engagement, we've realized there are a number of definitions regarding governance. Governance of EarthCube will have to take in account all of the interpretations of governance brought forward by EarthCube stakeholders, in particular as additional individuals are brought into the EarthCube community.

Determining EarthCube's Scope

We recognize that EarthCube is yet a nascent activity within a rapidly growing, diverse, and wide-ranging global environment that is moving forward without in many cases even awareness of EarthCube. Outside of EarthCube there are an untold number of people, programs, institutions, and organizations building and running systems, networks, and programs that make up the emerging cyberinfrastructure. These range from individual data sets and software applications to national and international cyberinfrastructure systems. Many of the contributors to cyberinfrastructure do not yet recognize their roles in the larger picture.

Emerging Governance Concepts

We presented three emerging governance concepts, brought forward by stakeholders in the growing EarthCube community, which influenced our thought process in creating this document. These concepts are not exhaustive, however, and as a wider spectrum of individuals and organizations are engaged in EarthCube, a plethora of potential governance concepts is likely to emerge. Although it will eventually be necessary to decide on a structure for EarthCube governance, we concluded that it is premature at this time to make any final decisions, given that only a sliver of the potential EarthCube community has been engaged thus far.

Community Building and Engagement Program

We presented an overview of our community engagement program to be carried out over the next six months. The overall community engagement strategy we plan to implement is conceptualized into three iterative, often organic phases: 1) Generating actionable intelligence, 2) Messaging, and 3) Engagement. Goals of this iterative community feedback process are to:

1. Vet the initial governance functions list and guiding principles presented in this document
2. Obtain website functionality and user requirements for online collaboration
3. Engage individuals and organizations not yet involved in EarthCube
4. Maintain momentum gained from the June 2012 NSF charrette
5. Identify gaps and overlaps in the existing EarthCube community
6. Provide for community engagement to minimize risk and ensure that EarthCube will be of use to the broader geoscience community.

Engagement activities will be focused on the broad spectrum of EarthCube participants, and will be tailored to reach out to specific communities, including geo, atmosphere, ocean, computer and information sciences communities, software developers, standards bodies, IT experts, private industry, government agencies, and the international community.

Community Building and Engagement Work Plan
We also presented a work plan for our community building and engagement program.

Part 1: Generate Actionable Intelligence

1. Leverage the EarthCube Stakeholder Alignment Survey to identify EarthCube stakeholders
2. Leverage the EarthCube Stakeholder Alignment Survey to gather community requirements on EarthCube governance
3. Create an EarthCube network map of EarthCube stakeholders
4. Develop EarthCube stakeholder profiles

Part 2: Develop EarthCube messaging strategies tuned to different stakeholder groups identified in Part 1.

1. Determine an EarthCube messaging strategy
2. Create marketing material to explain the EarthCube vision and introduce governance
3. Work with EarthCube Working Groups and Concept Teams to create short introductions to their groups, to be included in EarthCube marketing materials
4. Use online forums to market EarthCube
5. Identify community champions
6. Publish articles in scholarly journals and trade publications

Part 3: Engage EarthCube Stakeholders

1. Engage EarthCube stakeholders via social media
2. Engage EarthCube stakeholders via virtual meetings and workshops
3. Engage via in-person meetings and workshops
4. Gather EarthCube website requirements.

In Conclusion
This document serves as a summary of our initial Governance Framework recommendations to be vetted by the community. As we engage the broader cross-spectrum of EarthCube stakeholders,

the Governance Framework is likely to change. We plan to implement our community engagement plan throughout the second half of 2012, and present the findings to NSF and the greater EarthCube community throughout the process, with a report released in 2013. In this way, the body that implements Steps 3 and 4 of the Governance Roadmap may build on the community engagement program and Governance Framework that we recommend here.

APPENDIX 1: JUNE 2012 EARTHCUBE CHARRETTE MATERIALS

The NSF EarthCube charrette (community event) that took place on Jun 12-14, 2012, in Arlington, VA, provided a unique venue for the Governance Working Group to present Governance Roadmap findings to, and gather EarthCube governance wants and needs from, EarthCube Working Groups and Concept teams, in addition to a large group of domain and IT scientists that had not yet participated in EarthCube. Several activities facilitated discussion about EarthCube governance, including a presentation introducing basic concepts of governance and Governance Roadmap findings, and several question and answer sessions that allowed the Working Groups and Concept teams to ask each other pertinent questions regarding their particular area of focus in EarthCube. The outcomes of these activities are presented below.

EARTHCUBE GOVERNANCE PRESENTATION: JUNE 12, 2012

The EarthCube governance workshop introduced the Governance Roadmap to the EarthCube attendees and allowed audience members to raise questions and comments. It highlighted the principle milestones and tasks from the Roadmap, in addition to the 10 NSF guidance points as they relate to EarthCube governance. The audience was then tasked with coming up with pluses and challenges to EarthCube governance as presented in the Governance Roadmap, in addition to a community question and answer session following the presentation. Audience recommendations, pluses and challenges are presented below. Notes were recorded by Andy Burnett of Know Innovation and Genevieve Pearthree of the Arizona Geological Survey. For more detailed notes, please see:

1. http://earthcube.ning.com/group/governance/forum/topics/initial-analysis-of-ec-roadmaps-governance-wants-and-needs-and and
2. https://docs.google.com/document/d/10YnLTYEXemIYq16Jymv2MuR-MBHx8AZ8G15zKj_bZKA/edit

AUDIENCE RECOMMENDATIONS

Address Legal Issues

- There needs to be a legal framework.
- Foe legal elements, governance should be centralized.

Address Technical Components of Governance

- Software development and management needs to be agile.
- Technical governance must enforce policies and verify compliance.

Coordinate across Institutional and International Boundaries

- I hope EarthCube will be a component of an international system.
- Governance has to address responsibilities to multiple agencies.

- You need to coordinate to build synergies with other efforts.
- We need to pick a coordination model for joint development across agencies and systems.

Consider Guidelines for Setting up a Governance Framework

- You may need to setup one year tests to see what works and adjust after that.
- Governance should be a new kind of working group.
- Needed discussion on how infrastructure enables collaboration.
- You need confidence in security.
- Governance should be scalable.
- You should choose a successful venture and mimic it at the beginning.
- You should align governance with other initiatives with synergistic aims as a coalition.
- You should let existing cooperative groups propose how they will grow their governments up to a middle level - organic model.
- Governance of different components is different within the system.
- We may need heterogeneous models, depending on levels such as in systems of systems and ecology.

Include Centralized and Decentralized Governance Components

- The concept of a governance ecosystem is very useful. Ecosystems are resilient, but they have no top-down control.
- EarthCube needs centralized governance for the big tail of participants.
- A system of systems governance follows other success stories, like the internet, and the idea that governance may be centralized at the beginning, but will then naturally decentralize as EarthCube grows.
- There needs to be a central element in governance to serve as a synthesis for communities to come together.
- Having a small, dedicated program management office is a good idea.
- You need to determine what governance of a 'system of systems' means.
- Governance of EarthCube may have to include all governance models because it will be so complex.

Include Education and Outreach

- We need an education effort on governance to inform groups forming governance bodies.
- Educating the users needs to be a part of the governance plan.

Map the Stakeholder Community and Create a Stakeholder Roadmap

- It would be good to make a common roadmap for success to determine who the stakeholders are, and who should be engaged first. This roadmap will help navigate a multidimensional space.
- You should also make a multi-dimensional map of stakeholders.
- Community groups probably need better structured charters so people know roles and responsibilities, mission and vision, how to engage. This could be a rules of the road (1-3 page charter) about how groups orient and connect to each other within EarthCube. This document should feel inclusive, and it should be updated regularly.

Promote Buy-In by Domain Scientists

- People involved in EarthCube so far are mostly on the IT side. You need to engage scientists to bring them in.
- Scientists who are not IT knowledgeable still need to feel part of the governance.
- Recognized scientific leaders in geosciences should buy-in.

Promote Broad Community Involvement

- This group is only a subset of the community, so choosing a committee may not be representative.
- Need for broader community engagement. Only 600 of 14,000 responded to the survey.
- We should poll communities to measure perceived impacts and benefits of governance structures.
- EarthCube needs a simple collaborative decisions structure and a neutral synthesis of community input.

Refine EarthCube Vision, Mission, and Goals

- Knowledge management and behavioral science guidance is needed to guide the vision and understanding and collaboration.

GOVERNANCE ROADMAP PLUSES

Governance Models

- Good start on governance model types and the other possibilities too.
- Definition of governance exists in government styles have been listed.
- Two options laid out centralized versus diffuse; compromise organizations possible.
- I like NSF willingness to be challenged in terms of governance options.
- Review of different governance models is extremely useful.
- Variety of governance models each with pros and cons.
- Many good examples of successful initiatives to use as examples.

Leveraging

- Governance should leverage existing activities.
- I like the recognition leveraging in the middle level to set some standards.

Specific Roadmap Items

- There is an interim governance committee.
- It is an actionable roadmap.
- It elevates governance issues into view.
- The process was inclusive.
- I liked the recognition that governance has to be nimble.
- I like the recognition that governance has to evolve over time, reflecting changing needs.
- I like that the Roadmap broke down what you're trying to address by requirements. This promotes innovation and enables the governance team to look at how current social media and computing environment can help address challenges and requirements.

Timing

- The aggressive timeline is important.
- I like the streamlined timelines.
- Governance should be stood up quickly.
- It is a challenging timetable to catch up for community.

GOVERNANCE ROADMAP CHALLENGES

Choosing a Governance Model

- The governance model maybe too abstract to handle lower-down interaction and this must accommodate these realities.
- What governance model is appropriate for systems of system?
- Any choice of model for governance will make someone mad.
- How does governance apply to collaboration environments?
- How can we evolve the governance model?
- What is the governance of governance?
- The governance hierarchy is analogous to country, states, counties, cities.
- The constrained ecosystem that the NSF operates within will constrain progress.
- Avoid a tool driven approach.
- Who does governance group report to?

Community Involvement

- What is the relationship to a community driven identification research initiative?
- Can we assume "if we build it they will come"?
- What process will be used to engage the community in two months to form the governing bodies?
- Will individuals who spent many years developing their own data resources recognize the authority of EarthCube, whatever the governance model?
- Risk of scaring general geoscience population.
- Governance will be driven by IT/CI or will there be a scientific vision that drives governance?
- What are the scientific drivers of EarthCube?
- Many people at the Charrette already have mechanisms to find out what community wants.
- Have we given enough options for people to move forward? What are the options?
- How do we decide how to move forward? Who is involved?

Education and Outreach

- How do we get a training system to get people involved in science and IT?

EarthCube Mission, Vision and Goals

- How do you define governance without sufficiently defining EarthCube?
- The scientific vision that must underlie earth cube is not obviously represented in the governance model.

- What is a scientific vision? How will it be determined?
- How do we decide what and who are being governed?
- IT vs. Science
- EarthCube could be the greatest scientific instrument ever created for the Earth sciences, but it has to have a scientific vision to achieve goals.

EarthCube's Relationship to Other Initiatives

- Leveraging existing activities (not invented here).
- How is EarthCube different to other similar efforts? How does it relate to them?
- Other group roadmaps will be heterogeneous. This is hard to integrate.
- Subsystems already serve and are responsive to the needs of the science communities.
- There are interdependent NSF projects and deliverables and schedules.
- Getting independent agencies that already have governance structures to work together will be challenging in terms of communication, standards-building, consensus, etc.
- It's not like we're starting from scratch. Systems of systems – lots of systems are already serving communities.
- EarthCube is not unique. Other people are doing things like EarthCube (like in the UK)

Funding Sources

- There are multiple funding sources; each has different governance requirements.
- How will the two months of effort be funded?
- Is the NSF only a venture capitalist for EarthCube?
- Will the NSF accept an earth cube governance model that controls allocations?
- Volunteering-based governance is hard to maintain and be productive.
- Bottom-up federation of resources is driven by need; it's not top-down.

International Involvement

- How can we engage internationally with the various agencies and companies?
- International considerations – how will collaboration occur?
- Any effective data infrastructure must be international.
- Will EarthCube federate with international efforts?
- EarthCube once released will be global, and will have global funding, flexibility and legal frameworks.

Technical Components of Governance

- How will we keep in the sweet spot of evolving technologies?
- Will there be a central place for published data models, unique to each system?

Timing

- If August isn't enough time to agree on a permanent governance model, can an interim one be used?

GOVERNANCE QUESTION AND ANSWER SESSION: JUNE 13, 2012

The Governance Group was given the opportunity to develop a list of questions for the other Working Groups and Concept teams to answer regarding key issues brought to light by the Governance Roadmap. Although not all of the questions were answered by all of the groups, this activity provided important feedback on the other groups' governance wants and needs.

Part of the community engagement program to vet the initial EarthCube governance framework involves working with the other groups on a one-on-one basis to flesh out their governance wants and needs for EarthCube. These interactions will be based not only on the governance questions below, but also on the additional analysis of each group's roadmap, and the governance functions and guiding principles at the beginning of this document.

The original answers for this session are available at: http://earthcube.ning.com/page/materials-june-2012 (see *Google docs for Day 2 Morning Q and A*).

Groups that did not answer a specific question are removed from that question's answer table. As of August 15, 2012, the Workflow group had not answered any questions.

1. DO YOU AGREE THAT AN EARTHCUBE ENTERPRISE GOVERNANCE FRAMEWORK SHOULD BE IN PLACE FROM THE START?

Working Groups	
Data Discovery, Access and Mining	Can start in parallel, but needed sooner than later.
Semantics and Ontologies	Not sure.
Concept Teams	
Brokering	Yes…
Cross-Domain Interoperability	We see a transition period with introduction of governance elements in parallel with others; at the start, we need priorities for allocating resources for implementation; later, we need process governance; coordination of use-case collection and implementation across earth-cube
Earth System Models	EarthCube enterprise governance is not needed at start for ESM; indifferent.
Layered Architecture	No, it will evolve. Does governance include secondary institutions? yes it should
Web Services	Minimal at the start, lightweight and flexible framework. Should ideally be more of a bottom-up, emergent. With the caveat that we need a very good idea of where the overall goal/objective/focus of EC lies.

2. WHAT ELEMENTS OF YOUR GROUP'S ROADMAP ARE DEPENDENT ON THE INSTANTIATION OF AN ENTERPRISE-WIDE GOVERNANCE FRAMEWORK, AND WHAT ELEMENTS CAN BE DEVELOPED INDEPENDENTLY?

Working Groups	
Data Discovery, Access and Mining	Prioritization of science drivers, such as key use cases (available in roadmap documents)
Semantics and Ontologies	Organizational framework for maintenance and deployment of available ontologies. Need to form a collaboratory to capture existing ontologies and align them for use by community
Concept Teams	
Brokering	Governance is handled at a higher level, we just exchange data.
Cross-Domain Interoperability	see 1; in absence of GOV, keep going
Earth System Models	ESM will benefit from when processes get standardized, but it's not holding things back now. Standardization will benefit ESM, but not sure if the governance group will be a standards body or just a liaison to standards body.
Layered Architecture	System of systems approach, so not dependent on instantiation on governance.
Web Services	Recommending a process for "approving" best practices for deployment of SOA elements.

3. KEEPING THE ABOVE QUESTION IN MIND, IF ENTERPRISE-LEVEL GOVERNANCE IS NOT INSTANTIATED UNTIL AFTER JANUARY 2013, HOW WOULD THAT IMPACT YOUR TIMELINE?

Concept Teams	
Brokering	Sooner is better as it defines what is available for us to move.
Earth System Models	As of today, it will not impact ESM timeline.

4. WHAT ARE REASONABLE SCOPE ELEMENTS/ RESPONSIBILITIES FOR ENTERPRISE-LEVEL EARTHCUBE GOVERNANCE?

Who determines this?
Authority for funding to reduce risks and fill functional gaps?
Integration of piece-parts into an operational enterprise (EC architecture)?
Adjudicate resource-contention issues?
Set priorities for EC functional development?
Coordination, communication, education, outreach?

Concept Teams	
Brokering	Standards and interoperability issues are the most important to brokering group. Look to the governance group to prioritize resources to brokering.
Cross-Domain Interoperability	See road-map: we need a priority-setting body ("scientific governance") vs. technical body ("technical governance) for standards (geo-science interoperability inst.); priority includes use-cases.
Earth System Models	
Layered Architecture	NSF-funded project : coordination of activities with projects and agencies ; standardization; promote development
Web Services	Coordination, communication, education, outreach? Some level of leadership is needed to avoid the chaos of bottoms up development. Need both social and technical governance construct? There is value in some level of central authority to facilitate decision making.

5. WHAT ARE THE TOP THREE ATTRIBUTES/ACTIVITIES/DATA/PRODUCTS/SERVICES/CAPABILITIES THAT YOU EXPECT EARTHCUBE TO PROVIDE?

Working Groups	
Data Discovery, Access and Mining	DataSpace, i.e., conceptual infrastructure for distributed discoverable access for value-added innovation: key aspects: low barrier to entry for long tail of science; adding social glue, community space for sharing data, thoughts; feedback mechanism, w.r.t. readiness levels, technology, process of managing data.
Concept Teams	
Brokering	does not apply
Earth System Models	Ways to standardize models and model metadata, and identify associated use cases. Whatever standards are necessary for seamless integration.
Layered	layered is focusing on simplification so expect to get infrastructure to enable

Architecture	this
Web Services	Value of web services is enhanced by widespread and consistent adoption. Can EC make progress in identifying and prioritizing the need? Are the goals of EC amenable to a service level architecture?

6. WHAT DOES YOUR GROUP NEED FROM EARTHCUBE GOVERNANCE?

Does your group have any tasks or milestones that have a specific governance dependency? How can the governance framework prioritize tasks and manage allocation of resources?

Working Groups	
Semantics and Ontologies	SEMANTICS: policies for open data access; management guidance for legacy data; policies for international outreach and other stakeholders outside EarthCube; policies for interaction with other national interests
Concept Teams	
Brokering	Quick framework for the other layers.
Cross-Domain Interoperability	
Earth System Models	[Refer to ESM answer to question 2]: ESM will benefit from when processes get standardized, but it's not holding things back now. Standardization will benefit ESM, but not sure if governance group will be a standards body or just a liaison to standards body.
Layered Architecture	ANSWER: multiple layers of governance: NSF governance; technical governance especially if compliance with standards is part of governance; computer-enabled policy
Web Services	Require roadmaps and use cases from the community. Need to develop EC wide standards. Require prototyping of new services. EC governance needs to provide a structure that brings together the efforts of the community into a system.

7. HOW DO WE IDENTIFY, PRIORITIZE, AND ENGAGE EARTHCUBE STAKEHOLDERS?

Working Groups	
Data Discovery, Access and Mining	
Semantics and	outreach to communicate EarthCube vision to all geoscientists; engage

Ontologies	existing national initiatives according to greatest impact on broadest community
Concept Teams	
Brokering	Standards development and conformance. Under what conditions will the governance structure accept or a new brokering model. Will there be minimal requirements that new brokering schemes need to meet to be acceptable?
Cross-Domain Interoperability	community focus groups; polls are for measuring not communicating; rally people around interoperability issues; address people's pain-points >> use-cases >> find ways to empower starting with low hanging fruit; start with engaging with other fed agencies that are dealing with same issues; long-tail champions should also be identified
Earth System Models	*[Refer in part to ESM answer to question 11]:* Most model frameworks follow an agile development approach with respect to continual stakeholder input; most communities have a local governance to obtain stakeholder input. EC governance should endeavor to integrate with these community governance entities. Regarding how are ESM codes incubated and matured, currently most frameworks use code repositories with agile methodologies and stakeholder inputs integrated with these.
Layered Architecture	use-cases instantiated as exemplars with common themes of automation to minimize work for science; simplified access to resources
Web Services	Web services needs to be included in informing the overall design.

8. WHAT DO YOU NEED FROM EARTHCUBE GOVERNANCE IN TERMS OF RELATIONSHIPS AND INTERACTIONS WITH OTHER GROUPS?

Concept Teams	
Brokering	We believe that you need to implement a rating system like Better Business Bureau/Amazon ratings for funded scientist? How to police data sharing requirements?
Web Services	Need to enforce agreed upon standards and common infrastructures (i.e. authentication services). Privacy issues are also an important governance issue particularly in the global environment. Ran out of time here!

9. IF THERE IS AN INTERIM GOVERNANCE COMMITTEE TO BEGIN THE PROCESS OF IMPLEMENTING EARTHCUBE GOVERNANCE FRAMEWORK, HOW DO WE CHOOSE COMMITTEE MEMBERS?

Layered Architecture	Representatives from OOI, GEON, etc., NSF national scale projects that are implementing infrastructure; intermediaries who have already gathered requirements and reached out.

10. WHAT ARE YOUR EXTERNAL DEPENDENCIES THAT HAVE TO BE ACCOMMODATED BY EARTHCUBE GOVERNANCE (E.G., FUNDING FROM OTHER ORGANIZATIONS, STANDARDS SET BY EXTERNAL BODIES)?

Concept Teams	
Brokering	Use spiral development model - continuous improvement
Cross-Domain Interoperability	Not at this level; we have not discussed how to incubate and mature services; we considered this the purview of other groups; see pg. 39 of our doc for service requirements. A critical part is that domain systems should produce services that make it easy to create objects in the cross-domain layer (top); see note on governance here;
Earth System Models	Most model frameworks follow an agile development approach with respect to continual stakeholder input; most communities have a local governance to obtain stakeholder input. EC governance should endeavor to integrate with these community governance entities. Regarding how are ESM codes incubated and matured, currently most frameworks use code repositories with agile methodologies and stakeholder inputs integrated with these.
Layered Architecture	Hard to answer! Testing and bug fixes are prioritized; prioritization of requests by governance; prioritization of development goals for entire community; the 8 use cases are a great start
Web Services	Broker function needs to be funded as a core EC element. EC should rely on external standards bodies but Governance may need to identify a set of preferred standards.

Additional Questions Not Answered by Working Groups and Concept Teams

1. How does your group plan to implement internal governance?
2. How should this internal governance relate to "enterprise level" governance?
3. What attributes should be considered when evaluating different governance archetypes (models)?

GROUP MIXING: JUNE 13, 2012

On the second day of the charrette, the Governance Working Group was assigned to two discussion sessions: the first with the Semantics and Ontologies Working Group; the second with the Layered Architecture Concept Team. Each session focused on three areas in each group's roadmap: overlaps, potential conflicts, and shared concerns. For a complete list of Group Mixing activities and notes, please see: http://earthcube.ning.com/page/6435147:Page:21585

SEMANTICS & ONTOLOGIES AND GOVERNANCE

Notes were taken by Ray Idaszak (Renaissance Computing Institute, University of North Carolina at Chapel Hill) and Genevieve Pearthree (Arizona Geological Survey). For more detailed notes on this session, please see:
https://docs.google.com/document/d/1VTO8hAI8Ew4KITTelJMCGIJ4A0c6LtAeGuvHxyAGO6Q/edit

Possible Overlaps between Semantics & Ontologies and Governance

1. Facilitation of the infusion of semantics into the entire EarthCube enterprise
    a. But, infusion process is not defined at this point
2. Governance and Semantics/Ontologies have a symbiotic relationship
    a. EarthCube cannot work unless it has a robust semantics approach
    b. EarthCube cannot work unless there is a robust governance approach that facilitates the integration of all groups, including semantics
3. Potential governance mechanisms to infuse semantics into EarthCube include patterns
4. Semantics should be pulled, not pushed, into other groups

Potential Conflicts between Semantics & Ontologies and Governance

1. Overall goal of EarthCube is to provide data and model interoperability. This requires semantics support for every layer of EarthCube.
    a. Linked Data can fail with certain complex data types; need to constrain or declare conventions for limited use, for interoperability.
2. There are many ontologies; EarthCube may need to mediate among them.
3. Is there any conflict with how EarthCube participants see the role of semantics
    a. Why do we have to do that? (need 1-page brief explaining)
4. How can we mediate or unify approaches to interoperability across all the current national research centers?
5. How can EC governance enable the development of requirements for the semantics group?
6. If we try to agree on creating a common set of top-down domain-level and foundational ontologies, how do we handle conflicts in deciding how to formalize, what to formalize, and how and when to change these ontologies if our knowledge evolves over time? How may this hinder emerging research trends?

7. Semantics group has a way of governing its own tasks; how will this play with respect to EarthCube governance? See W3 Ontologies best practice around governance. Apache Foundation may be good governance model (e.g., based on meritocracy and other factors). Contracts define terms of relationships; could vary with different types or communities of users.
8. What level of conflict would trigger a governance action?
    a. If one community wants to use resources in a different way than another community.
9. Does semantics framework have to be ready and decided at the beginning of EarthCube?
    a. No, while it would be helpful, there will inevitably be use cases calling for different or new semantics than have been agreed in the past.
11. How do we accomplish interoperability across mature infrastructures with least damage to each? Could be within EarthScope or external.
12. Should NSF be requesting all proposals to include "semantics & ontologies plan" like the data management plan now required? Shouldn't this be part of the data management plan?
13. Investment in time by researchers vs. improved capability

Shared Concerns between Semantics & Ontologies and Governance

The discussion began with a clarification of what the EC Governance group was tasked with, understanding that the EC Governance group was not tasked at this time with prescribing what the EC Governance will be, but rather deriving the roadmap for how to get to this. Recognizing the range of governance in general can span from brokering to dictating, with respect to EC it is unanimous that governance must be open, address the long-tail of science, have a very low barrier to entry, facilitate democratized access to participation, and be very simple to deploy. EC shared concerns between the Governance and Semantics groups engendered the need for people to continually get together, community engagement, community building, priority setting, conflict resolution, resource allocation and decision making. Specific shared concerns between the EC governance and semantics groups included the following:

1. There is a fear is that EC might end up being monolithic when the desire is for EC to be broad. For example, one EC group has proposed an Interoperability Institute, another an "Apache-like" Software Foundation for EarthCube called "EarthCube Software Foundation." In a sense each of these could be an "EarthCube" in itself, however the desire is that EC must encompass the best of these while remaining broad.
2. Mechanisms for creating and expanding engagement of domain scientists with technologists.
3. Mechanisms for identifying and deploying cross-cutting methods/standards for EarthCube
4. Mechanisms for assuring ongoing communication among related groups in EarthCube
5. Mechanisms for increasing outreach of EarthCube beyond EarthCube engineering efforts per se-- e.g. to international scientists, incorporating legacy data.
6. Mechanisms for creating broader engagement and sustained communication with external related efforts-- e.g. NASA, EU, etc.

7. Mechanisms for "certifying" EarthCube technologies and approaches
8. Mechanisms for objective evaluation among various pieces of EarthCube
9. There are and have been many other collaboratory approaches. How is EC different, how does it improve upon these? That EC has been open and democratic is a good sign.
10. How do we get EC to a concrete development processes?
11. How open is EC to participation? What does EC entry membership look like? What does the whole process of EC participation look like?
12. How is EC going to assure that all relevant aspects of domain and CI areas are being fully represented by experts in those areas?
13. Is there going to be a process to identify and address boundaries of EC? Who is inside, who is outside the boundaries of EC, and how does one foster each with respect to EC governance? EC must map stakeholders to these boundaries, and must find synergies and missed opportunities with respect to these boundaries.
14. Researchers are very busy, and may not respond well to being legislated to by EC, but may respond better in understanding how they can turn to EC to assist them.
15. What about equity, vigilance, process? Governance should have representative expertise needed, be broadly representative, be comprised of a good group of people who can move things forward, and not be biased.
16. How are computing and earth science community paired? EC Governance should ensure a good pairing. In the long term, this is what is sustainable recognizing the concern that this often fails.
17. Identifying relevant science *and* governance use cases must be an early part of EC activities. Start with cross-cutting use cases so EC isn't a silo to begin with, and have frequent workshops so it remains "built-in" that this *is* a community.
18. Upon receipt of a lot of funding, a previously "open" NSF-funded effort may then instead focus inward. This is a concern of what will happen with subsequent EC funding. Therefore, a concern is that EC Governance must "police" this so this doesn't happen.

GOVERNANCE AND LAYERED ARCHITECTURE CONCEPT TEAM

Genevieve Pearthree (Arizona Geological Survey) took notes. For more detailed notes, please see: https://docs.google.com/document/d/1LQh2FyYU9EXt4LO-TY2bqut3gqp6ESakR2u-UFSeYXU/edit

Layered Architecture and Governance Overlaps:

1. Technical governance: Layered Architecture is focused on technical governance, which is one component within the greater EarthCube governance framework.
2. Identify and prioritize use-cases: Help explore the role of governance in terms of funding, priorities, conflict resolutions, etc.
3. EarthCube governance needs to go forward as quickly as possible to avoid conflicts between EarthCube and already-funded projects.
4. Need objective and quantitative approach to evaluate governance models, share them with the community, and get community feedback

Layered Architecture Governance Needs:

1. Strong consensus on set of technical policies to be enforced
2. Governance model for a distributed, system of systems approach to promote collaborative development of different components across independent projects and integrate different components.
3. Technical governance policies can be pushed up by the community for NSF approval.
4. Governance could provide a 'Good Housekeeping Seal of Approval'/ audit system to evaluate compliance on data and services. But, who decides what compliance means? Issue for the governance framework.

Shared Concerns between Layered Architecture and Governance:

1. What if NSF establishes data management plans and asks EC to enforce those?
2. How do we achieve consensus across diverse communities?
3. Balancing the needs of vocal individuals with resources allocations that will have the greatest impact systemically.
4. Identifying the right constituencies to make required decisions
5. Community governance vs. NSF governing EarthCube
6. What happens when multiple funding sources are involved?
7. Interoperability of existing organizations with limited resources
8. Mechanisms to give people credit and incentives to contribute, including program, data, and software citations
9. Communities of interest with micro-policies
10. Integrating already-funded projects into EarthCube
11. If multiple agencies contribute technology to EarthCube, which one gets credit?
12. How is authority vested?
13. How does technical governance fit in with overall governance framework for EarthCube?
14. How is funding controlled?
15. Adjudication: how is it decided what initiatives go forward within a climate of limited resources?

APPENDIX 2: EARTHCUBE WORKING GROUP AND CONCEPT TEAM ROADMAP SUMMARIES

This appendix is a summary of the draft Working Group and Concept Team roadmaps submitted to NSF prior to the June 2012 EarthCube charrette (accessible at: http://earthcube.ning.com/page/draft-roadmap). Although the roadmaps were updated and re-submitted to NSF on August 15, 2012, there was not sufficient time to analyze these roadmaps and include many changes from these in the summaries presented in this document. Throughout the next few months, these updated roadmaps will be read and analyzed for their governance wants and needs, and a summary will be included in a final version of this document.

NSF provided 10 guidance points for each of the EarthCube Working Group and Concept Team roadmaps. NSF directions for each of the 10 points are written below.

1. **Purpose:** Introduction, including community(ies) to be served, technical area(s) of the roadmap, and brief discussion what improvements in the present state-of-the-art in geoscience data discovery, management, access, or utilization it will enable. Also include examples of how the outcomes from your effort will enable the community to be more productive and capable.

2. **Communication:** Description of a communications plan with end users, developers, and sponsors, as well as links to and feedback from other EarthCube community groups and EarthCube concept projects to promote systems integration and accelerate development. Include a discussion of needed interactions with allied fields, agencies, and other related activities (present and desired).

3. **Challenges:** Description of major drivers, trends, and shifts impacting or that could impact the focus of a working group, including but not limited to changing technology, adoption culture, and community engagement.

4. **Requirements:** Process(es) to be used to get the necessary technical, conceptual, and/or community (i.e., end-user) requirements at the outset and during the life of the activity, including approaches to achieving community/end-user consensus.

5. **Status:** Description of the state of the art within the topical area of your roadmap. This should include approaches and technologies from geoscience, cyberinfrastructure, and other fields, the public or commercial sector, etc. that have the potential to benefit the EarthCube enterprise.

6. **Solutions:** Process for the identification and comparison (pros and cons) of approaches and technology solutions that will contribute to the EarthCube goal of satisfying current and future research needs of the geoscience end-user.

7. **Process:** Process(es) to develop community standards, protocols, test data, use cases, etc. that are necessary to mature the functionality of the topical area and promote interoperability and

integration between elements of EarthCube.

8. **Timeline:** Timeline for the project and all related sub-projects, including prioritization of activities and measurable milestones/major achievements and total resources (human and financial) required to achieve roadmap goals over a period of the next 3 to 5 years.

9. **Management:** Management/governance/coordination plan and decision-making processes necessary to successfully establish standing committee(s) and subcommittees (if warranted), including a plan to identify and respond to shifts in technologies and changing needs at the end-point of use. Include discussion of approaches to educating end-users and achieving community consensus on advancing the capability/technological solution.

10. **Risks:** Identification of risks and additional challenges to the successful establishment of any working group, and any unique risks associated with a working group associated with your topical area. With respect to identified risks, an approach to risk mitigation should be addressed.

DATA DISCOVERY, ACCESS AND MINING WORKING GROUP

1.0 PURPOSE

- "EarthCube DataSpace is our prominent vision of the landscape that will bring under a single umbrella the big-head of science and the long-tail geoscientists. It provides long-tail scientists with an easy way to upload, share, discover and analyze their data; It interfaces with the big head of science, both, in terms of bringing their data on a per-use-case need basis, and by leveraging data management practices followed by them over several years."[59]

- **Characteristics**
    - "Provide the much-needed **low-barrier for entry** and simplified data management by providing uniform services for storage, access, curation, discovery and analytics;
    - **Enable and democratize the community** by specifying a set of complaint capabilities and services which anyone can implement;
    - **Provide feedback in terms of readiness levels in** assessing a tool for compliance with existing capabilities and services, to enable technology insertion, especially in the changing data technology landscape;
    - Provide a **social glue** for the EarthCube community by encouraging dialogue and communication on issues." [60]
- Framework for data discovery, access, mining

2.0 COMMUNICATION

- **Communication tools:** Virtual meetings, EC Ning site, Google docs, ESIP wiki, etc.
- **Links to other communities and agencies**
- **Structured interactions:** Network map of participants
- **Community engagement:**
    - Improve engagement and fundamental conversations between computer scientists and geoscientists
    - Use collaborative tools
    - Training/education
    - Mailing lists to advertise upcoming events

3.0 CHALLENGES

- **Changing Technology** : Scale, data preparation

[59] EarthCube Data Discovery, Access and Mining Working Group, "A Community Roadmap for EarthCube Data: Discovery, Access, and Mining," draft roadmap released June 2012, 3.
[60] Ibid., 3-4.

- **Social challenges:** Data divided into distinct disciplines, but used for interdisciplinary policy decisions; Data crosses cultural and institutional boundaries
- **Adoption culture:** Adoption lethargy, tool sustainability, etc.

4.0 REQUIREMENTS

- **User needs identified thus far**
    - Cross-disciplinary access
    - Real-time data
    - Better structured and documented repositories
    - Access to more datasets
    - More web-based processing tools to analyze raw data

5.0 STATUS

- **Distinct challenges** in data discovery, mining and access
    - **Discovery**: metadata, standards, semantics
    - **Access**: formats, systems, technologies currently employed in ad-hoc way
    - **Mining:** technical challenges cause high entry barrier, need user training,

6.0 SOLUTIONS

- **DataSpace**
    - Low entry barrier
    - Data storage, access, curation
    - Capabilities and services so anyone can implement compliant services
    - Technology 'readiness levels'
    - Curation cooperatives
    - Decide what it means to publish data
- Establish a **common EarthCube Data Model**
- **Domain Specific Software Architecture (DSSA):**
    - Established process and infrastructure to develop domain model, reference requirements, and reference architecture for a software system
    - Will use to establish requirements for DataSpace, Workbench and other DDMA functionalities

7.0 PROCESS

- **Dependencies in integrating tools** with other elements of EarthCube
    - Require interactions with other groups to define interface requirements, develop test cases, perform integration testing
- Established process and infrastructure to develop domain model, reference requirements, and reference architecture for a software system

- Will use to establish requirements for DataSpace, Workbench and other DDMA functionalities

8.0 TIMELINE

- **First 3 months:** Establish working groups and logistics (communication, etc.); Launch reference architecture
- **Liaisons** with other EC groups and greater community
- **Governance group (in DDMA):** Develop Governance Plan to leverage lessons learned to ensure fairness, open participation, collaborative decision-making

9.0 MANAGEMENT

- **Curation cooperatives**
    - Within, among communities and sub communities
    - Evolve into trusted bodies and aid in establishing and infusing good practices around peer review of data and data citation
- Establish a **common EarthCube Data Model**
    - What would it enable?
    - Limitations?
    - Recommendations or requirements to inform EC design and development

10.0 RISKS

- **Risks**
    - Ensuring openness and fairness and address cultural issues
    - Evolving technologies
    - Funding
    - Technology adoption
    - Community participation
    - Fragmentation due to difficulties in linking roadmaps
- **Risk mitigation**
    - Working groups will represent stakeholders
    - Open communication
    - Respond to technological advances and changes in technology

SEMANTICS AND ONTOLOGIES WORKING GROUP

1.0 PURPOSE

- Development and use of semantics & ontologies to **standardize meaning of terms:** Resolve terms across subdomains (bridging)
- In EarthCube: S & O Promotes **interoperability and brokering ; Cross-cuts all groups**
- Semantics & Ontologies fosters **greater access to geoscience resources** (models, tools, data, workflow templates)

2.0 Communication

- **Liaisons**
  - Aggressive outreach plan to communicate semantic technologies application and development
  - Identify community members as liaisons with other EC groups
  - Outreach Committee
  - Additional committees
    - Develop use cases; Document status of current technology
- **Outreach Plan**
  - Sharing information and creating awareness of S & O and their benefits
  - Sustaining interest in S & O development through communicating updates and through virtual and face to face meetings
  - Development of teaching materials for geoscience community

3.0 Challenges

- Gaining **collaboration and coordination** of people and resources from different domains
- Geo-web and semantic web mostly **incompatible**
- **Benefits** of semantics work may be **incremental:** It will be difficult to encourage scientists and domain centers to contribute
- Existing semantic technologies may not support semantic infrastructure vision

4.0 REQUIREMENTS

- **Foundation and domain-based ontologies (developed top-down or bottom-up) key to EC success**
  - Interdisciplinary communities need to help clarify and agree upon concepts and technologies
  - Semantic infrastructure building must be community-based, long-term, and evolutionary
- **Need to organize existing ontologies**
  - New ontologies need to describe data products required for data discovery and integration
  - Data ontologies need to accommodate broad spectrum of formats across geosciences
  - Service ontologies need to be generic and flexible to support services in all domains and evolve as necessary

- Domain standards and guidance from OGC, FGDC, etc., should be considered

5.0 STATUS

- **Long tail scientists**
  - Getting data from individual scientists into EarthCube will be a challenge
  - **"What is needed includes the development of specific formats, ontologies in different fields, semantic similarity measures and matching, and mapping within linked data**. There could also be more communication among investigators and more integrated collaborative projects being funded."[61]

6.0 SOLUTIONS

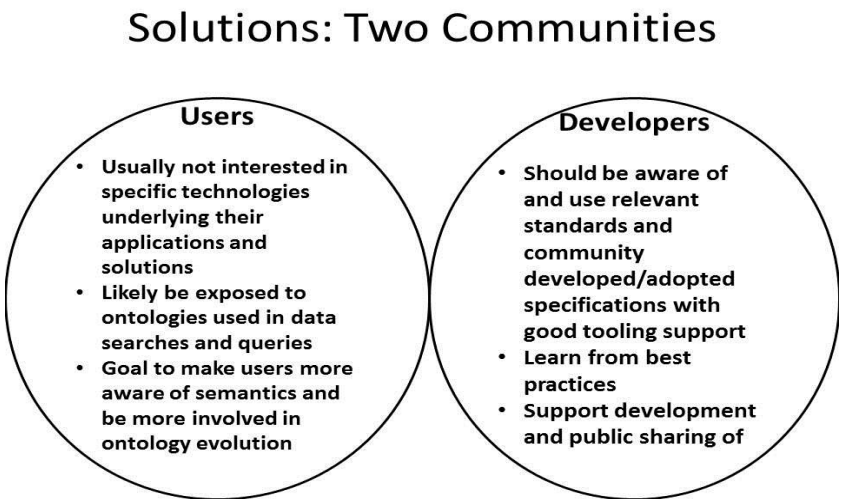

FIGURE 18: THE EARTHCUBE SEMANTICS AND ONTOLOGIES DRAFT ROADMAP DIVIDES THE EARTHCUBE COMMUNITY INTO TWO GROUPS: USERS AND DEVELOPERS

7.0 PROCESS

- **Introduce** semantics to EarthCube and geoscientific community
- Develop **Use cases**
- Develop ontologies: **focus on data discovery and sharing**
- Promote **community buy-in**
  - Process of Engagement
- **Information bulletin** to explain use of semantics
- **Guidelines** identifying useful technologies and pathways for data sharing
- **Working group** to focus on semantic infusion to bridge gap between science and IT communities

[61] Krishna Sinha et al., "Roadmap for Creating the Semantic/Ontologic Infrastructure for the Geosciences," draft roadmap released June 2012, 10.

8.0 TIMELINE

- **6 months – 1 year**
  - Semantics and Ontology workshops
  - Gap analysis & Use cases
  - Create initial foundational, domain and service ontologies
- **2-4 years**
  - Ontology repository
  - Community engagement; work with other EC groups
  - Ontology mapping
  - Ontology tools and software services

9.0 MANAGEMENT

- **Semantics and Ontology Manifesto:** explaining importance of semantics in data discovery and integration, to be distributed to geoscience community
- **Use cases and project registries**
  - Enable evaluation of existing semantics technologies, including vocabularies
  - Should create a semantic template to provide working solution for use case
  - Solved use cases should resolve communication with broader community since they can see the impact of semantic technologies in conduct of science
- **Joint domain/semantic workshops**
- **Capitalize on success** of DataONE, NEON, Data Conservancy, NCAR, OOI, IOOS, etc.
- **Websites/portals** to facilitate communication

10.0 RISKS

Risks of a Dedicated Semantics and Ontologies Group Risks

- A dedicated Semantics and Ontologies group is necessary for EarthCube
  - Addressing semantics and ontologies within other groups would lead to isolated solutions and redundancy, instead of creating a joint, semantics-enabled, EarthCube-wide infrastructure
  - Several prerequisites for other groups can only be addressed by a dedicated Semantics and Ontologies group, although there is a possibility to form working groups to allow other EarthCube groups to identify and meet their ontological modeling and reasoning needs.

Risks from the Semantics and Ontologies Roadmap

Domain experts may be turned away by technical language about ontologies

- Purpose of ontologies has to be defined based on scenarios and application areas and must strike a balance between too-defined and too-general
- The overall community might not accept proposed solutions.
  - Risk mitigation can be achieved by early and frequent usability testing, modularization, conformity testing, and semantic negotiation.
- Semantic aging and ontology evolution Investing in certain technologies over others (for example, Semantic Web)

WORKFLOW WORKING GROUP

1.0 PURPOSE

- **Workflow is a description of processes by which geoscientists work with data, models, and information to contribute to Earth Science knowledge**
- Goal **to improve interaction of scientists with workflow systems:** Easily create, validate, publish, visualize, reuse data

2.0 COMMUNICATION

- Need effective communication and mechanisms to **demonstrate benefits of workflow** technologies
- **Disseminate information** to community through
    - Accessible web information for geoscience researchers
    - Clear points of entry and participation for new participants
    - Virtual and face-to-face meetings
    - Measure community growth and engagement through time

3.0 CHALLENGES

- **Mapping science challenges into workflow technologies to improve process**
    - Diverse and dispersed workflow community
- **Technical**
    - Current requirements of scientific community are beyond reach of current technologies
    - Need to do basic research in EarthCube requirements context and transition to users
    - Partnership between scientists and developers as new workflow capabilities are addressed
- **Broader adoption**
    - Need to address tension between encouraging adoption of mature workflow systems versus development of lightweight customized systems / simple scripting solutions
    - Many geoscientists do not use workflow tools
    - Lost efficiency, lost metadata, lack of reproducibility, limited or no access to national geoscience datasets
    - Challenge to increase access and efficiency and efficiency of access of workflow technologies to geoscientists
- **Reproducibility**
    - will require semantic representations that document enough details about scientific processes in a reusable form to be easily re-run and adapted to other platforms
- **Rapidly-evolving technologies**
    - Impedances to integration
    - Difficulties of moving data, storage, computational resources, security mechanisms, etc.

4.0 REQUIREMENTS

- Need **ongoing process to obtain and understand geoscience community requirements**
  - Surveys
  - Process to obtain user requirements
  - Process to obtain technical requirements
  - Process for obtaining conceptual requirements

5.0 STATUS

See Status and Requirements task force in the Governance Needs and Mechanisms Section.

6.0 SOLUTIONS

- **Geoscientists need to be guided through the process of discerning their own workflows and how they can be automated**
- **Task Forces**
- **Technology Evaluation Framework**
  - Metrics
  - Conditions that must be satisfied
  - Evaluation process
  - Infrastructure
  - Prototyping needs

7.0 PROCESS

- **Define workflow types** and capabilities from community requirements
- **Identify relevant standards** from existing standards bodies
- **Classify workflow software:** Capabilities and scope, as they relate to project management and governance
- **Identify gaps** in standards, workflow tool capabilities, underlying infrastructure
- **Support use of workflow technologies** through training, documentation, outreach
- **Match appropriate workflow experts** to support NSF and other research organizations
- **Facilitate new collaborations**
- **Scientific and technical bridges** to other areas (semantics, data management, etc.)

8.0 TIMELINE

- **Step 1: Establishment (months 1-5)**
  - Steering Committee and basic operational infrastructure
  - Establish task forces
  - Establish Synthesis Center
- **Step 2: Awareness (months 6-18)**
  - Community Outreach and Requirements gathering
- **Step 3: Enabling (months 18-36)**

    - Prototyping proofs of concept and working with the community to disseminate workflow technologies.
- **Step 4: Infrastructure and Services (months 37-60)**
    - Develop community infrastructure to include workflow publication and citation, workflow sharing, workflow execution resources, and other community workflow resources.

9.0 MANAGEMENT

- **Steering Committee**
    - Open Community Model
    - Integrate new members
    - Leverage online tools
    - Specific decision-making processes
- **Organizational substructure**
    - Synthesis Center
    - Task Forces
- **Key issues**
    - Shifts in technologies
    - Changing needs
    - Education and outreach
    - Achieving consensus on advancing workflows

10.0 RISKS AND RISK MITIGATION

- **Risks**
    - Not establishing meaningful requirements
    - Substantive differences in user requirements
    - Not addressing workflow requirements
    - Inadequate communication with user community
    - Lack of adoption
    - Choosing wrong software engineering methodology
- **Risk Mitigation**
    - Establishing EC Workflow Working Group
    - Widespread Workflow Technology User Community
    - Channels of Communication between user community and working group
    - Promoting ease of use
    - Acquiring user feedback
    - Responding to user feedback
    - Effective user support

BROKERING CONCEPT TEAM

1.0 PURPOSE

- Brokering can **address challenges of sharing interdisciplinary data** – discovery, access, semantics, workflows, model web interoperability within EC
- Brokering **capabilities**
  - Dataset and Workflow Discovery
  - Data Access Services
  - Data Management within Workflows
  - Broad Participation: Enable Collaboration and Participation from International, Industry, Academic, NGO and other Domain Partners
  - Continuity, Sustainability, & Evolution
- Take into account **current and future needs**
- **Interfaces** between brokering and other groups essential

2.0 COMMUNICATION

- **Three communication phases:**
  - **Near –term** (before June 2012 Charrette)
  - **Medium term:** June 2012 through end of current contract
    - Focus on technology convergence and developing science users
    - Engage working scientists through partner orgs
  - **Plan through realization of roadmap**: 2013 and beyond
    - Integrated communication and outreach plan for EC
    - Wide audience
    - Uniform approach to national and international participants
    - What new capabilities do people need to see to participate in EC?
- **Communication Recommendations**
  - Needs visible top-down and bottom-up communication
  - Visible impacts of EC within 2-3 years
  - When services available should be clear in development plan
  - Visible advocacy from senior members of science research community beyond CI
- **Communication Mechanisms**
  - Communications plan should be part of milestones
  - Rollouts should be scheduled for highly visible events
  - Communications plan should address of current public interested – to be cited in science columns
  - Major project events should be highlighted in trade journals
  - Have high level keynote speakers for major conferences
  - Community workshops for students and early professionals are important – maintain sustainability
  - Outreach in the form of web formats – news, tutorials, interactive for a

- Highly visible demonstrations of community wide participation in EC development
    - Science teams and advisory boards will have positive impact
- Use of EC infrastructure by a major (>$1B) program should be a priority
- Network and clear communication with major CI activities with EarthCube

3.0 CHALLENGES

- **Challenges to EC as a whole**
    - need to support interdisciplinary research across traditional barriers between disciplines
    - Extreme diversity in data types, formats, data size, scale of measurement, etc.
    - Rapid technology change
    - breadth of the target community from large programs to small, individual scientists
- **Challenges for brokering**
    - Reaching and promulgating a common understanding of brokering and its role
    - Social challenges related to adoption
    - Determining suitable & sustainable brokering architectures for EC
- Addressing challenges will require **multiple iterations**
- Earth system analysis – requires distributed multidisciplinary collaborative teams
- Incorporate many, distributed, diverse data holdings of individual team sand scientists that are not widely accessible

4.0 REQUIREMENTS

- **Adaptive, organic requirements to help diverse systems organize into networks, and networks into broader infrastructure web**
    - Start simple and iterate towards complex solutions
    - Use real use cases to drive requirements
    - Use nested iteration cycles with multiple feedback loops to ensure that the highest value, yet achievable activities are targeted and that course corrections are made as soon as issues are discovered
    - Use pilot projects to verify directions while simultaneously delivering results to at least some portion of the EC community

5.0 STATUS

- **Two approaches to interoperability**
    - **Federation and adherence to common standards** – not possible given diversity of EC data and communities
    - **Mediation/adaptation between standards (brokering)**
- These approaches are not mutually exclusive

6.0 SOLUTIONS

- Hackathons, usability tests, pilots to identify strengths, weaknesses and gaps with existing broker solutions

7.0 PROCESS

- **Use cases /pilots**
    - small pilots to stress technology modules/component and assess strengths and weaknesses
    - broker framework assessments in specific domain sciences but focused on CI development
    - GEO community evaluations in active collaboration with science research community
- **Collect best practices**
- **Cross-EC testbed activity using commonly agreed upon criteria**
    - Interdisciplinary demonstration for diverse providers and users
    - "Full" range of CI exercised
    - Collaboration between EarthCube organizations
    - Provides a clear benefits demonstration
    - Can be continued as a practical implementation (to be determined)
- **Architectural Decision-Making Process**
    - Centralized or distributed brokers?
    - Tailored brokers or a single broker to rule them all?
    - Existing operators or new organizational entities?
    - Interface standards:
        - EC needs to deal with variety of informal standards (interoperability agreements) and variety in implementation of formal standards
    - Evaluation and validation process
        - Different levels of testing, use cases, pilots, etc.
- **Need decisions as part of EarthCube governance process about when to add new broker capabilities**
    - Part of process is determining when there is enough useful data available in a particular format or via a particular protocol to make it important enough to some community for it to become available via a broker

8.0 TIMELINE

- **Spiral development based on pilots, community feedback, identifying and developing components**
- **"July 2012:** CAs meet to harmonize roadmaps and identify joint use cases;
        - Develop an evaluation of current technologies including experiments with early use cases; recommendation on which technologies to pursue further; conceptual approach to a brokering Framework.
- **2013---2014**: expansion of systems connected through Arctic brokering infrastructure;
        - additional use cases for CI development are exercised;
        - evaluation with use cases stemming from NSF domains;

        - engage community to guide development; close collaboration with the various working groups to the Broker Framework with other modules and capabilities to allow in---depth understanding of evolving requirements and capabilities;
        - a decision is make on the broker construct in the System Architecture as part of a convergence on an overall architecture construct; move toward and initial operating capability;
        - A technical governing body for EarthCube is empowered.
- **2015---2017**: Increasing usage of EarthCube as a core service to the geoscience community enabling advances in education and science provides experience and feedback necessary to further advance the underlying technologies. A full Operating capability (IOC) is available in 2016---17."[62]

9.0 MANAGEMENT

- Roles and Responsibilities
    - **Broker Concept Group:** core team
    - **Contributors:** Community members who contribute in concrete ways to project through hack-a-thons and pilots; Give feedback to move brokering project forward
    - **Broker Maintainers: "Currently, there is no single broker provider or maintainer. The group will need to consider** these questions regarding **broker maintenance:**
        - Who maintains the broker?
        - How long does the broker commit to being operational?
        - Who decides what catalogs are brokered?
        - What are the requirements of data providers or catalog maintainers to the broker?"[63]
    - **Liaisons** to other EC groups
- **Support:** All participants encouraged to provide support to new uses and grow the community
- **Decision-making process**
    - **Near-term:** team decisions informed by community groups, partners and collaborators working on joint use cases
    - **Longer-term:** Rely on establishment of governance framework
        - Help prioritize work required to fulfill evolving brokering requirements
        - Establish processes to ensure that implementations of tasks selected by community employ standards and best practices to ensure interoperability and maximize their utility by, and integration with other components.
    - Use **lazy consensus** to make decisions
- Contribution process
    - Contributions through feedback on broker issues
    - Feedback collected through surveys and telecons

[62] Siri Jodha Khalsa, et al., "Roadmap for Developing of Brokering as a Component of EarthCube," draft roadmap released June 2012, 27.
[63] Ibid., 33.

10.0 RISKS

- **Risks/Challenges**
  - **Limited community consensus on meaning of broker** and its functionality
  - Should require **minimum effort to advance user objectives** – ideally with no learning curve
  - Providers prefer **stable interfaces and minimize impacts of upgrades** unless there is a significant improvement in performance
  - Users avoid **making changes in their data systems** and avoid more than nominal participation in activities that would demand those changes **unless are paid** or likely to be paid
  - Brokering component could provide all the **interface adaptations** for providers and users through adapters.
    - Minimizes entry barrier to user or provider.
    - Challenge to identify provisioning for broker efforts.
  - **Major repositories could address long-tail of curve**
  - Language, geospatial representation, and cultural patterns are **barriers to science uptake** in multi-disciplinary interactions
  - Brokers need to **adapt to evolving requirements** and support broad range of users.
    - Challenge to build in sufficient flexibility
  - How can **flexibility be built into brokering component** while maintaining a practical implementation within the architecture?
- **Risk mitigation**
  - "Constructing a **glossary** of key terms
  - Working with users and providers to have a **consensus set of requirements**
  - Implementing **components that require the minimum effort** on the part of the providers and users
  - Emphasizing **provider and user buy-in** through significant outreach and educational activities
  - Including an **international perspective** from the start."[64]

[64] [64] Siri Jodha Khalsa, et al., "Roadmap for Developing of Brokering as a Component of EarthCube," draft roadmap released June 2012, 5.

CROSS-DOMAIN INTEROPERABILITY CONCEPT TEAM

1.0 PURPOSE

- Cross-domain interoperability is a process rather than a product
- Communities to be served divided into discipline, organization, role in information lifecycle, stakeholder role

2.0 COMMUNICATION

- Three focus elements
  - Raise awareness and community support of cross-domain effort
  - Ongoing assessment of needs and progress
  - Encourage interoperability practice (adoption of technical governance outputs)
- Diverse community requires different communication approaches for different audiences
  - Spectrum of size, maturity, organizational level
    - Major Earth science data projects: IOOS, OGC, WMO
    - Individual scientists and small lab groups
  - Data access points
    - Major Data Assembly Center – CUAHSI, IEDA
    - Single-project portals
  - Cross-domain users
    - Scientists with expert IT skills
    - High school students
    - Several communication mechanisms available

3.0 CHALLENGES

- **Technical challenges/needs**
  - "Specifications of basic metadata content requirements
  - Standards for encoding basic metadata,
  - Services for cataloging and transmitting metadata,
  - Services for transmitting data, and
  - Semantic interoperability of names for variables, data entities, properties, and terminological values for data."[65]
- **Cultural challenges**
    - Need incentives to contribute to community data system or use new CI
    - Need incentives to share data
    - Proper citation/credit
    - Address questions of ownership of data acquired using public funding, and under what conditions data should be withheld

[65] Ilya Zaslavsky et al., "EarthCube Roadmap," draft roadmap released June 2012, 18.

- Implement policies to
    - require publication of metadata after moratorium period
    - require direct collaboration with primary data creators of data to reuse the data
        - would increase motivation to share data, but would not lead to free and open access and reuse of data (ultimate goal)
    - foster better professional rewards for data sharing
    - improve data citation mechanisms
    - tracking of data citations (similar to publication citation tracking)
- Must consider motivations and professional requirements of data creators
- Need incentives to document data (metadata)
- Institutional changes to motivate data documentation
- Giving credit for well-documented datasets
- Making adequate documentation of datasets a requirement to publish results in scientific journals
- Identify resources or motivations to compensate data providers for additional work to accommodate EarthCube rules and regulations
- Need to fulfill needs of individual researcher to achieve

- **Technical trends**
    - Growth of data volumes and complexity of information
    - Proliferation of new scientific data
    - Higher degrees of data availability
    - Proliferation of cloud-based solutions
    - Smart sensors
    - Increasing availability and reliance on real time data streams
    - Accelerating development and acceptance of formal domain models and semantic descriptions
    - Scalability

- **Technical needs**
    - compliance with community-adopted standard information models and services
    - efficient use of cloud resources

- **Social trends**
    - More open and transparent information management
    - Use of open source software
    - Crowd sourcing of data management and analysis
    - Citizen science and community engagement
    - Increased diversity in adoption culture and technologies

- **Social Needs**
    - Community consensus on standards for data model interchange and for automated standards-based interoperability solutions to deal with rapidly increasing scale of community involvement

4.0 REQUIREMENTS

- **Different levels of scope**
  - Enterprise (purpose, scope, system policies
  - Information describes information sources and models
  - Computational (services and protocols)
  - Engineering (architecture and interactions between system components)
  - Technology (hardware and software)
- Approaches to determining Cross-Domain Readiness Requirements
  - **"The bottom-up approach** is to study existing stories of cross-domain data use: success stories, in an effort to determine what strategies are effective and what requirements would enable *similar and foreseen data reuse*, and failure stories, in an effort to determine what circumstances block effective data reuse.
  - **The top down approach** is to consider and determine the qualities of data that – independent of the specifics of use – seem to best enable *unforeseen data reuse."* [66]
- Data needs to satisfy a spectrum of needs between the **lowest common denominator** needs of large populations of scientists to **highest common denominator** needs of a spectrum of scientists
- **Community Requirements: The consensus process**
  - Need to establish community consensus about data publication and exchange protocols
  - Process should be formal, open, transparent
  - Well-defined sequence of steps and broad community involvement
  - EC should leverage existing orgs (ESIP, OGC)

5.0 STATUS

The Cross-Domain Interoperability Roadmap focuses on the following status issues:

- Initial reference model and key CI components
- Generalized architecture
- Standards-compliance of key CI components across disciplines
- Readiness measures for CI components
- Readiness evaluation with respect to standards
- Cross domain integration initiatives
- Communities and projects
- Cross domain datasets

6.0 SOLUTIONS

- **Establish criteria** on how to move toward goals of cross-domain interoperability
- Define **measurable metrics** against criteria
- **Evaluate interoperability** solutions
- Evaluate **impact on community sustainability**

Ilya Zaslavsky et al., "EarthCube Roadmap," draft roadmap released June 2012, 29.

- Evaluate **interoperability readiness** model and assessment

7.0 PROCESS

- **Governance:** Choices for data products; Standards for metadata, information models, semantics, model/workflow execution
- **Curation of :** Standards, data products, data and resource catalogs, publications, change history
- **Infrastructure provision:** Reusable data products and processing, search and discovery services, software data publication tools, user training and documentation

8.0 TIMELINE

- **Inventory of available geoscience data and other resources**: already started, will be extended in year 1
- **Readiness assessment of domain infrastructures and resources:** already started, will be extended in years 1-3
- **Cross-domain pilots:** already started, will expand in years 2-3
- **Reference Architecture and Cross-Domain Interoperability Platform Development**: has already started, will expand in years 1-5
- **Outreach/Dissemination, Governance and Community Engagement:** expected to start in year 1 and continue on an ongoing basis.
- **Interchange formats, protocols and vocabularies**

9.0 MANAGEMENT

See **governance needs and mechanisms** above.

10.0 RISKS

- **"Conflict of interest** between data center domains and owners and the goal of enabling cross-domain data reuse
- **Insufficiency of solutions,** leading to lack of cross-domain data use due to remaining difficulties
- **Misuse of cross-domain data**, leading to inaccurate scientific conclusions
- **Lack of trust** in cross-domain data solutions, regardless of efficacy."[67]

[67] Ilya Zaslavsky et al., "EarthCube Roadmap," draft roadmap released June 2012, 103.

EARTH SYSTEM MODELS CONCEPT TEAM

1.0 PURPOSE

- **Optimize engagement and participation** across geoscience and end-user communities
- **Capitalize on** existing and long-standing **model development efforts**
- Use **standards-based approaches** to coding and semantic ontologies
- Fully **engage** current and future **computational capabilities**
- ESM modeling should be **adaptable and evolutionary**
- **'Geoscience Community Modeling Framework'**
    - accommodate models of the entire Earth system and small parts of
    - build upon and integrate several existing modeling frameworks and accommodate new frameworks

2.0 COMMUNICATION

- **Strategic communications plan**
    - Awareness of EC and ESM at all states to core user and developer communities
    - Foster active engagement, early adoption, feedback in developing ESM infrastructure
    - Create community of advocates to assist project team in disseminating vision and technologies for adoption by a broad geoscience community
    - Will need to evolve, adapt, develop in conjunction with technical aspects of project
- **Identify and engage core communities**
    - End-users, technical partners, data providers, other EarthCube groups

3.0 CHALLENGES

- "Solution for ESM…should be….adaptable and evolutionary in the sense that the solution can change over time in response to the demands of the community relying on it rather than presuming to design a singular end-to-end solution that will last in perpetuity.

- **Technical challenges**
    - "Increasing specialization within earth sciences
    - models that work on one piece of the earth system without due consideration to interdependencies on other Earth System components,
    - proliferation of technology through user and value chains
    - Moore's law of exponentially increasing computational capacity all must be fully acknowledged and appreciated."[68]
    - Semantic mediation
    - Standard names supported by underlying metadata (likes, assumptions, how quantities are measured, etc.)

[68] EarthCube Earth System Models Concept Award, "EarthCube Earth System Model Coupling Roadmap," draft roadmap released June 2012, 12.

    - Very few domain-specific controlled categories
    - Goal to "create seamless, automated systems and workflows that work as intended while requiring minimal effort from users."[69]
- **Social/institutional**
    - Institutional lethargy
    - Institutional mission
    - Institutional barriers leading to differences in coding, education, training, user communities, etc.
- **Other Challenges**
    - "(1) uncertainty tracking and analysis in modeling workflows (for which major advances are still needed),
    - (2) the need to calibrate new models that have been created by assembling components, and general tools for model calibration,
    - (3) a better and more general understanding of numerical instabilities that sometimes arise when coupling models that are individually stable,
    - (4) automated use of model metadata to alert users to model component coupling opportunities as well as potential mismatches/conflicts in terms of their individual assumptions, applicable scales, etc.,
    - (5) a seldom-acknowledged tension between the typical timescales for a) funding of cyberinfrastructure projects (3 to 5 years), b) maturation and adoption of new cyberinfrastructure (4 to 10 years) and c) the need to maintain cyberinfrastructure once it has matured and been adopted (4 to 20 years)."[70]

5.0 REQUIREMENTS

- **Adherence to a 'Community Modeling' Philosophy**
    - Source code should be freely available
    - Model development should be open to broad developer and stakeholder communities
    - Thorough documentation of model development
    - Standardized and community-defined modeling use cases and assessment benchmarks
- **Earth System Model Protocols**
    - Emphasis of Extensibility

6.0 SOLUTIONS

- **"Develop and integrate a set of standards-based protocols for linking component models of a myriad of Earth System processes into a coherent, conservative, multi-scale and high-performance computing enabled modeling system."[71]**
    - "We envision this solution evolving, not as a single monolithic system, but, instead, a layered architecture system upon which component models communicate with each other

[69] EarthCube Earth System Models Concept Award, "EarthCube Earth System Model Coupling Roadmap," 13.
[70] Ibid.
[71] Ibid., 1.

through the use of multiple, standards-based coupling interfaces whose structure and function vary in accordance with the nature of the coupling to be performed….
    - The proposed solution we advocate for is not to create a single standard or framework but, instead to advocate for a coupling programming standard that works with existing and emerging coupling architectures."[72]
- **"Develop an Earth System Model coupling testbed to facilitate standards-based approaches to ESM component coupling and to support broader engagement and participation of geoscience domain scientists in community-based ESM development and end-use applications.**
    - This testbed will be a community facility which provides updated repositories or links to repositories of
        - model component and coupling architecture code,
        - code documentation,
        - documentation on data semantic protocols (e.g. information on standard variable names, units and relevant metadata), and
        - Model benchmarking use cases from which new component model code and coupling architecture code can be evaluated."[73]

7.0 PROCESS

- Develop standardized questions to obtain input on immediate and broader coupling activities
- Evolve as use cases are developed
- Timeline of initial assessment activities to articulate model testbed attributes and metrics for EC ESM coupling applications
- Inventories
- Basic documentation and evaluation framework for assessing model coupling activities

8.0 TIMELINE

- **Timeline Markers**
    - "Articulation and refinement of testbed requirements (6-9 months)
    - Compilation of leading component models and coupling architectures (6-9 months)
    - Development of quantitative metrics for system benchmarking (6 months)
    - Development of a broad class of real-world coupling exercises to serve as benchmarks (9-12 months)
    - Demonstration of a first generation coupling benchmark activities (6 months)
    - Evaluation and synthesis of benchmark activities (6 months)
    - Refinement of the overall testbed architecture (6 months)
- **Timeline Milestones**
    - Draft publication of coupling architecture and coupling testbed requirements and posting for community feedback

[72] Ibid., 23.
[73] EarthCube Earth System Models Concept Award, "EarthCube Earth System Model Coupling Roadmap," 1.

- Component model and coupling architecture code repositories into a unified, online, searchable database
- Draft publication on quantitative metrics for system benchmarking
- Online, searchable archive benchmark use case descriptions and data sets
- Acquisition and activation of a community computational facility
- Delivery and posting of revised documents on requirements and metrics."[74]
-

9.0 MANAGEMENT

- **Focused on 2-solution strategy for ESM, not for EC as a whole**
  - Rotating steering committee of developers and 'knowledgeable' users
  - Open community group/forum of participants from entire enterprise
  - Core operational staff for operational management of ESM coupling testbed center

10.0 RISKS

- **"Science and technology will proceed at a pace that is too fast to successfully integrate into a coherent coupling framework or to reliably assess.**
  - Mitigation strategy: Incentivize, through funded support, model development and support centers as well as investigator teams to participate in model coupling and benchmarking activities through the proposed ESM coupling testbed.
- **Competition among model development groups both domestically and internationally will inhibit collaborative development, integration and assessment activities**.
  - Mitigation strategy: As with #1, incentivize collaboration through specific collaboration funding opportunities to encourage synthesis and synergy in parallel lines of development.
- **Budgetary support for integration and assessment activities is often tenuous and placed at a lower priority than foundational research or more 'edgy' model development activities.**
  - Mitigation strategy: Incentivize through funding opportunities for foundational model development activities to also consider and address model coupling architecture issues.
- **In broadening access to ESMs through multi-scale, multi-language and multi- platform functionalities, there is significant risk of ESMs being misused in a variety of manners.**
  - Mitigation strategy: Through support of the testbed center, foster a sustained focus on development of standards for model metadata and a Community of Practice Resource maintained by the center that fully archives and documents coupling efforts."[75]

[74] Ibid., 27-28.
[75] EarthCube Earth System Models Concept Award, "EarthCube Earth System Model Coupling Roadmap," 29-30.

LAYERED ARCHITECTURE CONCEPT TEAM

1.0 PURPOSE

- **Interoperability** between a research environment and community resources.
- **Definition of operations** that are performed upon shared name spaces.
- **Definition of components** needed to implement a distributed operating system that supports shared program execution and shared data."[76]

2.0 COMMUNICATION

- Existing NSF data CI projects
  - DataNet, XSEDE, Big Data Solicitation, DIBBS solicitation, CUAHSI, OOI, NEON, GEON, National Water Model
- Federal agencies provide CI and international collaborations
  - NOAA, NASA USGS, EPA, EUDAT, GEOSS
- "An architecture framework that federates across all of these initiatives can be successful if the focus is on interoperability mechanisms."[77]
- **Communication** across entities will ensure interoperability
- Requires **identification of liaisons** that track interoperability requirements
- **Tracking** types of entities, operations performed upon entities, virtualization mechanisms for operations
- Use **workshops** to compare architectures and interoperability mechanisms
- **Interoperability Testbeds** can accelerate communications process (proposed by Brokering Group)

3.0 CHALLENGES

- **2 layers of architecture can be implemented in parallel**
  - Brokers for community resources
  - Data grids for collaboration environments
  - Both require interoperability mechanisms
- **Emergence of new types of objects**
  - Mapping needed between streams and files for achieve operations
- **Integration with existing systems** without loss of info about provenance
  - Brokering can mediate differences in protocols and unify access mechanisms across existing systems
- Need to **share workflows**
  - Facilitate community engagement through workflows that capture processes steps for research questions

[76] EarthCube Layered Architecture Concept Team, "EarthCube Layered Architecture Roadmap," draft released June 2012, 3.
[77] Ibid., 10.

- The knowledge needed to implement a research analysis can be captured and shared for common applications by the same or other researchers.

- Challenge to **promote use of enabling data grid technology by non-federated providers** who may not see the utility beyond currently funded projects
- Need **education and outreach** to promote adoption

4.0 REQUIREMENTS

- **Use Cases – should illustrate:**
  - "Formation of a community consensus on infrastructure needed to support research
  - Integration with existing community resources, to ensure continued access to prior research
  - Codification of tacit expert knowledge necessary to enable reproducible science
  - Support for individual researchers as well as teams of researchers in collaboration
  - Support for frameworks for model benchmarking and model comparison
  - Support for data driven science
  - Support for transitioning current research results into new community resources."[78]

- Requirements to obtain interoperability between existing systems
  - **"Infrastructure interoperability:** Collaboration environment should be able to access a community resource or service using the protocol of that resource/service.
  - **Community consensus on semantics, ontologies, and procedures:** Each community needs a reserved vocabulary for expressing their semantic terms, a well-defined set of relationships between the terms, and services that manipulate data based upon the terms.
  - **Policy consensus on data governance and data sharing** between institutions and domains. Can be turned into policies that enforce required data format and semantics on any shared data."[79]

5.0 STATUS

- **Four approaches to integrate EarthCube community resources**
  - **Pair-wise integration** of technologies between applications and community resources
  - Adoption of **standard service interfaces**
  - Development of **middleware (brokers**) that mediate a specific service across current implementations
  - **Data grids** that mediate across multiple name spaces

6.0 SOLUTIONS

- **"The components of the layered architecture are also being explored by other Concept and Community groups:**

[78] EarthCube Layered Architecture Concept Team, "EarthCube Layered Architecture Roadmap," 11.
[79] Ibid.

- The Brokering group will test interoperability mechanisms required for brokers. On each testbed, use cases should be demonstrated and evaluated for successful support.
- The Governance group is identifying both technical and managerial governance policies. These should be automated when possible and enforced within the collaboration environment.
- The Cross-domain Interoperability group is identifying the variety of protocols and services used by existing community resources. These need to be accessible through both brokers and data grids.
- The Web Services Community group is developing standards for services. These should be integrated into the brokers developed by the Brokering Group technology."[80]

7.0 PROCESS

- Four types of use cases to explore the integration of community resources and collaboration environments. The purpose of the use cases is to:
  - **Demonstrate reproducible science**.
  - **Automate data retrieval**.
  - **Integrate community resources with collaboration environments**.
  - **Integrate multiple community resources.**

8.0 TIMELINE

- Community consensus can be obtained on the following activities:
  - **Entities that are being shared and manipulated** (files, databases, streams, semantics, etc.)
  - **Web service standards** for each entity (standard operations)
  - **Brokers** (virtualization services for the standard operations)
  - **Collaboration environment capabilities**
  - **Governance** (policies and procedures)
- 1-3 years to achieve community consensus for each activity
- Development process within each community consensus activity can be implemented for each consensus area
- Developers of community resources need to match community expectations

9.0 MANAGEMENT

- **Each community should be able to implement its own requirements within a common framework**

[80] EarthCube Layered Architecture Concept Team, "EarthCube Layered Architecture Roadmap," 21.

- Requirements can be characterized and implemented as policies and procedures executed in a common framework
- "If the generic framework allows each community to enforce their own policies, the management of changing objectives will be greatly simplified."[81]
- "The infrastructure needs to provide the extensibility required for management objectives to change over time."[82]

- **Within Layered Architecture**
  - Management objective is demonstration of highly extensible systems able to incorporate new technologies while supporting policies for multiple communities
  - Management decisions focused on visualization mechanisms needed for interoperability
    - Virtualization mechanisms can incorporate new technology and enforce technical governance policies that evolve over time

10.0 RISKS

- **Risks**
  - Interoperability of old and new components within an infrastructure
  - Evolution of the infrastructure through interoperability mechanisms
- **Risk Mitigation**
  - Identify and leverage existing NSF infrastructure initiatives
  - Select robust technology
  - Identify mechanisms to support research cyberinfrastructure requirements

[81] Ibid., 23.
[82] Ibid.

WEB SERVICES CONCEPT TEAM

1.0 PURPOSE

- **Service Based Integration Platform** for EarthCube to enable community-wide commitments to sharing via standardized methods
- "SBIP-E sees its role as a group that finds and assists in making a variety of data assets available through web services."[83]
- Focus on web services
- Provide new tools to help answer specific science questions

2.0 COMMUNICATION

- No information given.

3.0 CHALLENGES

- Need to **address adoption inertia**
    - Strong communication at design and development of web services abstraction layer and interoperability platforms built on top of services between users and developers
    - Manage user expectations through ongoing communication

4.0 REQUIREMENTS

- **Comprehensive survey of existing community data holdings and currently deployed web services**
    - Emphasis on those enabling return of high-level data and knowledge products
    - EC should learn from communities who already have experience in developing web services
- **Core list of requirements for web service-based component of EC**
    - "We anticipate that the core technical requirements of the service-based component of EarthCube, in general terms, will involve
    - a common form of service description that is human and machine readable,
    - a common method of invocation based on space-time constraints, such that the same space-time arguments can be successfully passed to and understood by all involved communities, and
    - A common form of output, including the possibility of output-type specification."[84]
- **Solicit needs of geoscientists and software developers** who will use web service component of EC
- **Mechanisms to engage long tail scientists**
- **Principles**
    1. Science value over technical strategy

[83] EarthCube Web Services Concept Award, "Web Services Concept Group Roadmap," draft released June 2012, 1.
[84] Ibid., 5.

2. Strategic community-wide goals over project-specific benefits
3. Intrinsic interoperability over custom integration
4. Shared services over specific-purpose implementations
5. Flexibility over optimization
6. Evolutionary refinement over pursuit of initial perfection

5.0 STATUS

- Types of web services
    1) End-to-end data analysis and visualization
    2) Workflow actors invoked as web services
    3) Compound web services
    4) Domain-specific examples
- Gap analysis
    1) Forms of service description
    2) Patterns of invocation (with arguments)
    3) Forms of output

6.0 SOLUTIONS

- "EarthCube technologies will always evolve and therefore we can never assume that EarthCube solutions will remain static."[85]

7.0 PROCESS

- **Review processes at large, successful existing organizations** (representative across academia, government, industry) e.g. OGC.
- Review processes at **existing smaller/community-based**

8.0 TIMELINE

- **Year 1**
    - *Technical Activities*
        - Development of OGC web services as needed
        - Development of Simple Web Services at Data Centers
        - Wrapping simple web services in OGC Processing Services (WPS)
        - Work with Brokering and Interoperability Groups within the Concept Awards to develop common methods of discovery and access to information assets.
    - *Community Activities*

[85] EarthCube Web Services Concept Award, "Web Services Concept Group Roadmap," 7.

        - Establish Discussion Forums specifically related to Web Services
        - establish web service short courses in conjunction with other national and regional meetings including AGU, UNAVCO annual meetings
        - establish EarthCube standards within governance committees
- **Year 2**
    - Continue the development of simple web services within the SBIP participants
    - Development of a turnkey system to enable deployment on systems managed by individual researchers
    - Work with Governance Group toward development of the EC Organization
    - Conduct surveys of the Geosciences to identify additional sources of data available in the community willing to expose data assets through Web Services
    - international coordination with European and Asian data centers in geodesy and
    - seismology to produce federated web services
- **Year 3**
    - Engage broader international geosciences community

9.0 MANAGEMENT

- No information given.

10.0 RISKS

- No information given.

APPENDIX 3: EARTHCUBE WORKING GROUP AND CONCEPT TEAM GOALS AND GOVERNANCE NEEDS

INTRODUCTION

This document reviews the goals, milestones and governance needs of each Working Group and Concept Team drafted roadmap (submitted prior to the June 2012 EarthCube charrette). Some of the governance needs were explicitly stated by the roadmap's authors; others we had to extrapolate. Additional materials (if provided by each roadmap) include the proposed internal governance framework (governance framework centered on that group's specific area of interest), reference architecture as it relates to governance.

The governance wants and needs from the draft roadmaps form the foundation of the EarthCube Initial EarthCube Governance Functions section of this document (See Section 3.1: EarthCube Governance Functions) A master list of these governance functions, and others gleaned from community engagement and research on governance, is presented in Appendix 4: EarthCube Governance Functions Master List.

Building an initial governance framework for EarthCube will require a deeper understanding of the governance wants and needs of the roadmaps, and how they interact with each other, in addition the many other stakeholders and communities not represented by these roadmaps. Thus, efforts will be made to follow up with each Working Group and Concept Team to better understand what they expect of EarthCube governance, in addition to an aggressive community engagement plan, in order to vet and refine the draft Governance Framework with the EarthCube community.

DATA DISCOVERY, ACCESS AND MINING WORKING GROUP

GOALS AND MILESTONES

ROADMAP GOALS

1. "To provide a timely and adequate response in terms of requirements, technology and policy issues to the challenge of ever-expanding practices of handling and managing data." [86]
2. Provide "robust infrastructure for distributed, discoverable data access and provide a core set of analytics capability." [87]

DataSpace Goals
"EarthCube DataSpace is our prominent vision of the landscape that will bring under a single umbrella the big-head of science and the long-tail geoscientists. It provides long-tail scientists with an easy way to upload, share, discover and analyze their data; It interfaces with the big head of science, both, in terms of bringing their data on a per-use-case need basis, and by leveraging data management practices followed by them over several years."[88]

- **Characteristics**
    - "Provide the much-needed **low-barrier for entry** and simplified data management by providing uniform services for storage, access, curation, discovery and analytics;
    - **Enable and democratize the community** by specifying a set of complaint capabilities and services which anyone can implement;
    - **Provide feedback in terms of readiness levels in** assessing a tool for compliance with existing capabilities and services, to enable technology insertion, especially in the changing data technology landscape;
    - Provide a **social glue** for the EarthCube community by encouraging dialogue and communication on issues." [89]
- Framework for data discovery, access, mining

DATASPACE MILESTONES

"…bring end-users and resource-constrained scientists at par with big heads of science." [90]
- Interactive publications
- Workbench of specialized analytics/visualization tools
- Learning nodes and examples

[86] EarthCube Data Discovery, Access and Mining Working Group, "A Community Roadmap for EarthCube Data," draft 1-2.
[87] Ibid., 3.
[88] Ibid., 3.
[89] Ibid., 3-4.
[90] Ibid., 9.

- Fully exploit big data and enable data-driven discoveries
- Best practices as an element and outcome to establish an interoperable data space
- Community engagement at existing community events (AGU, ESIP, etc.)
- EarthCube Data Coordination Committee
-

GOVERNANCE

GOVERNANCE NEEDS

- **Processes:** (human- best practices, patterns, etc. and automated – systems, algorithms, etc.) to add assurance, provenance, reproducibility to data preparation phase
- **Transparency and openness** to cross cultural and institutional boundaries
- **Mechanisms to make tools sustainable** (tools developed rapidly to fill a specific need and then seen to generally useful by scientists and students who understand them)
- **Consider sustainability from multiple angles**
- Decide **what it means to publish data**
    - Need **clear expectations**
    - May need to define multiple levels of releasing data to DataSpace with higher 'readiness' scores
- **Risks identification and mitigation**
    - Working groups will represent stakeholders
    - Open communication
    - Respond to technological advances and changes in technology
- **Address data issues for big head and long-tail scientists**
- **Assess need for data mining tools:** Map existing tools to functional categories and data structures
- Community discussion on how **data mining is related to licensing, standards, interoperability, semantics, brokering, governance, workflows**

PROPOSED DATA-SPECIFIC GOVERNANCE FRAMEWORK

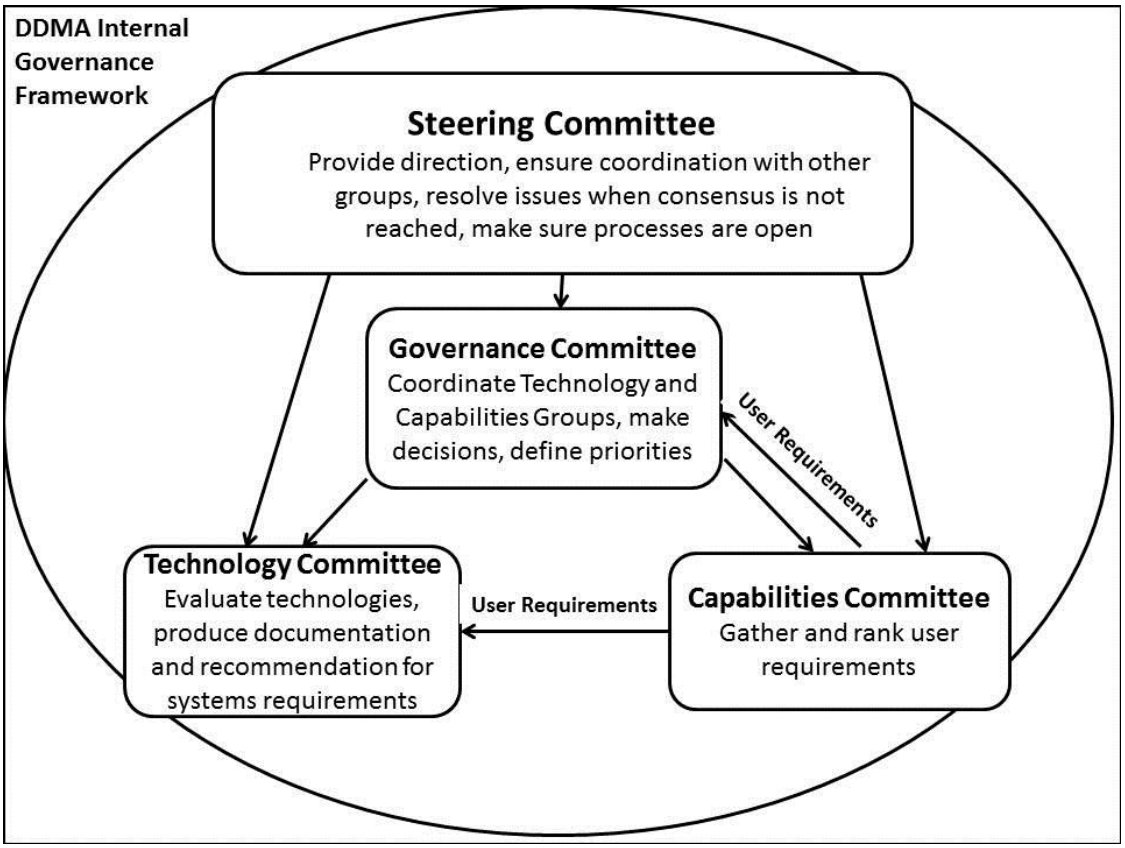

FIGURE 19. PROPOSED DATA-SPECIFIC GOVERNANCE FRAMEWORK

CAPABILITIES SUBGROUP

- 'continuous' requirements gathering activity
- Ongoing cycle to solicit, analyze and rank community needs
- Poll domain scientists
- Use cases – seek, analyze, categorize
    - Help form vision of seamless data discovery, access, mining capability
    - Derive functional requirements and map to formal technical requirements
- Gather user needs and make ranked list of requirements (according to degree of community preference)
- Seek input from data stewardship representatives and librarians to ensure data is managed properly
- Data centers
- NSF – concern towards furthering a particular scientific agenda

TECHNOLOGY SUBGROUP

- Evaluate technologies to map to functional requirements

- Identify gaps in coverage of needed capabilities and recommend how to fill them
- Enlist Capabilities members in user testing
- Set of readiness level definitions to asses technology's readiness for EC adoption
- Produce ongoing documentation, recommendations and DDMA systems requirements
- Evaluations fed back into EC development efforts

GOVERNANCE SUBGROUP

- Supporting role
- Coordinate Requirements and Technology subgroups
- Focus on data-specific governance issues, such as defining priorities
- Final decision-making body regarding requirements and other important DDMA decisions
- Prioritize ranked requirements, Integrate gap analysis, overall strategy, etc.
- Create a Technology Roadmap: living document to handle change

SEMANTICS AND ONTOLOGIES WORKING GROUP

GOALS

- "Reasoning and knowledge discovery (knowledge availability (versus data availability)…
- Complex analytics with information interoperable across all geoscience communities
- Improved data discovery technologies triggering increased community participation (long tail of science) for data and tool sharing."[91]

GOALS (6 MONTHS – 1 YEAR)

- "Continue and increase communication with other EarthCube groups, e.g., brokering, workflows, interoperability, layered architecture, data mining that also need semantics as part of their solutions
- Organize Semantics and Ontology workshops to identify geoscientists and semantics experts who would be charged with the task of starting the process of cataloging existing geoscience related ontologies, and their application in demonstration projects developed around the use case already identified in this report
- Conduct a Use for semantics technologies as applied to geoscience use case
- Support discussions through workshops with geoscientists to identify additional use cases
- Start creating initial foundational, domain and service ontologies, including Ontology Design Patterns for use in demonstration projects. Ontologies to be jointly created with domain experts and semantic/ontology scientists/engineers." [92]

MID-TERM GOALS (2-4 YEARS)

- "Continue with the short term goals
- Refine further use cases
- Enhance initial ontologies for data and services
- Develop mapping across common vocabularies and ontologies
- Develop demonstration projects utilizing geoscience community endorsed use cases
- Work on provenance and ontologies
- Establish and populate ontology repository; identify process of sustainability of all resources
- Begin putting geoscience data into linked format (RDF)
- Work with the brokering, workflows, interoperability, layered architecture, and other EarthCube groups to include semantics and semantic components
- Develop tools and other software related to semantic capabilities
- Begin to show increased data discovery, access, and interoperability
- Increase community engagement, including the long tail
- Develop and disseminate educational materials

[91] Krishna Sinha et al., 16.
[92] Ibid., 15.

- Engineer semantically enabled data, service and discovery engines to handle use cases" [93]

LONG TERM GOALS

- "Reasoning and knowledge discovery (knowledge availability (versus data availability) as per section 1)
- Complex analytics with information interoperable across all geoscience communities
- Improved data discovery technologies triggering increased community participation (long tail of science) for data and tool sharing"[94]

GOVERNANCE

PROPOSED SEMANTICS AND ONTOLOGIES GOVERNANCE FRAMEWORK

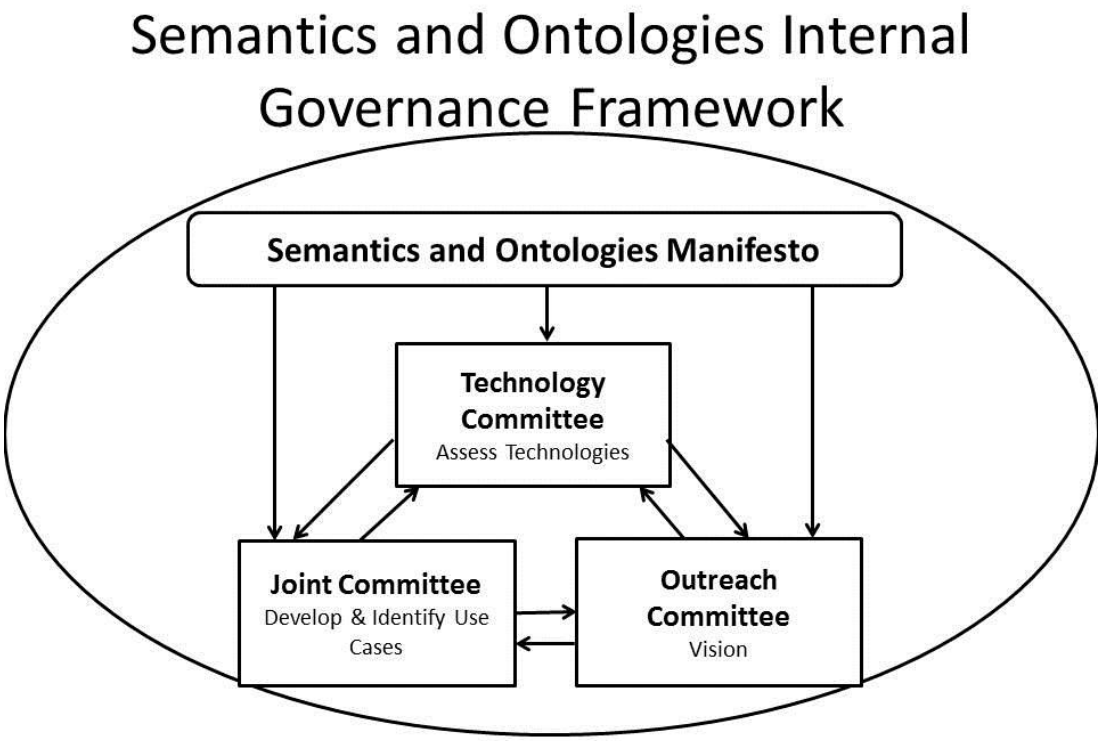

FIGURE 20. PROPOSED SEMANTICS AND ONTOLOGIES GOVERNANCE FRAMEWORK

- **Governance Framework**
  - **Semantics and Ontologies Manifesto:** Explain importance of semantics in data discovery and integration, to be distributed to geoscience community

[93] Ibid. 16.
[94] Krishna Sinha et al., 16.

- **Outreach Committee :** charged with sharing the vision and capabilities of the semantics/ontology group with other communities (e.g. ESIP/federal agencies; geoscience societies)
- **Technology Committee**: charged with assessment of current and future developments in semantic technologies
- **Joint (geoscience and technology) Committee**: charged with developing use case(s) and to identify selected use cases with a goal of delivering semantically enabled solutions to the community. This working group is also charged with identifying where infused semantic solutions directly address gaps in the science communities.

Proposed Semantic/Ontologic Based Infrastructure for Geosciences

WORKFLOW WORKING GROUP

GOALS

- "…make **workflows ubiquitous within the geosciences** and to further develop or enhance the workflow tools to meet the needs of geoscience."[95]
- **Demonstrate workflow benefits** to geoscientists
- **Create broad community within geosciences to identify short and long term problems for scientific workflows.**
    - Education and outreach
    - Understanding of different types of workflows
    - Better collaboration between workflow software developers and geoscientists
    - Identify gaps
    - Vision for grand challenges that no current workflow technology can address
- **Major achievements**
    - Establish divers and interdisciplinary Steering Committee
    - Establish and operationalize a Workflow Synthesis Center
    - Establish and operationalize Workflow Group Task Forces
    - Create an online workflows community with active member participation

GOVERNANCE

GOVERNANCE NEEDS

- An effective organizational structure to enable efficient strategizing
- An effective operational structure to ensure smooth and timely operational activities
- Effective processes for creating groups, organizations, etc.

[95] Yolanda Gil, et al., 2.

- Effective processes to facilitate consensus and enable efficient decision-making

WORKFLOW-CENTRIC GOVERNANCE FRAMEWORK

The following diagram is a visual representation of the workflow-centric governance framework proposed by the Workflow group in the Workflow Roadmap.

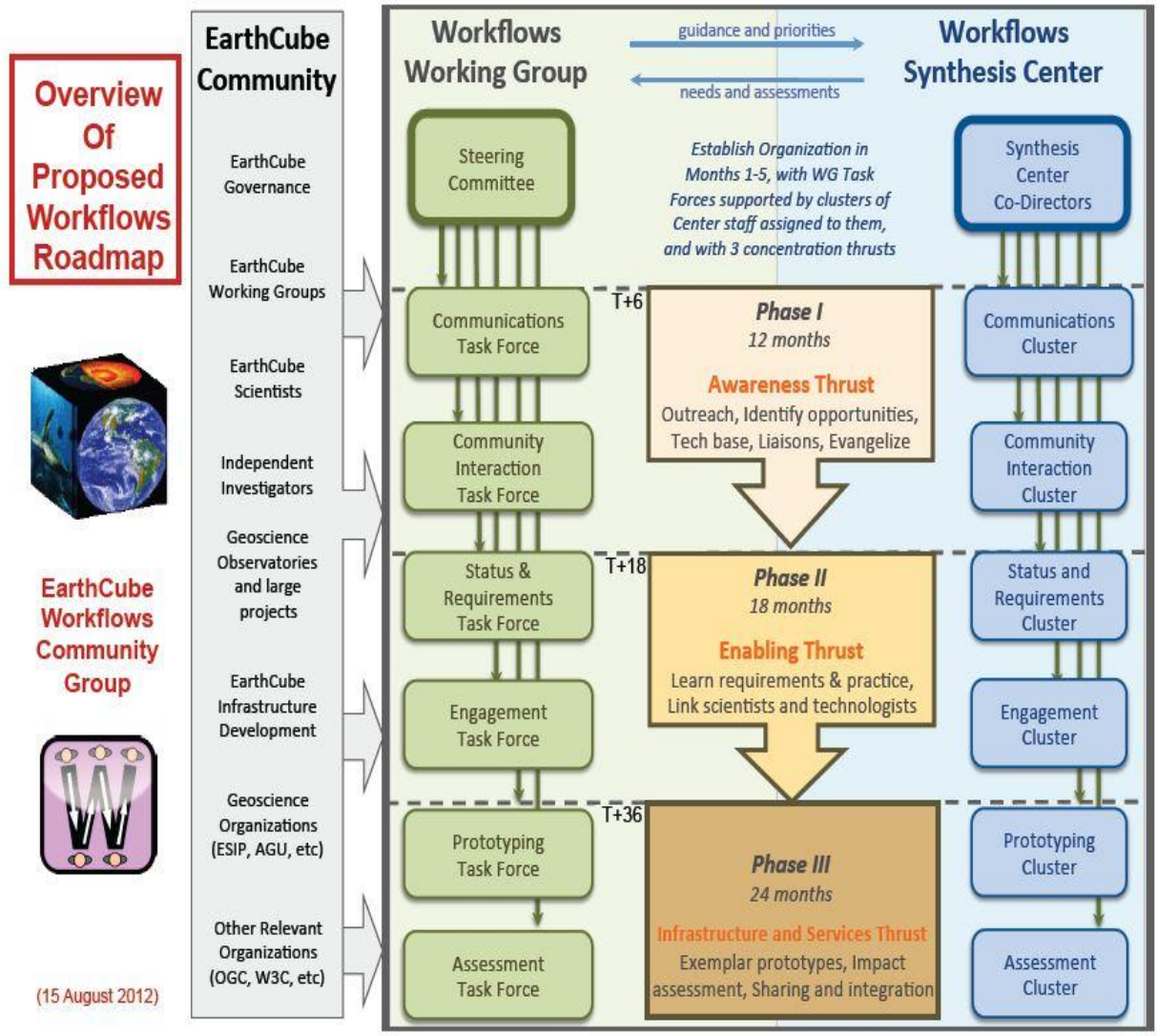

FIGURE 21. PROPOSED WORKFLOW-SPECIFIC GOVERNANCE FRAMEWORK FOR EARTHCUBE.[96]

[96]EarthCube Workflows Community Group Roadmap, Version 2, released August 2012.

WORKFLOWS SYNTHESIS CENTER

- Facilitate activities and operational tasks
- Full-time programming staff and post-doc students
- Leverage existing organizations to create a distributed center strongly connected with tight communication mechanisms
- Domain and IT expertise represented
- Tasks:
    - Develop workflow material, tutorials, use cases
    - Attend geoscience meeting and evangelize the workflow community mission
    - Help geoscience groups prototype workflow applications
    - Organize and conduct summer/winter schools
    - Create and maintain FAQs
    - Collect and track use cases, success stories, feedback from end users
    - Organize outreach workshops
    - Develop approaches to 'workflow citation,' link to publications, etc.
    - Recommend to Steering committee creation of new task forces and merging old ones
    - Leverage and participate in activities of ongoing groups
    - Reflect on past initiatives and analyze impact on workflows
    - Coordinate with government organizations
    - Coordinate with other EC groups

STEERING COMMITTEE

- Carry out Workflow goals
- Open community model
- Mechanisms to integrate new members
- Leverage online tools
- Consider and blend recommendations from Governance Working Group
- Flexibility to allow members to take initiative to address problems
- **Decision-Making Processes**
    - Public vote on an open SC mailing list
    - Each SC member has one vote
    - 72 hour voting period
    - Decision is made when quorum is reached and number of positive votes greater than negative votes
    - Decisions communicated through open, archived discussions on mailing lists

TASK FORCES

Communications Task Force

- Disseminate information to the community through

    - Web presence with accessible information for geosciences research
    - Clear points of entry and participation to engage newcomers
    - Pursue synergistic and opportunistic meetings for virtual and physical participation
    - Measure community engagement and growth over time
- Organize and actively participate in community-wide EC workshops and other community events
    - Define goals and expected products from a workflows session at a workshop
    - Prepare materials to facilitate engagement
    - Establish strategic and sustainable lines of communication within community
    - Leverage existing events (AGU, XSEDE, etc.)
- Pursue publications
    - Submit regular reports in newsletters and journals
    - Encourage participants to publish accomplishments
    - Collect and make available published materials

Prototyping Task Force

- Pursue workflow technology prototypes in various geoscience domains that
    - Demonstrate workflow capabilities for doing science
    - Gain scientific advocates
    - Identify difficulties in adoption and need for outreach
    - Disseminate expertise
    - Create opportunities for interactive educational material for students in various geoscience domains
    - Produce scientific advances that would not occur as rapidly or easily without workflows

Interaction Task Force

- Pursue and coordinate interactions with major funded projects, standards agencies, community groups in geosciences to:
    - Uncover use cases and opportunities to disseminate workflow technologies
    - Collect success stories and lessons learned, future requirements, challenges, etc.
    - Align success stories, lessons learned, etc. with Workflows Roadmap
    - Leverage and participate in events where larger communities already participate

Status and Requirements task force

- Assess current workflow technologies in various domains
- Evaluate available workflow technologies in the commercial sector
- Obtain and understand requirements of geosciences community
- Investigate and compare overlapping workflow technologies
- Assess trends and distill shifts in technologies related to workflow
- Disseminate information in geoscience workflow needs and state of the art workflow technologies

Engagement Task Force

- Work from Status and Requirements Task Force is passed to Engagement Task Force
- Provide guidance to geoscientists to identify ways to address their workflow needs
- Assist scientists to evaluate potential workflow technology solutions
- Get support from Status and Requirements Task Force and the Prototyping Task Force when necessary
- Disseminate expertise in workflow solution approaches
- Develop "Workflow Capability Maturity Model"
- Could define process to identify an approach to introduce workflow orchestration and automation capabilities
- Define levels of maturity to guide directions of growth in workflow orchestration and automation
- Tool to define organizations progress to effectively use workflows in their domain-specific goals

BROKERING CONCEPT TEAM

GOALS

- **Goal to identify path forward to build a CI that best serves the science communities and allows them to be more productive and capable within and between disciplines.**
- **Principles**
    - Collaboration, flexibility, iteration
- Goal to use **spiral approach – iterative process** to deliver services, get feedback, and expand the cyberinfrastructure further.
- Brokering is **one element of architecture**
    - **Integration** with other elements is essential

GOVERNANCE

GOVERNANCE NEEDS

- Need governance to help make architectural decisions
    - Centralized vs. distributed brokers
    - Brokers tailored to communities or a single broker to serve all communities
- Governance issues
    - Existing operators vs. new organizational entities
    - Threshold at which a new format or protocol has become popular enough that it should be added to existing brokers
- Decisions should be shaped by technological trends and community practices

BROKERING MANIFESTO

"Preamble: The goal is to enhance the ability to predict and understand the Earth system by increasing the capability and productivity of science.

Whereas (problem statement)

- Earth system science data are extremely diverse (in format, description, scale, precision, etc.)
- Different data users discover, access, interpret, and manipulate these data with a wide variety of tools and services that may be specific to the data and each user's needs.
- Interdisciplinary science requires people to access and use data from very different disciplines and communities.
- Data providers are increasingly expected to serve communities outside their normal clientele.
- Both users and data providers experience impediments in achieving multi--- disciplinary interoperability.
- Connectivity for multi---disciplinary interoperability should not restrict the autonomous nature of discipline---specific systems.
- Technology evolution and maintainability is a great challenge for cyber--- infrastructure sustainability and usability.
- Brokers should be capable of improving and facilitating access to both "big data" and "long tail" resources, in real time or static environments

We believe (assertions)

- No single information technology or standard will serve all user needs.
- Data services (discovery, access, processing, semantics) should be openly accessible in well--- defined, machine—interpretable ways.
- Cyberinfrastructure should maximize data and service usability for both providers and users (end point use).
- Flexibility is necessary for incremental expansion of data service mechanisms and to easily achieve scalability.
- Brokers can be an effective way to achieve this flexibility particularly when multiple brokers with different characteristics are included in a brokering framework.
- A brokering framework provides the cyberinfrastructure that allows providers and users to better take advantage of the open services.
- A Brokering framework is best managed and maintained by the cyberinfrastructure (rather than Users or Data Providers).
- A Brokering framework should be transparent to Users and Data Providers.
- Not all existing systems will continue, but it is more sustainable to supplement not supplant systems mandates and governance arrangements"

Definition of Brokering Framework:

- Brokers are middleware interconnecting client and server components in the EarthCube cyberinfrastructure.
- Brokers are Services facilitating the run---time interconnection (sharing of resources) among users and providers in a way that requires little effort on the part of either.

- A brokering framework can consist of multiple brokering components to support different capabilities.
- A Brokering framework may include discovery, semantic and natural language enhancements, data access, processing and publishing."[97]

BROKERING REFERENCE ARCHITECTURE

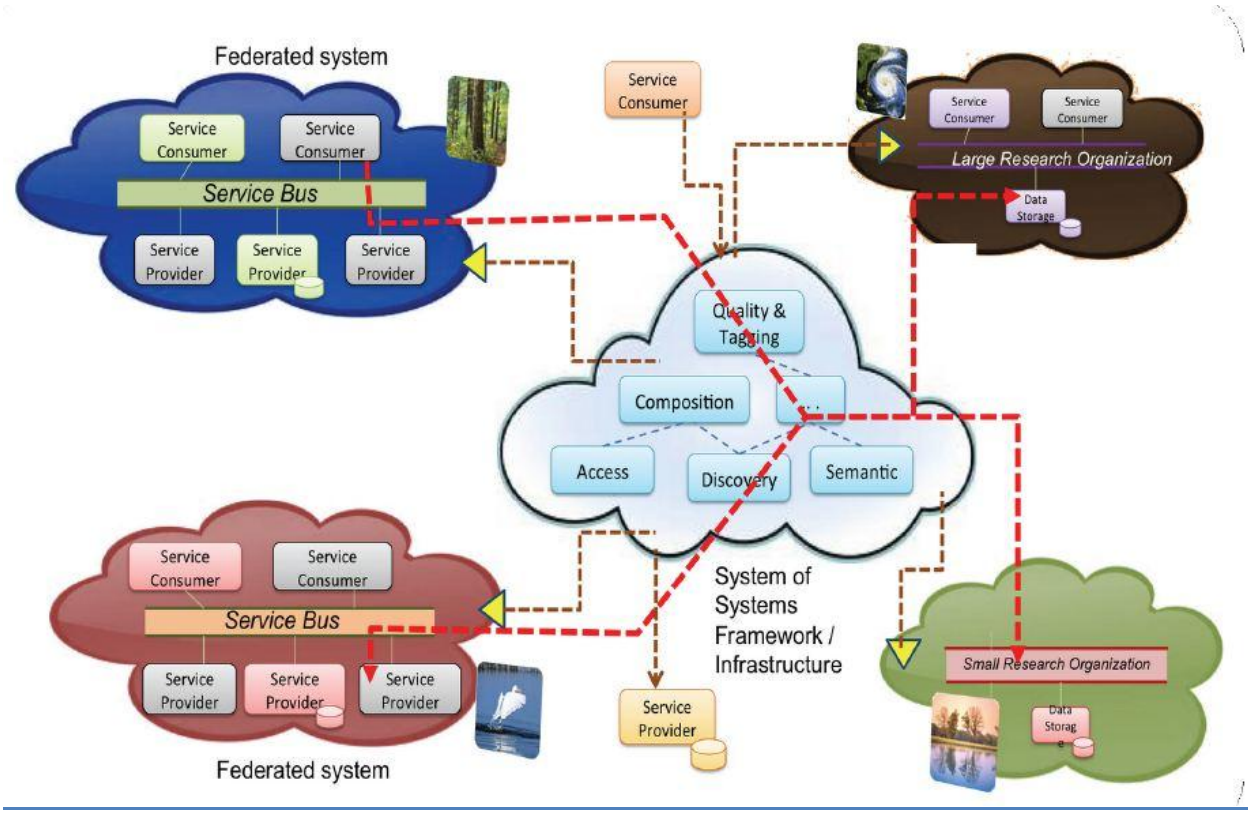

FIGURE 22. BROKERING FRAMEWORK ARCHITECTURE COMPONENTS AND INTERFACES[98]

The updated Brokering Roadmap (released August 2012) describes the proposed reference architecture in this way:

> "This figure illustrates the flow of information in a brokering framework. The dashed red lines show a canonical use case of a service consumer accessing data from a service provider. The discovery of the service provider occurs when the service consumer performs a query against the brokering framework. Then, initiated by the service consumer, the broker accesses the data from the discovered service provider on behalf of the service consumer. This may involve accessing other requested data from research

[97] Siri Jodha Khalsa, et al.,, "Roadmap for Developing of Brokering as a Component of EarthCube," 3-4.
[98] Siri Jodha Khalsa, et al., "Brokering for EarthCube Communities: A Road Map," updated roadmap released August 2012, 20.

organizations and further processing of all the data by the broker for semantics, quality, etc. prior to being made available to the service consumer." [99]

CROSS-DOMAIN INTEROPERABILITY CONCEPT TEAM

GOALS AND MILESTONES

GOALS

- "Enable researchers and students to combine information from different domains via a system of shared standards-based software environments that facilitate discovery, interpretation, access and integration of data…
- Facilitate and promote the formation of cross-domain teams, and foster better understanding and communication within such teams
- Educate new Renaissance-type scientists for whom cross-disciplinary research is the norm, supported by a reward system for such scientists."[100]
- Advance improvements in state of the art in cross-domain interoperability.

KEY MILESTONES

- "Development of a cross-domain interoperability readiness assessment procedure
- Application of the assessment to the current NSF geoinformatics portfolio
- Collection and documentation of cross-domain use cases
- Gap analysis: development of requirements and comparison with current capabilities
- Specification and demonstration of cyberinfrastructure (CI) components supporting cross-domain interoperability, including fitness-for-use assessment
- Iterative implementation of the CI components, including community validation and testing
- Organization of a continuous cross-domain interoperability testbed process that bridges needs of geoscience users with advanced technical solutions
- Development, management and curation of a cross-domain interoperability platform and content supporting the platform
- Validation of the platform and its content in a series of research scenarios
- Ongoing collaboration with other EarthCube groups to avoid duplication of effort and assure a system of interoperable components to implement EarthCube function" [101]

[99] Ibid.
[100] Zaslavsky et al., "EarthCube Roadmap," draft roadmap released June 2012, 5.
[101] Ibid., 2.

ACTION ITEMS

- "Formulate evaluation metrics for interoperability readiness, and identify gaps, development priorities, and risks for cross-domain infrastructure development.
- Identify tools and workflows necessary to capture a high percentage of data, use cases, and innovative ideas produced by modestly funded research projects that constitute "the long tail of science."
- Enable and demonstrate relevant, community-agreed, standards-based interoperability models for selected aspects of catalogs, vocabularies, services and information models.
- Create a prototype data integration platform designed to bridge user needs with advanced standards-based technologies to enable data reuse for new applications.
- Articulate differences in research paradigms, accepted norms of scientific explanation, patterns of organization, data collection and sharing practices, and the interrelated technological strategies and governance arrangements these dictate.
- Establish a community-guided process to identify cross-domain use cases, capability gaps, and development priorities that integrates technological advances with community adoption and broad engagement." [102]

GOVERNANCE

GOVERNANCE NEEDS

Scientific and Technical Governance

- Scientific governance to determine needs and priorities.
- Technical governance to decide upon implementation plans, including standards for data format, services, and discovery services.

[102] Zaslavsky et al., "EarthCube Roadmap," 6.

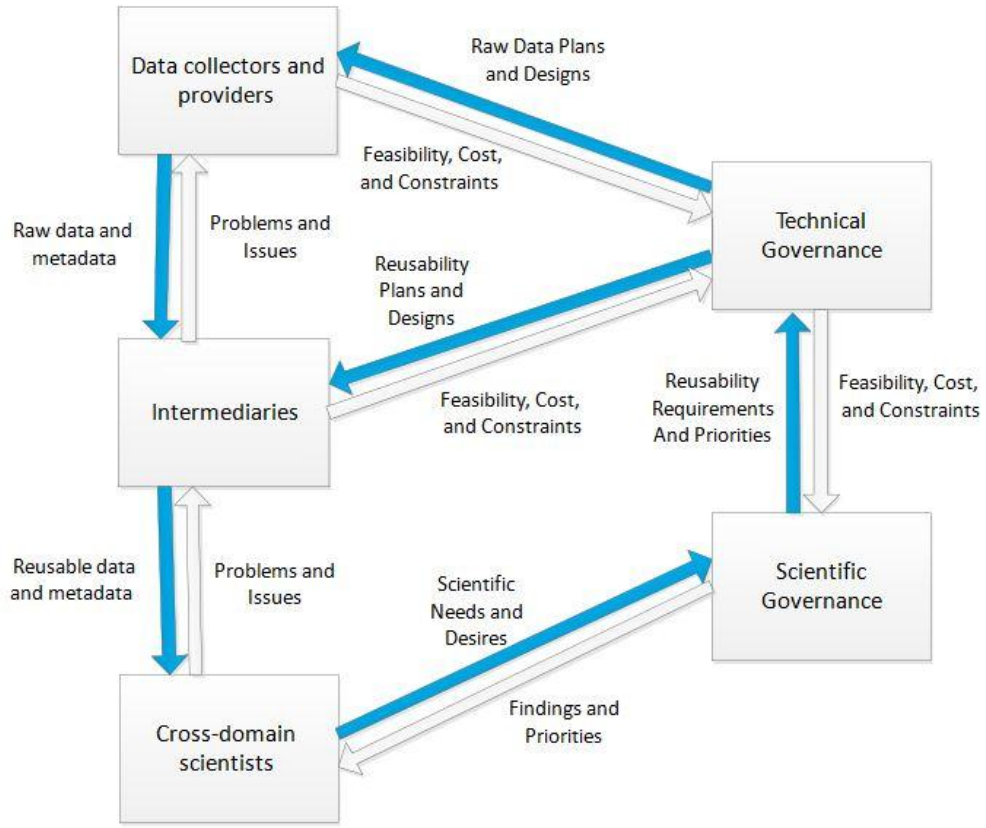

FIGURE 23 : COMMUNICATION PATHS TOWARD CROSS-DOMAIN INTEROPERABILITY.
BLUE ARROWS REPRESENT PRIMARY COMMUNICATION PATHS, WHITE ARROWS REPRESENT RESPONSE AND FEEDBACK MECHANISMS.[103]

COMMUNITY

- Interoperability platform is to bridge to cross-domain research scenarios and applications, and advanced computing technology pursued by other EC groups
- Interoperability agenda and technical programs should be complementary to other groups

DATA INCENTIVES

- Need incentives to **contribute** to community data system or use new CI
- Need incentives to **share data**
  - **Proper citation/credit**
  - Address **questions of ownership** of data acquired using public funding, and under what conditions data should be withheld

[103] Zaslavsky et al., "EarthCube Roadmap," updated roadmap released August 2012, 19.

        - Implement **policies** to
            - require **publication of metadata** after moratorium period
            - require **direct collaboration with primary data creators** of data to reuse the data
                - This would increase motivation to share data, but would not **lead to free and open access and reuse of data (ultimate goal)**
        - foster **better professional rewards** for data sharing
            - improve data citation mechanisms
            - tracking of data citations (similar to publication citation tracking)
        - Must consider motivations and professional requirements of data creators
    - Need incentives to **document data (metadata)**
        - Institutional changes to motivate data documentation
            - Giving credit for well-documented datasets
            - Making adequate documentation of datasets a requirement to publish results in scientific journals
            - Identify resources or motivations to compensate data providers for additional work to accommodate EarthCube rules and regulations

ADDRESS RISKS

- "**Conflict of interest** between data center domains and owners and the goal of enabling cross-domain data reuse.
- **Insufficiency of solutions,** leading to lack of cross-domain data use due to remaining difficulties.
- **Misuse of cross-domain data**, leading to inaccurate scientific conclusions.
- **Lack of trust** in cross-domain data solutions, regardless of efficacy." [104]

OUTREACH/DISSEMINATION, GOVERNANCE AND COMMUNITY ENGAGEMENT

"The **goal** of this activity is to ensure that **broad geoscience audience is closely involved in all phases of EarthCube CI design and development**, from feeding the development with use cases and providing feedback on implementation to participating in software development teams. The components include:

- **Identifying and nurturing EarthCube communities of practice (CoP)**, in particular:
    - CoP meetings to inform development priorities;
    - CoP meetings for User Interface design feedback and user input for incremental releases;
    - Road shows at professional meeting
    - Organizing a bi-annual EarthCube Interoperability Conference, and interoperability workshops (twice a year) focused on pilots and use cases, as well as technology demonstrations

[104] Zaslavsky et al., "EarthCube Roadmap," draft roadmap released June 2012, 103.

- **Developing governance plans, in collaboration with the EarthCube Governance team:**
    - Elaborating charters for Cross-domain advisory board and committees
    - Implementing the governance structure, in particular the establishment of separate technical and scientific committees
    - Ensuring that the governance processes are in place and active, and evaluate efficiency

- **Establishing a web presence, including:**
    - Defining policies for participation, initiation and management of blogs, wikis, mail lists, etc.
    - Website development and maintenance
    - Development of online tools for user annotation and crowdsourcing
- **Creating documentation, tutorials, and workshops, and setting up a help desk**

- **Conducting annual reviews of emerging solutions**
    - assessing them in terms of interoperability capacity, technical and community sustainability
    - making recommendations for priorities in interoperability research
    - refining and adapting capacity and sustainability metrics

- **Developing business plans for data preservation and data/model management** (in particular, in conjunction with NSF-mandated Data Management Plans)

- **Developing curricular materials**; accepting and curating contributions of relevant educational materials
- **Developing international collaborations."**[105]

LINKS TO OTHER EARTHCUBE GROUPS

- Diverse EC community – requires different approaches for different audiences
    - **DDMA:** central to cross domain use cases
    - **Layered Arch.** – emphasize layers in stack specifically needed to support cross-domain integrations: Catalogs that can be federated for cross-domain search
    - **Brokering –** need brokering services to provide mappings between common protocols and encodings
    - **Web Services** – "vocabulary services that can support interpretation of measurements across domains, and data access services working across datasets from different domains."[106]
    - **Earth System Models** – "earth systems models often represent materialized and well thought out cross-domain use cases." [107]

[105] Zaslavsky et al., "EarthCube Roadmap," 96-97.
[106] Zaslavsky et al., "EarthCube Roadmap," 8.
[107] Ibid.

- **Governance:** "key foundation for cross-domain interoperability, as mediation between different domain data systems and community-guided interaction between data providers, consumers and intermediaries require specialized governance arrangements." [108]
- **Workflows:** "key component in the analysis of domain readiness using replicable and repeatable processing chains to assess fitness for use. A cross-domain interoperability platform would provide shared expertise and an environment where these technologies can be further explored, adapted and applied in cross-domain research scenarios." [109]

GATHERING CROSS-DOMAIN INTEROPERABILITY REQUIREMENTS

Bottom-Up vs. Top down Approaches

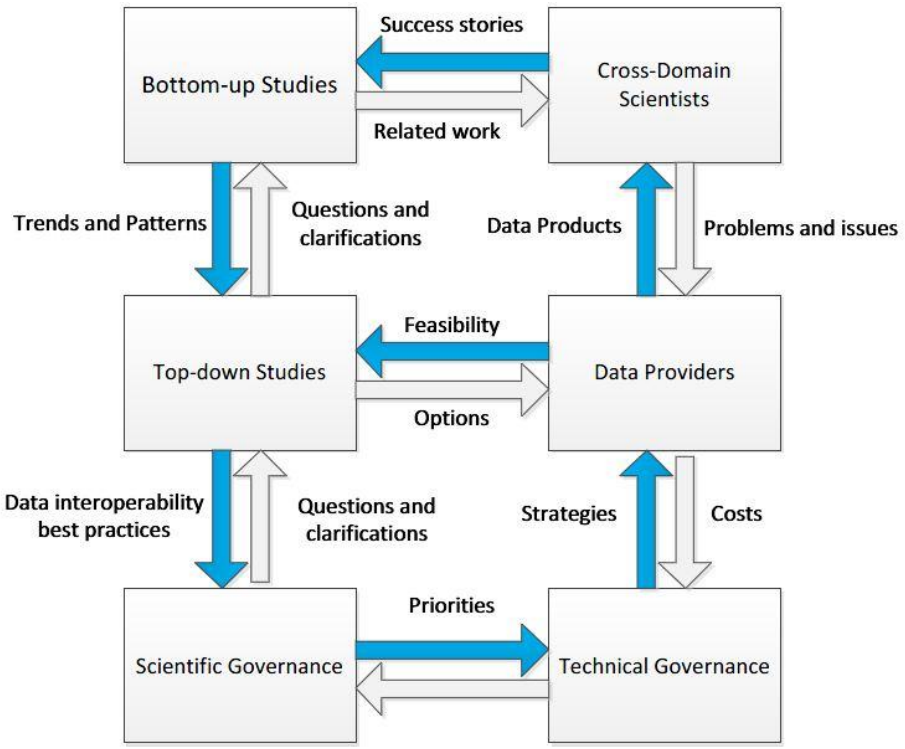

FIGURE 24. REQUIREMENTS GATHERING AND FULFILLMENT FOR CROSS-DOMAIN INTEROPERABILITY[110]

[108] Ibid., 7.
[109] Ibid., 7.
[110] Zaslavsky et al., "EarthCube Roadmap," updated roadmap version released August, 2012, 30.

CROSS-DOMAIN INTEROPERABIILTY GOVERNANCE FRAMEWORK

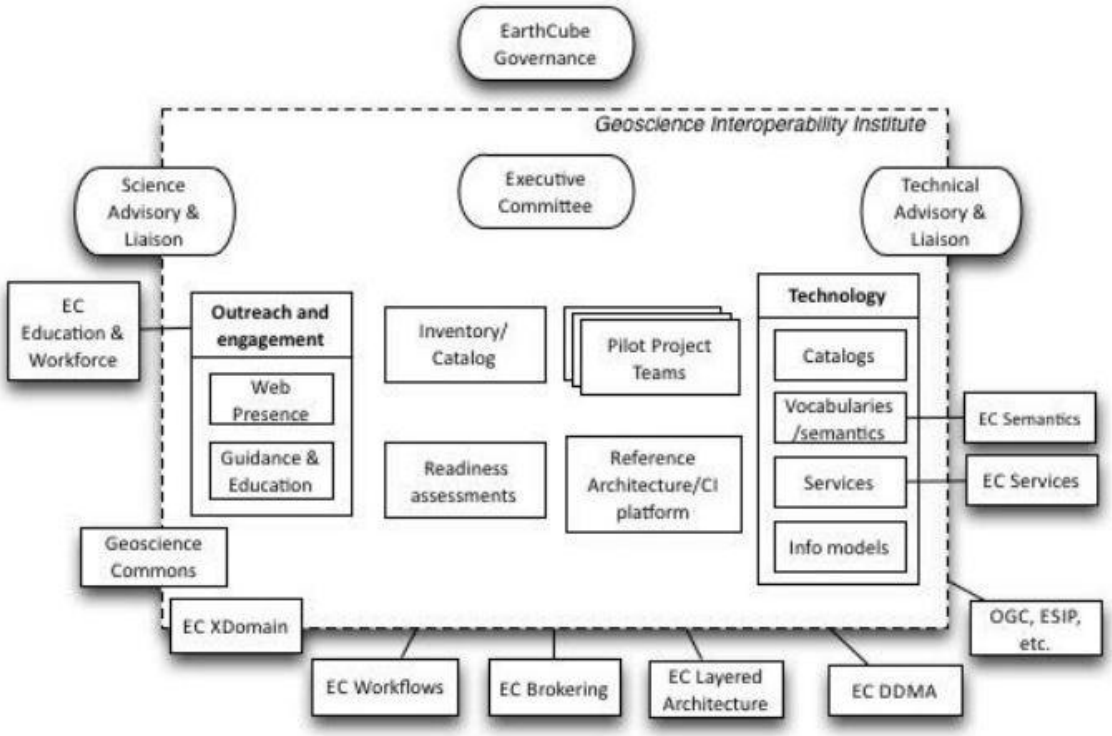

FIGURE 25. MANAGEMENT FUNCTIONS FOR CROSS-DOMAIN INTEROPERABILITY PROJECT[111]

Geoscience Interoperability Institute (GII)

GII Principles include:

- Governed by dialogue between scientists who determine priorities and IT experts who provide for those needs
- Study problem and propose solutions and standards
- Interoperable data sources via data brokering
- Implement best practices for data sources and metadata

GII tasks include:

- "Maintaining an inventory of geoscience data sources, other resources, and models of interest
- Maintaining an inventory of information system architecture diagrams
- Assessing each resource according to a readiness model described in the report

[111] Zaslavsky et al., "EarthCube Roadmap," updated roadmap, 106.

- Performing pilot deployments of new approaches as "proofs of concept", which are then evaluated for inclusion in a set of "best practices."
- Developing a reference architecture and cross-domain interoperability platform.
- Developing a governance plan for the above in collaboration with EarthCube governance.
- Developing interchange data formats, protocols, and vocabularies.
- Ongoing outreach and community engagement."[112]

ADDITIONAL COMMITTEES

Executive Committee

- Responsible for GII and GII outcomes
- Composition
- PIs and co-PIS of funded GII grants (contractual obligation to produce results
- Ex-officio representatives of Science and Technical Committees
- Additional important members
- Should be community-driven

Science Advisory and Liaison Committee

- Determine user needs and priorities
- Set overall science direction and priorities
- Responsibilities
- Advise working groups focused on meeting science needs or requiring input from science users
- Membership from geoscience projects, disciplines, organizations, and domains

Technical Advisory and Liaison Committee

- Decide on implementation plans (standards for data formats, services and discovery aspects)
- Responsibilities
- Advise Executive Committee
- Advise working groups most focused on technology (reference architecture, cross-domain interoperability, platform development, outreach and education, interchange formats, protocols, vocabularies)
- Membership from technical projects, organizations, domains

Additional **Working Groups** to address

- Readiness assessment of geosciences infrastructures and resources
- Cross-domain pilots

[112] Zaslavsky et al., "EarthCube Roadmap," draft roadmap, 3.

- Carried out by small, multi-disciplinary teams of experts
- Reference architectures and cross-domain interoperability platform development
- Outreach/dissemination and community engagement in conjunction with the Technical Committee
- Interchange formats, protocols and vocabularies

Liaisons

- Designated liaisons will identify touchpoints and potential synergies with other EC groups
- Groups outside of EarthCube (OGC, ESIP, etc.)
- Government agencies

SCOPE OF EARTHCUBE

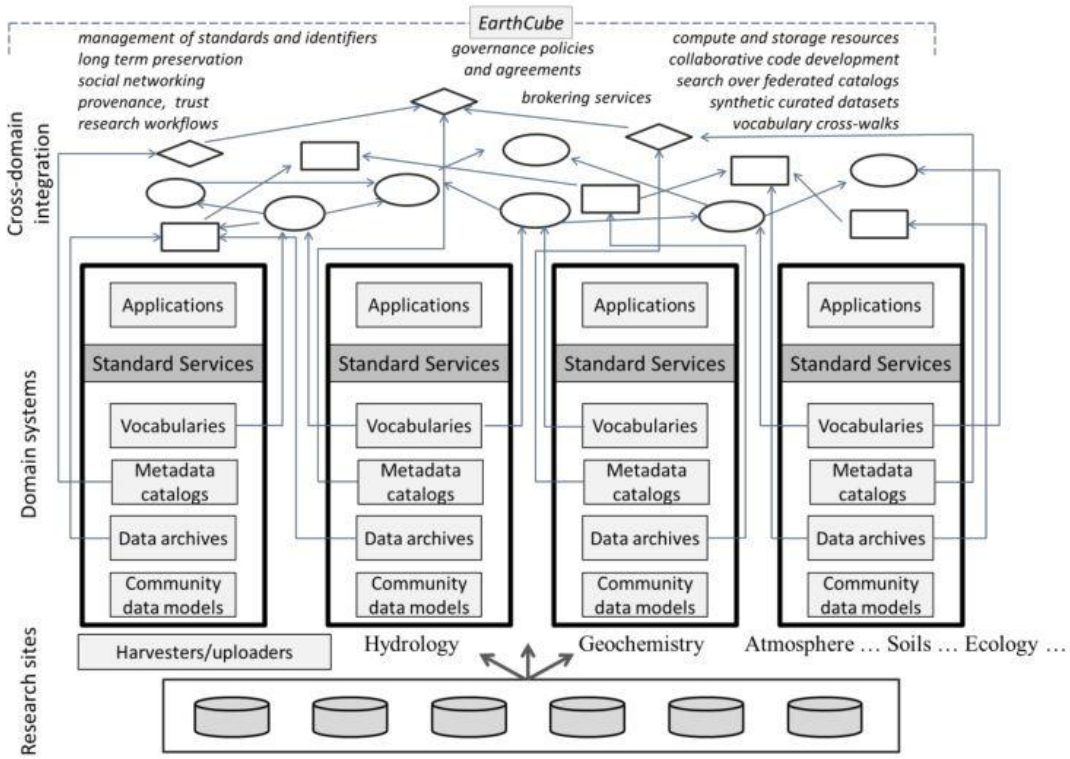

FIGURE 25. INITIAL MODEL OF AN EARTHCUBE LOGICAL ORGANIZATIONS[113]

The updated Cross-domain Interoperability Roadmap describes its vision for EarthCube:

> "A general vision of EarthCube logical organization…is of an integrated information system (or a "system of systems") that includes research observatories generating large volumes of observations and analytical/simulation results, domain systems that publish

[113] Zaslavsky, Ilya, et al., "EarthCube Roadmap, Version 1.1," 43.

the information according to community conventions about data models, vocabularies and protocols, and a cross-domain knowledge layer that includes federated catalogs, normalized and curated datasets integrating data from domain systems and observatories, cross-linked vocabularies, service brokers, as well as social networking, governance and compute infrastructure. This conceptual diagram is consistent with the consensus EarthCube diagram presented earlier[114]…and emphasizes the central role of the cross-domain interoperability layer enabling discovery, interpretation, data access and integration across domain infrastructures: the component named "EC Infrastructure[115]… has similar content and functionality to the upper layer in Figure [25 in this document], while "domain clouds"[116]…emphasize the same functions of cross-domain systems as the vertical "domain boxes" in Figure [25]."[117]

[114] See Figure 14 in this document.
[115] See Figure 14 in this document.
[116] See Figure 14 in this document.
[117] Zaslavsky, Ilya, et al., "Cross-Domain Interoperability Roadmap, Version 1.1," 42.

EARTH SYSTEM MODELS CONCEPT TEAM

GOALS

1. "The core of the ESM concept group proposal is to **provide definitive guidance on how to accelerate model development and coupling efforts**. As such, this component of EarthCube has the potential to put powerful new hypothesis –testing capabilities into the hands of geoscientists that should enable new discoveries and pathways to understanding our earth system."[118]

2. Create a **Geoscience Community Modeling Framework**
    - Accelerate progress related to modeling
    - Accommodate models of the entire Earth system and small parts of
    - Build upon and integrate several existing modeling frameworks and accommodate new frameworks
    - Need for cooperation among modeling framework projects to identify and adopt (or mediate between) common standards for interoperability

3. **Two-Solution Strategy**
    1. "Develop and integrate a set of standards-based protocols for linking component models of a myriad of Earth System processes into a coherent, conservative, multi-scale and high-performance computing enabled modeling system, and
    2. Develop an Earth System Model coupling testbed to facilitate standards-based approaches to ESM component coupling and to support broader engagement and participation of geoscience domain scientists in community-based ESM development and end-use applications." [119]

GOVERNANCE

GOVERNANCE NEEDS

- **Overarching principles**
    - Optimize engagement and participation across geoscience and end-user communities
    - Capitalize on existing and long-standing model development efforts
    - Use standards-based approaches to coding and semantic ontologies
    - Fully engage current and future computational capabilities
- **ESM modeling should be adaptable and evolutionary**
    - Solution can change over time in response to community needs
- Want **community based modeling**
    - Supports open model development processes
    - Community-defined modeling use cases and benchmarks

[118] Earth System Model Concept Award, "EarthCube Earth System Model Coupling Roadmap," draft roadmap released June 2012, 5.
[119] Ibid., 1.

EARTH SYSTEM MODEL-CENTRIC GOVERNANCE FRAMEWORK

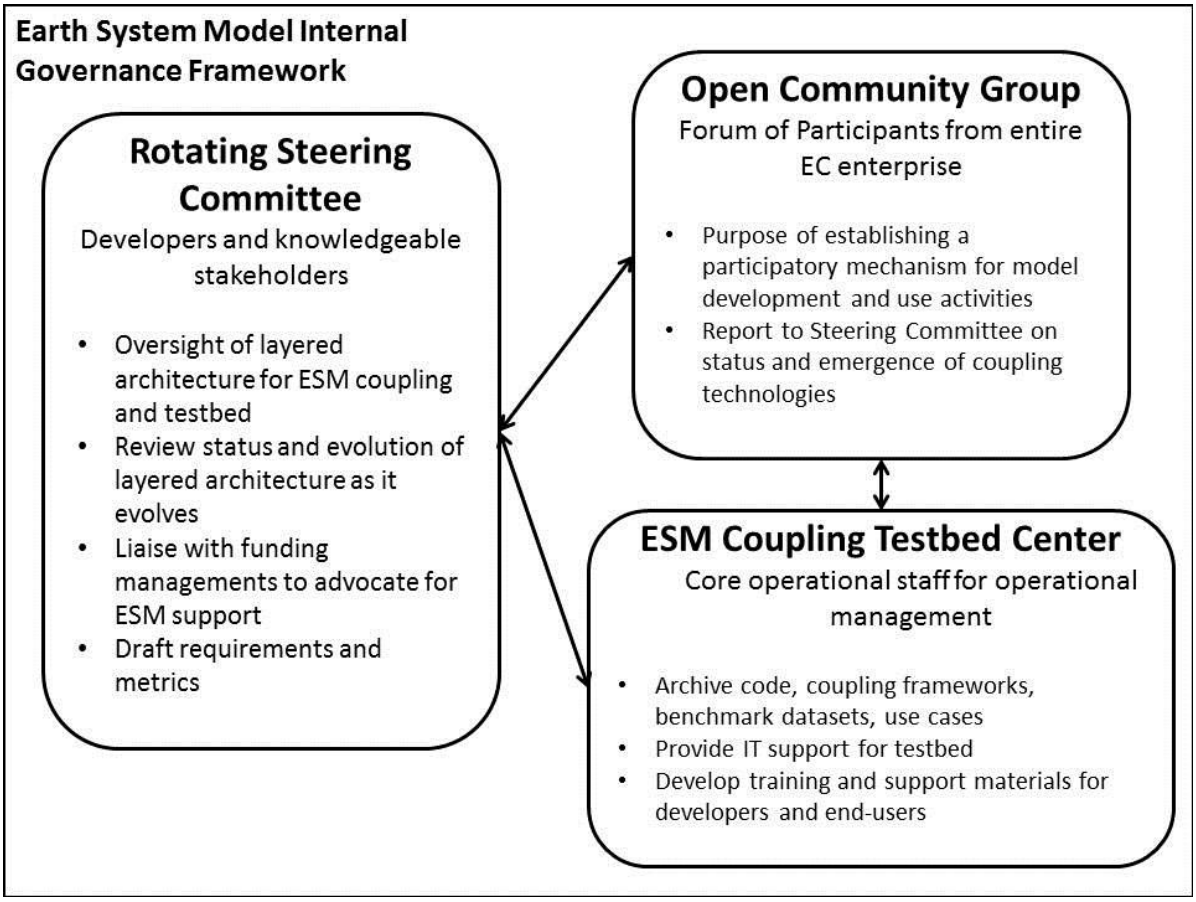

FIGURE 26. EARTH SYSTEM MODEL-CENTRIC GOVERNANCE FRAMEWORK[120]

[120] Graphic developed from written text in "EarthCube Earth System Model Coupling Roadmap."

LAYERED ARCHITECTURE CONCEPT TEAM

GOALS

Layered Architecture-specific goals for EarthCube were not mentioned.

GOVERNANCE

GENERAL GOVERNANCE NEEDS

- "For governance, a requirement should be the **establishment of an active, collaborative, software development community** for EarthCube. The first steps toward forming this community were taken at the Charrette by bringing together software engineers, informatics professionals, computer scientists, and domain scientists with an interest in building EarthCube infrastructure. We propose to continue this progress by **establishing a Community of Practice (CoP) as a more formal collaborative software engineering and informatics community that is informed by the geoscience community and that lays the foundation for a software governance organization**."[121]
- "Thus the end goal is the ability to specify the policies that control interactions between community resources and collaboration environments, verify that the policies have been enforced, and enable reproducible science."[122]
- "A community resource needs to verify that the policies governing use of the resource are enforced."[123]

TECHNICAL AREAS

- Distributed system of systems
- Assess community needs, requirements, and new opportunities for collaboration
- Explore architectural "layers" that can be integrated within higher level collaboration environments and functionality provided by each layer
- Define mechanisms for interactions among the layers
- Understand brokering, data and web service standards and other cross-domain interoperability mechanisms
- Explore mechanisms to make community resources available to researchers within all geoscience sub-domains

ENTITIES TO BE MANAGED WITHIN GEOSCIENCE INFRASTRUCTURE

- **"Users:** single sign-on environment
- **Objects:** files, databases, web sites, workflows, sensor streams
- **Collections:** logical arrangement of objects
- **State information:** metadata that tracks results of operations on name spaces
- **Storage resources:** collective operations across resources such as load leveling

[121] EarthCube Layered Architecture Concept Award, "EarthCube Layered Architecture Roadmap," draft roadmap released June 2012, 11.
[122] Ibid., 22.
[123] Ibid., 57.

- **Policies:** governance enforced through computer actionable rules
- **Procedures:** workflows composed from operations on name spaces
- **Physics model composition**: transformations on shared data to desired geometry, coordinate system, physical variables, etc.
- **Ontologies:** semantic mapping between domains, reserved vocabularies
  Grid applications: computing, workflows
- **Domain knowledge**: standard analyses cast as workflows that evaluate a research question"[124]

WEB SERVICES CONCEPT TEAM

GOALS

- Deploy web services at the various 9 partners in the Web Services group (geoscience domains of marine geophysics, atmospheric sciences, geodesy, hydrology, seismology, geochemistry, bathymetry, petrology, geodynamics, geological maps)
    - Extend simple web services to other data providers that the group engages.
    - Develop a turnkey system that can be easily deployed to a group without a significant IT staff

GOVERNANCE

Align with other EarthCube groups, especially Governance.

[124] EarthCube Layered Architecture Concept Award, "EarthCube Layered Architecture Roadmap," 7.

APPENDIX 4: EARTHCUBE GOVERNANCE FUNCTIONS MASTER LIST

The following is a master list of governance functions collected throughout the roadmap and governance framework writing process. Sources include background research on governance, the EarthCube Working Group and Concept Team draft roadmaps released prior to the June 2012 EarthCube charrette, community input during Governance Steering Committee virtual webinars, and the June charrette, a July 2012 EarthCube Principle Investigator workshop, and the EarthCube sessions at the ESIP Federation Mid-Year Meeting in Madison, WI, July 17-20.

This list forms the foundation from which the EarthCube functions and super-functions list was developed (see Section 3.1: EarthCube Governance Functions) which is a cornerstone of this initial EarthCube Governance Framework draft we are presenting to NSF and the EarthCube community. The master functions list represents everything we've collected since March 2013, while the Governance Function list presented in Section 3.1 is a synthesis and consolidation of these functions.

GOVERNANCE FUNCTIONS FROM THE EARTHCUBE GOVERNANCE ROADMAP

1. **Address barriers to participation:** EC governance should identify and mitigate barriers to participation.
2. **Collect and incorporate best practices and lessons learned**: EC governance should incorporated from case studies and historical infrastructure development into an EC governance framework.
3. **Identifying and address problems:** EC should incorporate processes for identifying critical, unsolved social, technical, organizational, cultural, and other problems that impede EarthCube from moving forward.
4. **Address path dependence:** EC governance should build-in mechanisms to avoid early technical, organization, social, etc., choices constraining available options later on.
5. **Address issues of data ownership and investment:** EC Governance should address tensions of data ownership and investment including intellectual property rights, data citation/credit.
6. **Address issues of data storage and meaning:** EC governance should address the technical, organizational, social and culture issues regarding data storage, preservation and curation.
7. **Promote data sharing.** EC governance should promote data sharing, including incentives for scientists to share their data, and tools that have the ability to accommodate and translate different needs from distinct domain communities.
8. **Promote collaboration and coordination:** EC governance framework should foster collaboration and coordination between Communities of Practice (CoPs), different organizations, government agencies, and individual scientists, among others.
9. **Foster open and transparent communication:** EC governance should foster communication across EC, including among EC stakeholders and those in leadership positions.

10. **Promote community-building:** EC governance should promote the building of an EC community, and foster community ownership of EC and EC governance.
11. **Promote education and outreach**: EC governance should foster education and outreach.
12. **Enable periodic evaluation:** periodic evaluation of the EarthCube governance framework is necessary to see if it is working to help achieve EarthCube goals
13. **Enable evolution:** EarthCube governance framework should enable its own evolution to meet changing community needs and emergent technologies.
14. **Define funding mechanisms and resource allocation:** EC governance should determine funding mechanisms and resource allocation across EC (at least initially – during the build phase).
15. **Create and maintain goals and objectives:** EC governance should create and maintain EarthCube Goals and objectives, and determine who is responsible for creating and maintaining EC goals and objectives.
16. **Promote integration:** EC governance should create mechanisms to deal with the technical, organizational, social and cultural challenges of integrating existing communities of practice, organizations, long-tail scientists, etc., into EC.
17. **Promote interoperability:** EC governance should promote interoperability of EarthCube CI components and EarthCube CoPs.
18. **Define leadership:** EC governance should define leadership roles and how those roles may be filled
19. **Address legal issues**: EC governance should address possible legal issues and how those issues will be handled.
20. **Create and update metrics:** EC governance should build in quantifiable metrics to measure success of EarthCube and effectiveness of governance framework.
21. **Ensure broad representation:** EC governance should include broad representation of EC community
22. **Identify and mitigate risks**: EC governance should build in a process to anticipate and manage possible risks.
23. **Foster sustainability:** EC governance should build-in mechanisms to increase the likelihood that EC is sustainable in terms of funding and community buy-in.
24. **Foster Trust:** EC governance should foster trust
25. **Create and maintain EarthCube vision**: EC governance should build-in mechanisms to create and maintain the EC vision.
26. **Adjudicate disputes.** EC governance should define and comply with a process for resolving disputes.
27. **Prioritize tasks**

GOVERNANCE FUNCTIONS FROM OTHER EARTHCUBE GROUPS

DATA DISCOVERY, ACCESS AND MINING WORKING GROUP

28. **Address data issues for big head and long-tail scientists**
29. **Processes:** (human- best practices, patterns, etc. and automated – systems, algorithms, etc.) to add assurance, provenance, reproducibility to data preparation phase
30. **Mechanisms to make tools sustainable** (tools developed rapidly to fill a specific need and then seen to generally useful by scientists and students who understand them)
31. **Clear expectations on what it means to publish data**
32. **Process for identifying technology 'readiness levels'**
33. **Process to gather and prioritize requirements on users' data needs**
34. **Creation of an EarthCube Data Coordination Committee**
35. **Transparency and openness** to cross cultural and institutional boundaries
36. **Risk identification and mitigation**
37. Better understanding of how **data mining is related to licensing, standards, interoperability, semantics, brokering, governance, and workflows**
38. Process to **evaluate technologies, produce documentation, and recommendations for systems requirements**
39. Process to **determine technological 'readiness levels' and evaluate if technologies meet these requirements and are ready for EC adoption**
40. Process to **collect, analyze, and categorize use cases**
41. Process to **gather input from data stewardship representatives and librarians to ensure data is managed properly**
42. Produce ongoing **documentation, recommendations and DDMA systems requirements**
43. Identify **gaps in coverage of needed capabilities and recommend how to fill them**
44. **Creation of a technology roadmap** -living document to handle change
45. Process to define and address **data-specific governance issues**
46. **Value scale to encourage contributions of higher quality data and focus on the value that scientists get from moving up the continuum by providing additional capabilities**
47. **Process to decide on what a low barrier to entry means, where the pitfalls are, and anticipate the pitfalls**
48. **Process to encourage people to participate/demonstrate benefits of participating in EC**

SEMANTICS AND ONTOLOGIES WORKING GROUP

49. **Processes to introduce and infuse semantics and ontologies across EC**
50. Creation of an **Ontology repository**
51. Promote use of **ontology tools and software services**
52. **Cataloging** of existing geoscience related ontologies
53. Semantics and ontologies **use cases**
54. **Mapping** across common vocabularies and ontologies
55. **Assessment** of current and future developments in semantic technologies

WORKFLOW WORKING GROUP

56. **Process to define workflow types** and capabilities from community requirements
57. **Identify relevant workflow standards** from existing standards bodies
58. **Process to classify workflow software:** Capabilities and scope, as they relate to project management and governance
59. **Process to identify gaps** in standards, workflow tool capabilities, underlying infrastructure
60. **Support use of workflow technologies** through training, documentation, outreach
61. **Match appropriate workflow experts** to support NSF and other research organizations
62. **Facilitate new collaborations**
63. **Scientific and technical bridges** to other areas (semantics, data management, etc.)
64. Process to **demonstrate workflow benefits to geoscientists**
65. An effective organizational structure to enable **efficient strategizing**
66. An effective **operational structure to ensure smooth and timely operational activities**
67. Effective processes for **creating groups, organizations,** etc.
68. Effective processes to **facilitate consensus and enable efficient decision-making**

BROKERING CONCEPT TEAM

69. **EarthCube-level protocols, interchange formats**, etc. to link communities of practice, existing organizations, etc.
70. **Interfaces** between brokering and other groups
71. Process for determining how EC CI brokers should be **deployed and maintained.**
72. Process for determining **requirements for broker deployment**- who sets them, what processes are used?
73. Process for determining **requirements for maintenance** - who sets them, what processes are used?
74. Process for determining **funding mechanisms**
75. Policies for **brokering across boundary between communities and EC cyberinfrastructure**
76. **Standards for publishing data, info, knowledge**
77. Process for **evolution and innovation of brokering for EC cyberinfrastructure**

CROSS-DOMAIN INTEROPERABILITY CONCEPT TEAM

78. Create a **Geoscience Interoperability Institute**
79. **Catalog federation**
80. **Service brokers**
81. **Information model profiles**
82. **Standards and identifiers**
83. **Collaboration**
84. **Archiving**
85. **Grids**
86. **Workflows**

87. **Long-term data preservation**
88. **Social networking**
89. **Provenance**
90. **trust**
91. **Compute and storage resources**
92. **Collaborative code development**
93. **Search over federated catalogs**
94. **Synthetic curated datasets**
95. **Vocabulary cross-walks**
96. Efficient way to **integrate data across domains**
97. Process to define **rules for interaction between domains** because managed by different domains
98. Process to **foster data and model reuse**
99. Process to **create documentation that can be understood outside of original context**
100. Process to develop **system to accumulate different uses of data and make it available to users**
101. Process to **give feedback to data providers**
102. Process to **determine scope of EarthCube**
103. Process to market EC and **demonstrate importance of EC to users**/why they should join
104. **Incentives to contribute data and use EC**
105. Process to address **questions of data ownership**
106. **Incentives to document data (metadata)**
107. Incentives to give proper **citation/credit in reuse of data**
108. Process to **identify and nurture EC communities of practice**
109. **Separation of technical and scientific decision-making**

EARTH SYSTEM MODELS CONCEPT TEAM

110. **Process for continued support and evaluation of model coupling activities**
111. **Process for developing a new layered architecture for ESM coupling**

LAYERED ARCHITECTURE CONCEPT TEAM

112. Promote **reproducible science**
113. **Community consensus** to determine **entities that are being shared and manipulated** (files, databases, streams, containers, semantics, ontologies)
114. **Community consensus on web service standards for each entity** (standard operations such as format transformation, coordinate transformation)
115. **Community consensus on brokers** (virtualization services for the standard operations)
116. **Community consensus on collaboration environment capabilities**
117. **Community consensus on governance** (policies and procedures)
118. **Interoperability mechanisms** to link existing infrastructure to EC

119. **Policies to control interactions between community resources and collaboration environments**
120. Process to **verify that the policies have been enforced**
121. **Process to commonly agree upon criteria for a cross-EC testbed activity**
122. **Decisions** about when to **add new broker capabilities** as part of EarthCube governance process
123. Process to facilitate the development of an active, collaborative, **software development community of practice**
124. Process to **assess community needs, requirements, and new opportunities for collaboration**
125. Explore **architectural "layers"** that can be **integrated within higher level collaboration environments and functionality provided by each layer**
126. Define **mechanisms for interactions among the layers**
127. **Determine how the following entities will be managed:**
    a. **Users:** single sign-on environment
    b. **Objects:** files, databases, web sites, workflows, sensor streams
    c. Collections: logical arrangement of objects
    d. **State information:** metadata that tracks results of operations on name spaces
    e. **Storage resources:** collective operations across resources such as load leveling
    f. **Policies:** governance enforced through computer actionable rules
    g. **Procedures:** workflows composed from operations on name spaces
    h. **Physics model composition**: transformations on shared data to desired geometry, coordinate system, physical variables, etc.
    i. **Ontologies:** semantic mapping between domains, reserved vocabularies
    j. **Grid applications:** computing, workflows
    k. **Domain knowledge**: standard analyses cast as workflows that evaluate a research question
128. 'Good Housekeeping Seal of Approval'/ audit system to evaluate compliance on data and services.
129. Process to determine who decides what compliance means and how this seal is implemented?

WEB SERVICES CONCEPT TEAM

130. **Work with Brokering and Interoperability Groups** within the Concept Awards to develop **common methods of discovery and access** to information assets.
131. **Community vetting and feedback** on web-services standards/specs
132. **full-function, sustainable user-support systems**, sufficient to foster large-scale, long-term deployments
133. Conduct **surveys of the Geosciences** to identify additional sources of data available in the community willing to expose data assets through Web Services

134. **Address knowledge management:** EarthCube governance should address how knowledge is managed within EarthCube.
135. **Collect and incorporate use cases, and address use case implementation:**
    a. Focus on pilots and use cases that need to be prioritized for CI development.
    b. Governance use cases should leverage and engage consistent with the three level model adopted at this (PI) meeting
    c. When possible, use cases for governance should coordinate with (and use?) system and component level uses cases mentioned in the PI meeting during use cases discussions
    d. Interactions with outside groups need to be major component of use cases.
136. **Connect Resources:** EC governance should connect resources across EC.
137. **Leverage existing investment:** EC governance should leverage CI investment over the past two decades
138. Process for the community to **determine research and funding priorities**
139. **Deal with cross-agency and interagency issues**
    a. **Create acceptable use policies** for NSF directorates, federal agencies, and international resources
    b. Identify how EarthCube will work across different agencies and what kinds of governance need to be in place for that to happen.
    c. Address the avenues for **international collaboration**
140. **Manage how domain components interact with each other**
141. **Manage EarthCube-wide functions, such as issues of lifecycle, catalogs, information models, data centers, etc.**
142. **Define metrics for CI adoption and readiness**
143. **Evaluate extensibility**

APPENDIX 5: COMMUNITY INPUT AND ENGAGEMENT

WORKSHOPS & MEETINGS HOSTED

DEVELOPING THE GOVERNANCE ROADMAP

March 23, 2012 through June 14, 2012

All meetings were open to the public with agendas and Webinar call-in information posted to Ning:

- Governance Steering Committee Call, March 23, 2012
- Governance Steering Committee Call, March 30, 2012
- Governance Steering Committee Meeting, Denver, CO, April 4-5, 2012, 12 participants
- Governance Steering Committee Call, April 6, 2012
- Social Media Conference Call, April 9, 2012, 4 participants
- Virtual Governance Plenary Session, April 11, 2012
- Governance Steering Committee Call, April 13, 2012
- Virtual Governance Plenary Session, April 17, 2012
- Social Media Conference Call, April 19, 2012
- Workflow Working Group Weekly Call, Governance Briefing, April 19, 2012
- Governance Steering Committee Call, April 20, 2012
- Governance Steering Committee Call, April 27, 2012
- Governance Steering Committee Call, May 4, 2012
- Governance Steering Committee Call, May 11, 2012
- Governance Steering Committee Call, May 18, 2012
- Governance Steering Committee Call, "Roadmap Review," May 22, 2012,
- Governance Steering Committee Call, May 25, 2012
- Focus Group, "Engaging International Partners on Governance (Pacific Rim) Webinar," May 30, 2012, 5 participants
- Focus Group, "Engaging International Partners on Governance (Europe) Webinar," May 31, 2012, 6 participants
- Focus Group, "Engaging Industry Partners on Governance Webinar," May 31, 2012, 4 participants
- Focus Group, "Engaging Federal Agency Partners on Governance Webinar," May 31, 2012, 8 participants
- Governance Steering Committee Call, June 1, 2012
- Governance Steering Committee Call, June 8, 2012
- Governance Steering Committee Meeting, Roadmap review with EarthCube Working Group and Concept Team Liaisons, June 11, 2012
- EarthCube Charrette, June 12-14, 2012

DEVELOPING THE GOVERNANCE FRAMEWORK

June 22, 2012 – August 31, 2012

- Governance Steering Committee Call, June 22, 2012
- Governance Steering Committee Call, June 29, 2012
- Governance Steering Committee Call, July 6, 2012
- Governance Presentation at the EarthCube Concept Award Principle Investigator Meeting in Boulder, CO, July 10, 2012
- Governance Steering Committee Call, July 13, 2012
- EarthCube Governance Workshop at the ESIP Federation Mid-Year Meeting, Madison, WI, July 17, 2012
- EarthCube Governance Presentation at the ESIP Federation Mid-Year Meeting, Madison, WI, July 19, 2012
- Governance Steering Committee Call, July 27, 2012
- Governance Steering Committee Call, August 3, 2012
- Governance Functions Editing Session, August 7, 2012
- EarthCube Governance Workshop at Geosciences Australia Headquarters, Canberra, Australia, July 30, 2012
- EarthCube Governance Presentations at the 34th International Geological Congress, Brisbane, Australia, August 5-10, 2012
- Governance Steering Committee Call, August 10, 2012
- Governance Steering Committee Call, August 17, 2012
- Governance Steering Committee Call, August 24, 2012
- Governance Framework Editing Session, August 27, 2012
- Governance Steering Committee Call, August 31, 2012

COMMUNITY INPUT TO-DATE: WORKSHOPS AND CONFERENCES

Community building and engagement of EarthCube stakeholders is a key component of the EarthCube Governance Working Group's efforts prior to and following the June 2012 NSF EarthCube charrette. As described in the framework narrative at the beginning of this document, an aggressive community engagement plan is being developed and implemented with the goals of educating EarthCube stakeholders about EarthCube and EarthCube governance, and gathering input on their governance wants and needs.

EARTHCUBE CONCEPT AWARD PRINCIPAL INVESTIGATOR MEETING BOULDER, CO, JULY 10, 2012

The EarthCube Brokering Concept Award organized a meeting of EarthCube Principal Investigators (PIs), mostly representing the Concept Awards, although a representative from a few of the Working Groups was also present. The goals of this meeting were to build consensus on community views and the direction of EarthCube development, focusing particularly on pending issues that were not resolved at the June 2012 EarthCube charrette, including merging and synthesizing the draft roadmaps to synthesize over-arching goals and create a timeline for EarthCube development in the near-term.

Outcomes of the PI meeting focused on governance, EarthCube reference architecture, use cases, and timelines. For a detailed report on this meeting, please see: https://docs.google.com/folder/d/0B6ovZrDPKFGuc0xtMHp2YTA3R3M/edit?docId=0B6ovZrDPKFGuOFlFNTBDSEl3d0E

EARTHCUBE GOVERNANCE RECOMMENDATIONS

1. "There is a need for near term governance guidance so that the EC community can be aligned behind a set of priorities established by NSF.

2. Participants agree that an operational governance framework should be in place by the end of the year.

3. In formulating Governance, the following should be considered:
    a. Prioritize research initiative use cases
    b. Define metrics for CI adoption and readiness
    c. Incorporate technical and scientific fora with appropriate representation from the EC community
    d. Address use case implementation
    e. Evaluate extensibility
    f. Consolidate needs for sustainability as part of working with NSF

4. Support NSF in cross-agency and interagency issues (in these, it was not clear the role of the EC program office, NSF program offices and the EC governance function).

a. The anticipated activities, by the NSF EC office or others may be to:
    i. Create acceptable use policies for NSF directorates, federal agencies, and international resources
    ii. Identify how EC will work across different agencies and what kind of governance needs to be in place for that to happen.
b. Address the avenues for international collaboration
c. Support community deliberations for identifying research priorities

5. Use cases should have priority in assessing program directions and governance
    a. Focus on pilots and use cases that need to be prioritized for CI development.
    b. Governance use cases should leverage, engage and be consistent with the three level model adopted at this (PI) meeting
    c. When possible, use cases for governance should coordinate with (and use?) system and component level uses cases mentioned in the PI meeting during use cases discussions
    d. Interactions with outside groups need to be major component of use cases.

6. Governance Structure Considerations
    a. A recommendation from the X-domain interoperability CA was to create a science committee and a technical committee (as noted in the x-domain interoperability roadmap). These committee structures imply a specific governance model. Further examination is needed to assess whether such a governance model can be broadly applied in the EC paradigm.
    b. If a two-committee structure is adopted, it is necessary to have a process/structure to integrate the outputs and recommendations of these committees.
    c. The breadth of representations on the committees or within the governance structure in its final instantiation should be large and bring in the GEO communities and scientists

7. Scope of EarthCube – issues that should be addressed:
    a. It is important to determine the scope of EC, because the scope so far has been defined differently by different groups.
    b. Will EC governance encompass the entire information management landscape of geosciences or will it only be a more superficial layer that manages how domains interact with each other?
    c. If there is an "interoperability middle layer of the CI" and, if so, how will it be governed and maintained?

8. Scope Recommendation – participants at the meeting agree that
    a. Each science domain maintains its own internal governance framework, catalogs, services, information models, data centers, etc. It is not the intent of EC to displace these.
    b. EC governance would recommend and oversee how the domain architectures and information systems are connected with each other, and would manage EarthCube-wide issues of lifecycle, catalogs, vocabularies, including the management of an interoperability CI middle layer, etc. Further discussions on details are necessary to address the scope.

c. To assess the scope, we need to look at the scenarios among groups within EC and how these groups interact with groups outside of EC."[125]

EARTHCUBE REFERENCE ARCHITECTURE RECOMMENDATIONS

Another outcome of the PI meeting was a recommendation for a possible model for an EarthCube reference architecture. Although it combines elements from all the Concept Award draft roadmaps, it is based on the original Brokering Concept Award reference architecture, first introduced in the Brokering Draft Roadmap presented to NSF prior to the June 2012 EarthCube charrette.

The PI meeting group came to consensus on a high-level architecture for EarthCube via a 3-phase process: 1) Outline of the benefits of a conceptual-level architecture framework; 2) Review of current recommendations; and 3) Formulation of a common model. This model is presented in Figure27 below.

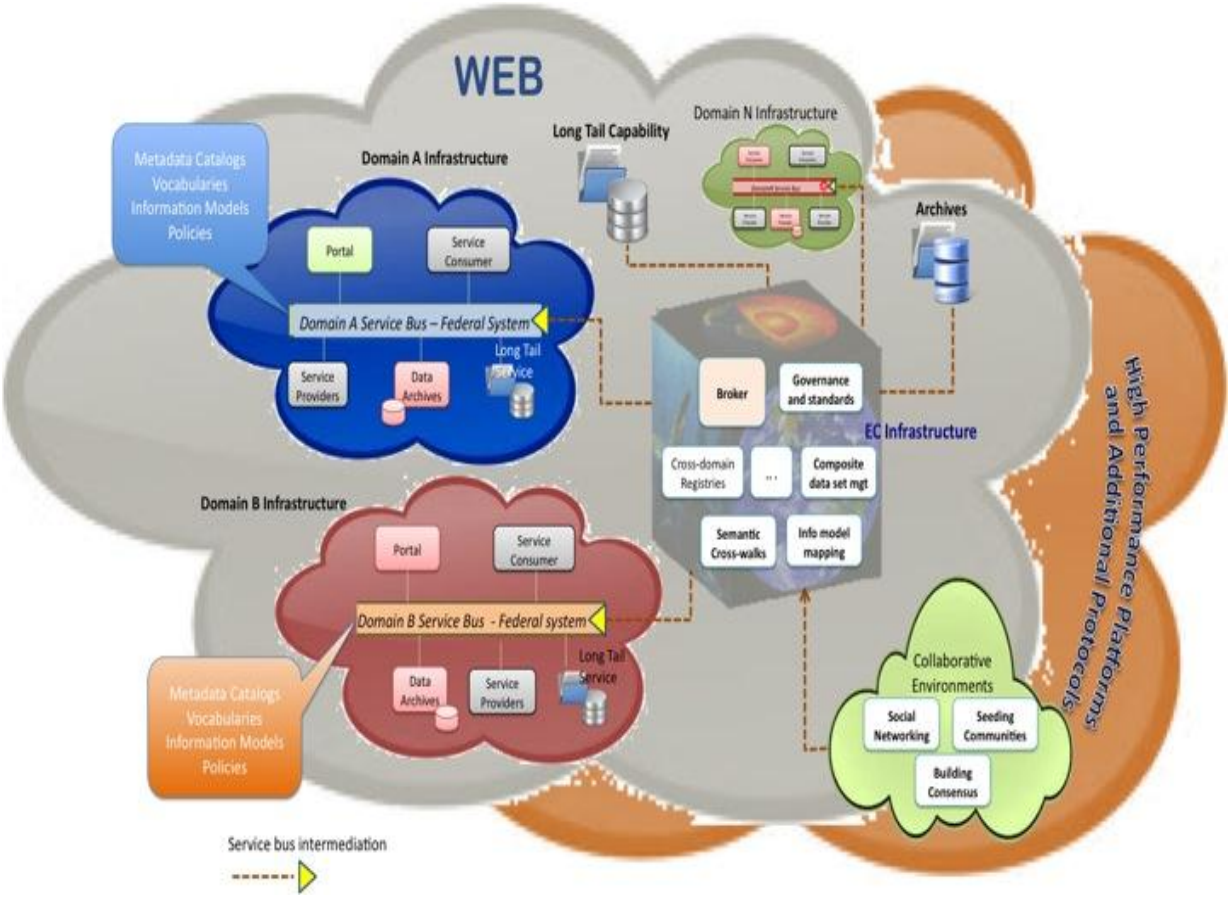

FIGURE 27: EARTHCUBE ARCHITECTURE RECOMMENDATION RECOMMENDED AT THE EARTHCUBE CONCEPT AWARD PRINCIPAL INVESTIGATOR MEETING IN BOULDER, CO, JULY 10, 2012.[126]

[125] Siria Jodha Khalsa et al.,"., "Summary and Recommendations from EarthCube CA PI Meeting, 5-6.
[126] Ibid., 4.

The PI Meeting Report describes the architecture model for EarthCube in this way:

> "Discussions of the EC architecture frameworks from the various EC roadmaps led to a common vision for a high level EC architecture construct given in the figure below. The reference architecture can be characterized as an environment enabling effective collaboration across multiple science domains and domain data systems to address challenges in the geosciences through efficient discovery and re-use of data, information and knowledge. Interactions within existing domain infrastructures use established catalogs, semantics, data access mechanisms, information models and policies. These components, when exposed via community standards, facilitate interactions between different domain infrastructures. In particular, brokering services and related cross-domain components (semantic cross-walks, cross-domain registries, information model mappings) are employed to mediate between community resources and users/researcher clients that follow different domain specific models. In this manner, the EC infrastructure facilitates cross-domain science through creation of mechanisms for discovery, access, processing and semantic mediation; workflows services are provided as well for distributed process chaining…
>
> …The figure shows various disciplines and domains with their own community infrastructure. Both the "long-tail" and smaller archives are addressed. The system includes both web-based and high performance platform and other protocol support. The fundamental attributes of the EC Infrastructure element allow for services that link multiple domains without imposing additional burdens on the participants, be they archives or users or others. The EC Infrastructure addresses not only current services such as discovery and access, but desired capabilities such as quality and provenance and should be designed to evolve as EC progresses."[127]

EARTHCUBE USE CASE RECOMMENDATIONS

The PI Meeting Participants also developed a series of recommendations on collecting and analyzing use cases. One of the functions of EarthCube will likely be to collect and analyze use cases, and governance will help guide how this process occurs. Use case-specific recommendations include:

1. "Establish a use case management system including a use case repository. The system should include:
    a. An inventory of use cases, including use cases developed elsewhere (ESIP, OGC, agencies)
    b. An inventory of geosciences models
    c. Standardized description and community tagging
    d. Standardized use case analysis methodology and results reporting
    e. The establishment of a use case model and templates describing how to populate the model

[127] Siri Jodha Kahlsa et al., "Summary and Recommendations from EarthCube CA PI Meeting," 5-6.

2. Use 3-level approach for use case management [see Figure 28 below]
3. Recommended set of use cases is formulated based on road maps noted at the beginning of this section
4. Extend use cases to include social, organizational, cultural aspects
5. Define metrics and governance for use case development, selection, and analysis across all EC groups –test with initial set of use cases listed above.
6. Provide community system for eliciting additional use cases consistent with the 3-level
7. approach
8. Start use case work with large science projects to actively participate in use case development & evaluation
9. Create examples use case implementation tutorial."[128]

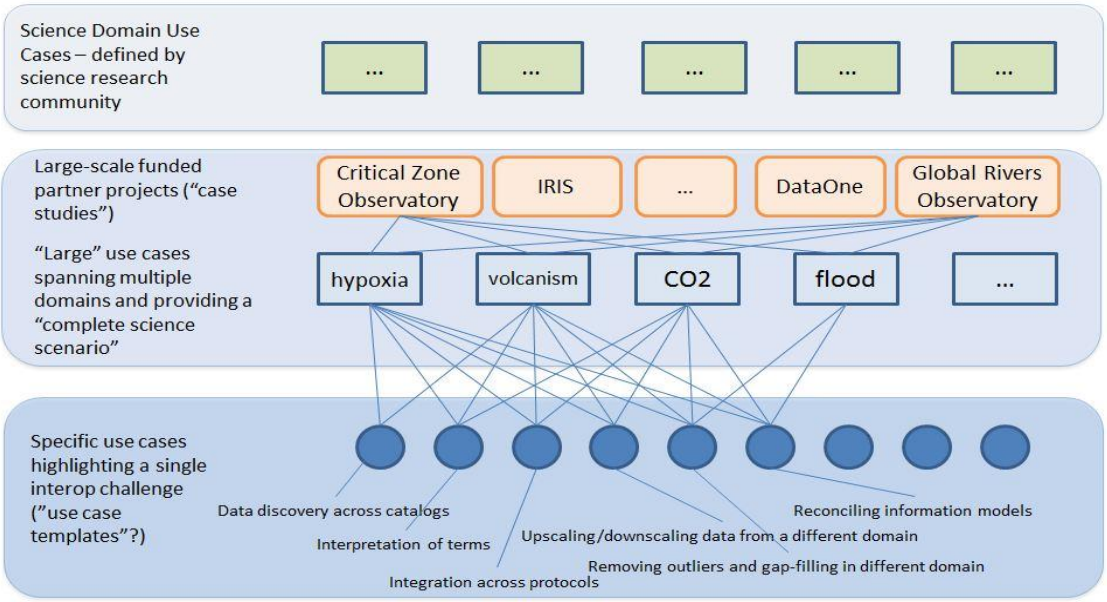

FIGURE 28. THREE LEVEL USE CASE PARADIGM[129]

EARTHCUBE TIMELINE RECOMMENDATIONS

The EarthCube Concept Award PIs also proposed a timeline that synthesizes the milestones and identified in the Concept Award and Community (Working) Group draft roadmaps (Table 3).

[128] "Summary and Recommendations from EarthCube CA PI Meeting," 11.
[129] "Summary and Recommendations from EarthCube CA PI Meeting," 8.

Concept Awards	Community Groups	Resequence Analysis - New resources
Catalog existing prototypes and pilots	Governance – Science grand challenges, prioritization of research	Science Use cases (6-12 months); Cyberinfrastucture integration (2 months)
Catalog and characterize use cases/governance	Use case process	Requirements (incl. reference architecture); 4 months for 1st cycle
Define metadata req'ts	Reference arch blueprint	Inventories (12 months)
Catalog existing/proposed resources for interface	Evaluation metrics	Governance: Scientific framework (6 months); Technical (12 months)
Reference architecture	Technical task forces – community input & eval	User interface/outreach/ engagement (12 months)
Define and implement testbeds	Technical Prototyping/dev	Standards / registries (24 months)
Outreach/engagement	Education group assumed	Pilots/testbeds (12 months)
Define and implement production testbeds		Metrics (12 months)

TABLE 3. ACTIVITY SEQUENCE SUMMARY AS IDENTIFIED BY CA AND CG TEAMS INTEGRATED INTO A SINGLE TABLE.[130]

This timeline makes several assumptions regarding governance:[131]

1. The governance framework will be operational by the end of 2012
2. Governance will prioritize research initiatives and use cases
3. Governance will define metrics for evaluation of cyberinfrastructure adoption, use, efficacy and readiness
4. Governance will consist of a technical committee and a scientific committee, each with appropriate representation
5. Governance will evaluate use case implementations
6. Governance will evaluate potential for extensibility

ESIP FEDERATION MID-YEAR MEETING, MADISON, WI, JULY 17-20, 2012

The Earth Science Information Partners (ESIP) Federation Summer Meeting in Madison, WI, July 17-20, provided the first venue to introduce EarthCube governance to domain scientists not already involved in EarthCube. The Governance Group held an all-day, working meeting to begin development on an initial governance framework for EarthCube, in addition to a Governance presentation for ESIP community, introducing EarthCube and basic concepts of governance. In addition to the EarthCube Governance sessions, there were several other EarthCube sessions at ESIP: Brokering, Cross-Domain Interoperability, Data Discovery, Access and Mining, and Workflows.

[130] "Summary and Recommendations from EarthCube CA PI Meeting," 13.
[131] Ibid., 14.

Both the July 17th Governance Workshop and July 19th Governance Presentation are summarized here, in addition to meeting minutes from these and the other EarthCube sessions.

EARTHCUBE GOVERNANCE WORKSHOP, JULY 17, 2012

The July 17th Governance Workshop served as an open, working meeting for the Governance Steering Committee, with participation and input from interested ESIP members.

Workshop Goals:

1. Determine a path forward in order to meet steps 1 & 2 of the EarthCube Governance Roadmap (determining an initial EarthCube governance framework and determining and engaging the EarthCube stakeholder community)
    a. Develop a list of concrete tasks and a timeline to move forward on choosing a governance framework
    b. Identify gaps in information and in steps needed to move forward
2. Determine appropriate use-case templates and governance scenarios to evaluate potential governance frameworks for EarthCube.
3. Move forward on a community engagement and outreach plan.
    a. Determine how community engagement plan will move forward
    b. Determine what materials are needed to market EarthCube to new users
    c. Initial discussion of upcoming conferences/venues to present EarthCube Governance to new users
4. Review and consolidate initial list of EarthCube governance functions
    a. Use these to help evaluate potential governance frameworks
5. Address remaining issues:
    a. How different EarthCube groups and CI components interact with each other.
    b. Scope of EarthCube
6. Determine how potential governance models will be evaluated
    a. What metrics will be used? etc.
7. Plan for Thursday, July 19th EarthCube Governance Workshop
    a. Determine workshop goals, presenters, and materials needed

Although the group came into the meeting with a set agenda, the course of the meeting altered dramatically from the original plan. The group anticipated that by mapping governance models to governance use cases, the optimal governance framework for EarthCube would become clear. Upon intending to do this, however, it became obvious that this method would not provide the answers the group was looking for. Additionally, the group realized that a recommendation of a specific governance model for EarthCube was too prescriptive at this point in EarthCube's development, and that it would be better to develop a framework to give guidance for developing EarthCube governance, while leaving room for many different options to actually implement EarthCube governance in the future.

Thus, the group decided to develop an initial list of functions that should be carried out by EarthCube governance, in addition to a list of guiding principles, and a series of general governance

recommendations. These lists ultimately became the foundations of the initial governance framework recommended earlier in this document.

MEETING MINUTES

Here are the meeting minutes from the EarthCube Governance Workshop, which took place on July 17, 2012. Notes were taken by Genevieve Pearthree (Arizona Geological Survey). Meeting materials for the workshop can be found here:
https://docs.google.com/#folders/0BxE8IKrxOMWncFNDWlBoTU12ZFE

In-Person Participants

Not a complete list – more people joined the meeting throughout the day and not all were included on this list:

1. Lee Allison, AZGS
2. Chris MacDermaid, Colorado State University
3. George Percivall, OGC
4. Kim Patten, AZGS
5. Genevieve Pearthree, AZGS
6. Jennifer Schopf, NSF
7. Brian Wee, NEON
8. Siri Jodha Khalsa, Brokering PI, NSIDC
9. Steve Browdy, EC Brokering Team
10. Caron Moe, NASA Earth Science Technology Office
11. Mike Engel, Kodak
12. Charles Thompson
13. Dave Meyer, USGS
14. Steve Allenbach
15. Mike Daniels
16. Kim Patten

Virtual Participants

17. Jim Bowring, CERDLES
18. Mohan Ramamurthy
19. Steve Kempler
20. Cecelia DeLuca
21. Geoffrey Fox
22. Jay Pearlman
23. Carroll Hood

July 10th PI Meeting in Boulder General Governance Recommendations

- Governance is important
- Group would like governance framework in place by end of year and agree with August 15th deadline
- Group would like governance group to provide input on priorities for governance
- Roadmap alignment exercise

Need to define scope of governance and what will be governed.

- Cross-Domain Interoperability group has a good model of cross-domain resources that should be managed
  - Cross-Domain group also has a come forward with a potential governance model
- Scope of EarthCube is a major decision, and what is the scope of governance?
  - Cross-Domain approach:
    - EC is interoperability and coordinating layer
    - Each domain manages itself
- Top-down vs. bottom up approach

- Not necessarily consensus from July 10th PI meeting for governance of domain systems, but guidance on best practices to lead to convergence on standards to achieve better integration
    - Initially components will develop independently, but later they converge and are made interoperable
- Will governance put in place a process to define EC scope?
- 2 scopes
    - x-domain resources (top layer)
    - To what extent will EC governance penetrate domains (second layer)?
        - How much does EC put requirements on domain systems?
        - there will be domain systems that NSF has no control over
        - if you can get best practices out there, many organizations will want to be interoperable
        - but, there are different layers of interoperability
- Have we considered evolution of EarthCube over time?
- Governance will change as EC gains critical mass
    - EC won't be able to change existing orgs, like NOAA or NASA
- Inception phase - interoperability layer
- CI community - second phase
    - Create own capability and identity

Developing an Optimal EarthCube Governance Framework Document

- Many different definitions of EC
- Possible to enable interoperability framework based on common functions and touchpoints (at least initially)
- Eventually, EC-specific research initiatives will emerge (once EC is robust)
- This will Determine what governance pieces need to be in place
- A high-level reference architecture needs to be developed to determine what components need to be governed
    - Determine 7-10 high-level governance functions

Use Cases

- Use cases to evaluate and limit governance models to recommend for EC
- Need to consider community values when evaluating governance models
    - Possible that governance models considered so far are underspecified
    - Need to come up with minimal specifications for how groups can participate in EC
- Current governance group is separate from governance itself
- Workflow Software Release (sample use case)
    - Expected outcomes: define requirements
    - Need to connect use case template with 9 governance functions (from Developing an Optimal EarthCube Governance Framework document)
- Need list of low-level governance functions to evaluate different governance models

    - Terms of reference for committees and processes, etc.
    - Typical data flows and routines that EC is supposed to support
- Use workflow software use case to determine governance functions
- Boehm study as an example of governance archetype evaluation
- Maybe could evaluate specific frameworks against metrics and criteria

EarthCube Governance Functions

- We need to define functions, maybe not specific framework
    - But this may give NSF acquisition arm the authority to determine framework
    - Nut also may give NSF the tools for individual groups to develop own governance models
        - Governance structure will emerge from operational governance system, based on how each group interacts with each other
        - Risk in uncertainty of development process, but in line with typical agile development processes
        - follows infrastructure development process where it starts small and grows
            - initial functions identified and go from there
            - allow for governance aspect to have parallel efforts in certain domains that eventually merge to a more comprehensive governance structure

what you're building and how you build it are two different things

what about non-functional requirements - maybe are more important than actual functionality

Maybe need a list of guiding principles and functional requirements, EC values

Excel spreadsheet to evaluate archetypes compared to functions

Next steps for today's workshop

- Consolidate EC functions into a shorter list of 'super functions'
- then look at what processes go along with each function
    - decision-making, alignment and communication processes
- then evaluate against governance archetypes

Re-evaluation of Framework to present to NSF

- Instead of presenting pros and cons of specific governance models to NSF, better to present list of functions and associated processes
- NSF can then solicit proposals to fill these functions
- But, the framework won't define what type of governing body should carry out these functions
    - Expectations from EC Other Groups (from roadmaps)
        - There will be an umbrella governance group/functions
        - There will be working groups

- Will there need to be an umbrella group (s) / governing body, maybe for the next 3-5 years?
    - like DARPA (at the beginning of the internet)
        - This follows the typical infrastructure development process
- Is there a different group to fill each EC governance function?
    - could be centralized or decentralized
        - single central body with a series of working groups within
        - or a series of distributed bodies with some sort of coordinating committee to coordinate their interactions
- An over-arching umbrella group with separate groups to fill each function?
- EC framework deliverable to NSF
    - what we anticipate is going to happen, in terms of governing bodies, such as an umbrella committee/organization/management body to bring all the pieces together
    - expectation that within 5 years, EC will turn into a classic infrastructure model
    - deliverable doesn't have to name a specific archetype for EC
        - up to proposers to create organizational structures to achieve EC goals and EC functions
- NSF solicitations to design and set up an organization to manage NEON
    - did not specify what type of organization or type of governance mechanisms that organization would have
    - make use of partnerships with other orgs

Potential Initial Recommendations for NSF: Document as the front of the original Governance Roadmap

1. Umbrella organization/body
2. Specific approach to carrying out specific processes may take many different forms (up to proposers), but they must be compatible with EC goals and EC community
3. articulated set of values/guiding principles to go along with framework to inform how framework would be realized
    a. for NSF to state explicitly what they've been saying all along
        i. Such as cross-communication, open-source software, community-based etc.

Governance constraints on other projects (?) - maybe can come later

. might need to write explicitly some assumed constraints

a. could be seed for evolution of other requirements

b. or say that community groups may bring forward constraints that the umbrella group may propagate

c. Related to NSF's data management plan? Maybe being EC-compliant fits in with management plan

Guiding principles

1. Science-driven objectives and development
2. Open and transparent processes

3. Globally-distributed and diverse and developer base
4. Sustainability, reduce environmental footprint as much as possible
5. Scalability
6. Searching for and apply best ideas regardless of source
7. Collaboration among the IT and science people
8. Community engagement at every opportunity
9. community-based governance for direction and priority setting
10. Free and open sharing of data and software (after how many years? - determined by governance)
11. platform-independent tools and interoperable frameworks
12. use of open and community standards
13. adopting, adapting, and only as a last resort, duplicate existing or develop new capabilities

Near Term Next Steps

1. We believe we can synthesize governance functions and develop framework by August 15th.
2. But, can we properly vet our recommendations with the community by August 15th?
    a. Will leverage existing venues
        i. ESIP Thursday July 19th workshop
        ii. IGC August
        iii. NEON workshop in October
        iv. ASIS&T : Oct 26-30
        v. GSA early November
        vi. AGU December
        vii. OGC meeting at ESRI - January
        viii. iConference 2013: Feb 12-15
        b. in addition to community engagement plan

Maybe a good idea to deliver as a draft to NSF by August 15th, and continue to vet it with community - beta version

Goal to keep EC and NSF moving forward - 6 month plan

Work with EC groups as a reference source to help them consider their governance needs

1. one-on-one webinars and briefings with other EC groups

a. A lot of groups are still working on their roadmaps, so governance guidance and advice might be useful for them

EARTHCUBE GOVERNANCE PRESENTATION, JULY 19, 2012

Prior to the EarthCube Governance session (which was the second-to-last of the EarthCube sessions at ESIP), several important points became clear. First, each of the EarthCube presentation gave a slightly different definition of EarthCube, depending on that group's particular area of focus. Second, the audience members had many pre-conceived ideas of EarthCube, in addition to several shared concerns.

Thus the EarthCube Governance Presentation focused on what EarthCube might be, with the objective of giving a bigger picture of EarthCube, and presenting governance concepts in general. The presentation is accessible online at: http://commons.esipfed.org/node/474.

AUDIENCE QUESTIONS AND COMMENTS

For more information, please see the ESIP Commons EarthCube Governance Session at: http://commons.esipfed.org/node/474. Kelly Monteleone took notes.

Access

1. Access may not be enterprise level function… computer and visualizing are specific
2. EarthCube may consist of different registries that are connected
    a. Each domain may have different data model, especially if visualizing in 2 or 3D
3. There needs to be agreement on what is an enterprise level function
4. Some functions will exist at different levels
5. There will be infrastructure for community – ex. environments

Evolution of Governance

1. The concept of expecting evolution of governance and being able to let that happen is good
2. But, it is also a problem to put boxes around things that may not need boxes
3. There is a need to balance boxes and innovation
4. Touch points was not meant as an organization diagram – functions not organizations – problem is there are so many preconceptions as to what is coming forward – functional description
5. It will be hard for governance to morph with changing system

Interoperability

1. NSF has a wide variety of communities that use community level standards
2. There is a need some for interoperability
3. Governance should not impose a profile for community, just the minimum or optimum amount
4. Governance could make recommendations, but expect at least a minimum amount

Feedback

1. What is the feedback from other communities on EarthCube governance so far?
2. Community engagement has been largely within EarthCube community – Around 250 in-person and virtual participants at the NSF EarthCube charrette
3. The Governance Group has given some webinars, but a limited amount
    a. There were a few webinars for funding agencies and international community (which were recorded and posted to the Ning site)
    b. There were good representatives from each community but not good representation.
4. The Governance Group is working to design this framework

OTHER EARTHCUBE ESIP WORKSHOPS

These questions and comments arose during all the EarthCube Data Discovery, Access and Mining, Workflows, Cross-Domain Interoperability, and Brokering sessions at the ESIP Federation Mid-Year Meeting, which took place July 17-20, 2012. Community input was recorded by Governance Group Research Assistant Genevieve Pearthree (Arizona Geological Survey), and ESIP Fellows Kelly Monteleone (University of New Mexico), Angela Murillo (University of North Carolina at Chapel Hill) and Sarah Ramdeen (University of North Carolina at Chapel Hill).

WORKFLOW

For more information, please see the ESIP Commons EarthCube Workflow Session page at: http://commons.esipfed.org/node/485. Angela Murillo took notes.

Engaging the Community

1. Are there specific ideas on how to engage people invited to workshops?
2. Participants do not want to be inundated with questions
3. If I'm a scientist, do I have a brokering problem? I know I have data problems
    a. Everything is a data problem in some way

eScience

1. Europeans have been very involved with eScience, have they done anything with workflows?
2. No one has done anything with workflows
3. See more eScience for social science and humanities
4. There has not been this kind of consorted effort to create workflows for science

DATA DISCOVERY, ACCESS AND MINING

For more information, please see the ESIP Commons EarthCube Data Discovery, Access and Mining Session page at: http://commons.esipfed.org/node/461 and notes at: https://docs.google.com/document/d/1mksljGbFVnL5fFYE8bqFBwlQV5vgZ1uUN_dFx6ACj2Q/edit Kelly Monteleone, Genevieve Pearthree and Sarah Ramdeen took notes.

Governance

1. There is a need for governance policies regarding data sharing
    a. Policies to choose who you're going to share data with, and for how long
    b. Data sharing changes, depending on the audience.
        i. Some audiences need more technical components, while others need more visualization tools.
    c. A potential answer is data cooperatives – you don't share data with everyone at once
2. What does it mean to participate in EarthCube and contribute data to EarthCube?
    a. Who manages or decides if EarthCube is going to archive data?
        i. How do you come up with the framework to help the community make these decisions?
    b. It needs to be decided what low barrier to entry means, where pitfalls are, so that a governance framework can avoid pitfalls
        i. The Data Working group can decide what it need in terms of data-specific governance, and the Governance group would provide a larger strategic governance to bound the scope of EarthCube.
    c. Governance should bring communities together, but not force policies down from the top
        i. Or perhaps be a blend of top-down and bottom-up governance
        ii. Overall EarthCube governance can determine strategic priorities, there can also be internal governance for each element within EarthCube

What is EarthCube?

1. It seems like everyone is thinking of EarthCube as an artifact or place to put stuff, which is very much like a data center.
2. Or, is EarthCube a set of entities supporting specific services?
3. Or, is EarthCube is a set of capabilities that allow different resources to be accessed.
    a. It is not a repository or fixed set of standards
    b. Brokering can allow each community to decide how data is exposed
4. Different groups have different goals and scope of EC
    a. You will hear different perspectives from each group depending on where they're coming from
5. Is EarthCube an Angie's list for data?
    a. In this way, only members can comment and participate

6. EarthCube is not a single data center
    a. Instead, it's more like the internet/cloud and users/data is subject to different standards, depending on what level you enter

Prioritization

1. It would be good to focus on low-hanging fruit, such as connecting data repositories, and using common standards to allow long-tail scientists to make data available.

Low barrier to entry

1. Is it possible to have a low barrier to entry and encourage people to contribute high quality data?
2. What does a low barrier to entry mean?
    a. Audience opinion 1:
        i. It implies people entering into a system, which can lead to low-quality, poorly document data.
        ii. There is no evidence that poorly-documented data leads to reproducible science.
    b. Audience opinion 2: A lower barrier to entry is more comprehensive than simple metadata standards
        i. It does not necessarily have to do with the quality of data
        ii. Instead, it is a goal regarding the system, and guides the system, encouraging a change in mindset on data sharing
3. Metadata
    a. If the entry barrier is too high, people will not contribute
4. EarthCube could be like the EBay of data sharing platforms
    a. EBay has a low barrier to entry, but there is a required amount of metadata
    b. Seller ratings helps others evaluate quality - social feedback mechanism to move people up the ramp

Data Quality

1. Maybe EarthCube is analogous to the web
    a. At first, the content was low quality, but as the web grew, people were better able to sift good content from bad content
    b. Maybe community consensus will help sort good data from bad data
2. Data documentation
    a. Poorly documented data sets in data centers doesn't work
        i. There needs to be data and documentation
        ii. Data value depends on documentation
        iii. Creating another large collection of poorly documented data sets won't help anyone

        iv. Instead of minimum standards, what are the maximum standards?
3. Finding data
    a. It is difficult to know what data is out there
4. In terms of data quality, where do you start?
    a. How do you determine minimal standards?
    b. EarthCube would be more useful and would accelerate faster if data quality is higher
        i. Better to start off with higher quality data
        ii. If time and data contributor are the only identifiers, data will likely not be used
5. User-support services is another means to improving data quality
    a. NASA and NOAA already have a structure for doing this
    b. NSF should take advantage of existing structure instead of inventing another place that people can upload data to
    c. But, how to make that structure work is a good question

Data sharing

1. Assumption: people want to share their data
    a. Many researchers don't want to share data at this point in time
    b. It doesn't matter if there is an NSF mandate
    c. How do you facilitate a mindset change so that people want to participate?
        i. A potential solution is to demonstrate that their data can be used effectively and well.
        ii. Some data producers want to share with people who will use data well
            1. There are many different users; some use data well, but others don't
2. Assumption: separating data providers from consumers
    a. But, providers are also consumers
    b. There should be a feedback mechanism
        i. Improve metadata to allows providers and users to access more of other people's data too - go up data continuum and have access to more and more capabilities
3. A lot of work is needed to make data more easily understood by the general public so that they use it well
    a. At this time, data providers don't have time and funding to meet all requirements unless there are funding incentives
    b. A change will come when funding agencies focus on metadata and sharing
4. It is better to focus on data sharing than spend 10 years determining minimum/maximum standards for contributing data

Compliance and or 'value' scales

1. Solution: value scale continuum
    a. Low entry barrier for sharing and accessing data at first, but possibility to move up continuum

    b. As scientists move up by achieving data or technological 'readiness levels,' they can access more capabilities
    c. Large, projects can pull the long-tail scientists up the continuum
2. The word 'compliance' turns people off. It would be better to focus on the value that scientists will get out of using EarthCube. People need to be convinced that participating at all is a good thing, and the benefits of moving up in readiness levels need to be demonstrated.

Prior NSF Investments

1. How much people at EC have thought of prior NSF investments?
2. DataNet and other infrastructures are relevant to EarthCube
3. 4-d Weather Cube
    a. It would be good to see where things went wrong and not repeat past mistakes
    b. The Weather Cube principles are in alignment with EarthCube, but the scope is much smaller
    c. The project still exists, but there is a holdup in getting RFPs out to build system

Semantics and Ontologies

1. A critical aspect is collecting and cataloguing vocabularies
    a. EarthCube should avoid creating new vocabularies
    b. What kind of languages to encode meaning?
    c. Need an inventory of existing semantics and ontologies
        i. There should be a template for to catalog existing Semantics and Ontologies
        ii. People can fill in the spreadsheet

CROSS-DOMAIN INTEROPERABILITY

For more information, please see the ESIP Commons EarthCube Cross-Domain Interoperability Session at http://commons.esipfed.org/node/468 and additional notes at https://docs.google.com/document/d/1lT0cTwNsIE5LRvH6foKW_w2TeGXCU2tgT9hNV_13GXE/edit . Kelly Monteleone and Genevieve Pearthree took notes.

Governance Issues

1. How do you combine the Data Discovery, Access and Mining Working Group, and the Cross-Domain Interoperability Concept Team?
    a. Governance needs to scope the problem
2. Governance and Geoscience Interoperability Institute
    a. It looks like small groups contribute to large group in the center
    b. Training and outreach are very important
    c. Focus on geospatial standards
    d. Some pieces are governed and some are not… depends on scope of EarthCube

    e. Governance that tracks metrics, shows success, etc. should be part of governance

1. The Workflow group
    a. This group has a similar diagram
        i. There is some confusion as to which group does what – each group is involved in governance
2. Governance Group
    a. The Governance Group has a liaison from each of the other EarthCube groups
    b. Governance Group members are also PIs of other work groups
    c. There were a number of webinars/workshops to analyze governance aspects and extract governance needs from each group's roadmap.
        i. The Governance Group is starting to follow up with groups asking what they need to operate internally
        ii. This information is then fed back into other groups, thereby providing services back to each group, which can help extract their requirements to develop a system to facilitate interaction between groups
3. EarthCube Governance
    a. Start building the platform for EarthCube & establish agreements about interfaces to create and manage the data
    b. There are technical and social means to convince people of the benefits of using EarthCube
        i. This is in concordance with research on historical infrastructure development, in which solutions have a technical or social component or both
    c. It would be good to first define domains of ownership, then what is shared across domains and how it is shared
        i. Governance will weigh in on how it happens

Additional Comments and Questions

1. Would the GII be a representative for the semantic group?
    a. The semantic group would be integrated into the GII.
2. Would brokering be part of the solution?
    a. Yes, because brokering can work with any type of input.
3. Data
    a. Can we assume archives will be part of the architecture?
    b. There needs to be a way to get feedback to the data center to improve access and tools for the next user
    c. One option is a "fitness for use workflow" which will provide additional annotation (what the schema will be) –to render data reusable
    d. There are mechanism for users to record measures for quality and results of test or compliance against those datasets but they need user input
    e. Is a low barrier to entry a good thing?
        i. It could be an evolving process

BROKERING

For more information on this workshop, please see the ESIP Commons EarthCube Brokering Session at http://commons.esipfed.org/node/482 and notes at: Kelly Monteleone took notes.

Governance questions

1. How should the brokers be deployed and maintained?
2. Policies for brokering across boundaries?
3. Published data, information, and knowledge standards – is there a convergent set of standards?
4. What is the process for evolving and innovating cyberinfrastructure?
5. Everyone in EarthCube seems to be waiting on some other group
    a. EarthCube needs to be community driven and engaging the community to make progress will take time
    b. The Governance team will likely have something in place this year

Additional Questions

1. How loosely or tightly coupled will the components be?
    a. This is not defining that – semantic capability – it does not have to be there
    b. Semantics should be outside the broker
2. In EarthCube – is there a broker or a type of broker?
3. In the future there will be interoperable brokers that are used by EarthCube
4. Have you thought about existing standards for brokers, for example Wps?
    a. Are there existing mappings on how to chain brokers?
    b. Existing standards/processes need to be used but this hasn't been done yet
5. Do you have a registry of brokers?
    a. There is a list, but it is not a large list
    b. It has been around for a long time
    c. There are other areas outside of the domain sciences that have used them for a while
6. Are there common use cases for EarthCube?
    a. Some groups have them and some don't
    b. Everyone is working on compiling the use cases, but there needs to be a primary activity to inform the collection and analysis of use cases.

GEOSCIENCES AUSTRALIA SEMINAR AND ROUNDTABLE

CANBERRA, AUSTRALIA, JULY 30, 2012

Lee Allison, chair of the EarthCube Governance Steering Committee, gave an invited seminar on cyberinfrastructure in the U.S. at the national headquarters of Geosciences Australia (GA) in Canberra, Australia in advance of the International Geological Congress in Brisbane. He was specifically asked to discuss developments in EarthCube. Afterwards, there was an informal roundtable with GA representatives and from a number of key leaders from other federal agencies that Lesley Wyborn of GA organized. Participants spent about 4 hours talking, with numerous questions about EarthCube. GA's Chief Scientist and CIO both participated. They shared their strategic planning on national infrastructure including cyberinfrastructure and expressed strong interest in further collaborations.

34TH INTERNATIONAL GEOLOGICAL CONGRESS

BRISBANE, AUSTRALIA, AUGUST 5-10, 2012

Informatics was a major theme throughout the program. Arizona Geological Survey (AZGS) representatives gave 6 informatics talks and posters, and questions about EarthCube were raised in every one of them. Also, the OneGeology Operational Management Group met on Sunday before the conference for a day-long strategic planning session. The group is very receptive to a more formal collaboration with EarthCube through the Governance Framework engagement process, so Lee Allison will be writing a letter to Dr. Alex Malahof, CEO of GNS (New Zealand geological survey) who chairs the OneGeology Steering Committee, outlining a proposal. OneGeology has 117 countries that have signed the international protocol, through their national geological surveys, so it provides a tremendous global network in the solid earth sciences. On August 5, OneGeology held a Town Hall Forum that drew principally directors of national geological surveys from around the world, and others, into a discussion about sustainability of OneGeology as a component of geoscience cyberinfrastructure. Dr. Allison gave an impromptu summary of the research into infrastructure described in the EarthCube Governance Roadmap, and its application to OneGeology.

The bottom line is that Australia (and the rest of the world) appears deeply interested in EarthCube and there are some fantastic efforts underway in Australia, the European Union (particularly via the INSPIRE initiative) and globally that can complement beautifully with EarthCube. EarthCube should expect a receptive audience based on initial responses.

COMMUNITY INPUT TO–DATE: EARTHCUBE GOVERNANCE NING SITE

Community input was a key component in developing the initial EarthCube Governance Framework Document. Although most input occurred during the weekly Governance Steering Committee virtual webinars, several important comments were posted to the EarthCube Governance Ning site since the June EarthCube charrette. These comments are pasted below.

IS THERE A MEMBERSHIP MODEL FOR EARTHCUBE?

Posted by Bruce Caron on August 2, 2012 at 2:37pm in Governance

One of the foundational moments for any "community led" organization is to determine the membership model for the community. This will not only announce who (what individuals or organizations) can be members, but also a set of responsibilities and benefits for membership.

Without this moment, the community remains external to the conversation about governance. After this moment, the members become owners of the process.

Without a membership model, a virtual organization is never community "led", but rather it assumes a posture of being of some benefit to "the community," which is described in some way. Meaningful community participation can only happen when the community is invited to become members in the organization, and when membership is given the keys to this car and the right to drive it in any direction it desires.

So, is there a membership model for EarthCube?

Reply by Sky Bristol on August 2, 2012 at 4:01pm

My take is that EarthCube is a multifaceted and still evolving organism that involves at least the following distinct units, which all relate in some way to the question of membership:

- NSF as an organization putting money and energy into leadership and facilitation along with developing funding opportunities associated with the concept
- Other organizations like my own (USGS) deciding that EarthCube sounds like a good concept that aligns with our own mission goals and then create the opportunity for people in our organization or funded by us to take part in EarthCube
- Other communities such as ESIP or standards bodies such as OGC that find resonance with the concept of EarthCube and can articulate value for their organizations or individual members to be involved
- A Community of Practice with interested practitioners who get involved to greater or lesser degrees with an interest in leveraging the community to further their own practices and in turn contribute to the furtherance of the community as a whole

- Individuals or groups of entrepreneurs who gravitate toward the possibility for funding resources or partnerships which may help advance their agendas, which may be in direct line with or tangential to the evolving vision of the whole

Each of these aspects of the EarthCube picture has its own dynamics associated with participation, contribution, and "membership." Some or all of them may have a time horizon (isn't EarthCube a 10-year initiative from NSF's perspective?).

Of these, the Community of Practice aspect probably lends itself best to establishing a membership model. I personally ascribe to and have used many of the tenets from Etienne Wenger's work in this area. I particularly like his fairly simple diagram showing the dance between formal organization and communities of practice in Cultivating Communities of Practice.

Wenger points out the difference between communities of practice and institutional entities and argues that both have an important symbiotic role to play in advancing an enterprise (Wenger 1998). Our own experience in USGS has experimented with varying levels of "power under" vs. "power over" methods in tuning a community toward institutional objectives, and we've found marked differences in community participation and output when too much executive direction is applied. My point here is that there is a relationship between the institutional entities participating in the EarthCube enterprise (NSF chief among them) and the community of practitioners forming up to determine the shape and direction of that enterprise. I believe it would behoove the enterprise to formally establish the community of practice aspect and develop the operational model.

ESIP offers a pretty highly functioning model and story in the evolution of a community from its roots at NASA in the late 90s. It has both a formal institutional entity in the ESIP Federation with a membership model and process for participating organizations (including my own). Those organizations contribute to greater or lesser degrees for the institutional entity to operate in facilitating the Earth Science Information Partners, the Community of Practice, which is principle business of the ESIP Federation. Part of that facilitation has been in establishing and maintaining governance concepts such as a constitution and bylaws, and from the Community of Practice standpoint, the useful structures of Committees, Working Groups, and Clusters to help focus and channel creative energy toward mutually beneficial outcomes (visit http://esipfed.org/).

I would even go so far as to ponder whether or not EarthCube would be better off to simply put more energy and other resources into ESIP to help bring about the Community or Practice aspects needed by EarthCube. Do the missions and aspirations align closely enough to entertain the idea?

At any rate, I do believe that there ought to be a formal Community of Practice entity for EarthCube that we start evolving. It should have a membership model that includes both organizational membership in the institutional entity supporting the Community so we have a mechanism for investment from organizations like mine and some of those other types I mentioned along with a model for how community members best fit in, contribute, and gain benefit from the community. ESIP and others should be examined for partnership and applicability of concepts.

Reply by SiriJodha Khalsa on August 2, 2012 at 4:28pm

Interesting question to ponder. Is EarthCube an organization? Or is it a collection of resources (data, information and knowledge made accessible and usable through ICT) with stakeholders - implementers, beneficiaries, sponsors and the like? I would agree that the governance body for EarthCube will be an organization, with membership. But what of the EarthCube community - people participating in the process of realizing the EarthCube vision? Membership required if funded to be part of this community, or voluntary if not. Certainly no one with ideas to offer is turned away. I can't envision a circumstance, beyond a governing body, in which formal membership would be required to be part of the EarthCube community.

Reply by Bruce Caron on August 6, 2012 at 2:38pm

This might also raise the question of who selects the governing body? Who votes? Who can be elected? It also gets to the scope and rewards for people volunteering their time. If the only organization is a fully-funded project, then this becomes external to the stakeholders it claims to support. The project becomes a "community-for" effort, not a "community-led" effort.

Maybe doing work "for" a broad, not-entirely-defined EarthCube community is how this effort will move ahead. That has certainly been the way the NSF has funded networked groups in the past.

REVISED CLASSIFICATION OF EARTHCUBE GOVERNANCE FUNCTIONS - OPEN FOR EDITING!

Posted by Genevieve Pearthree on August 3, 2012 at 4:51pm in Governance

These documents represent a first step in developing the Governance Framework we plan to submit to NSF on August 15th.

Please edit these documents between now and Tuesday, August 7th; at this time we will be holding a virtual meeting to collaboratively edit these documents.
Meeting time: Tuesday, August 7th, 11 am EDT (8am (PDT) – call in details to come…

- Document 1: Editing guidance for EarthCube Functions and 'Super-Functions' - PLEASE READ
- Document 2: EarthCube Functions and 'Super-Functions'
- Document 3: Governance Guiding Principles
- Document 4: General Governance Recommendations

Thanks for all your help thus far!

Reply by SiriJodha Khalsa on August 10, 2012

The function "Community-based repositories, standards and governance structures for EC compliant" needs an action verb like "establish". Does the following accurately convey the meaning?:

Establish governance structures that will determine standards upon which EC-compliance can be determined for community-based repositories.

It's a bit hard to imagine that a top priority is to set up standards by which community repositories will be evaluated for EarthCube compliance. Is this really what's intended here?

Reply by SiriJodha Khalsa on August 10, 2012

I'm afraid I don't fully understand the super-functions. As categories of governance functions I can understand "set policies" and "oversee/manage infrastructure development and operations." Infrastructure comprises many different kinds of services, so why is a "service" category necessary? i.e. what doesn't fall under either policy or infrastructure development and maintenance?

Reply by SiriJodha Khalsa on August 10, 2012

It seems to me that the crux of the touch-point paradigm is how the functions/services (Compute, Visualize, Publish, Discover, Archive, Access/integrate, Sense/collect Manage) are categorized as being either enterprise, touchpoint or community governed (with both enterprise and touchpoint functions being managed by the enterprise, i.e. EarthCube, and touchpoint being tailored/optimized locally).

Is it expected that any initiative or community with an established infrastructure that wishes to be part of EarthCube will have to agree to have those functions deemed enterprise or touchpoint managed by EarthCube? If so, seems that a high priority policy governance function is to establish what these enterprise and touchpoint functions are.

What will be the mechanism that will "logically connect" catalogs and repositories? Will it be mandating conformance to standards and protocols? Or will it be by providing a mediation service that will allow existing community infrastructures to continue doing business as they do now? If history is any indication, the former approach is likely to meet with much resistance. If the latter approach is being considered, then I would promote the unclassified functions related to brokering (47-51), such as "Policies for brokering across boundary between communities and EC cyberinfrastructure" to classified under policy and infrastructure super-functions.

Reply by Erin Robinson on August 11, 2012

SJ - These are all valid points and questions. The functions are evolving through conversations like this one and will continue to evolve rapidly over the next few weeks and in a more controlled manner over the next six months as we continue to actively engage with the broader Earth science community. The documents are editable, so if you have a suggested prioritization, I would encourage you to make changes directly as you see fit.

EARTHCUBE GOVERNANCE ROADMAP - FINAL VERSION

Posted by Genevieve Pearthree on August 15, 2012 at 5:40pm in Governance

Here is a final version of the EarthCube Governance Roadmap.

Although this version is more polished than the initial roadmap released June 7, 2012 prior to the June 2012 EarthCube Charrette, there are very few substantive changes in content. Instead of making substantive changes, the Governance Working Group felt that it would be fitting to include new material in the initial governance framework document that will be presented to NSF and the EarthCube community no later than August 31, 2012.

This governance framework document will reflect the evolution in the group's thinking, particularly as new individuals have joined the discussion and a broader spectrum of EarthCube stakeholders have been engaged. Components of this document have already been posted to the EarthCube Governance Ning site and are available for public editing and comment.

Reply by Bruce Caron on Friday, August 17, 2012

There is a lot of great work in this document. In fact, if this was just delivered to an assembly of EarthCube members (however defined) as a starting point for their deliberations on governance, then this committee will have done their job.

I see a danger here of the initial governance deliberations becoming not just preliminary research for EarthCube members to use, and, instead, forging a governance plan before there are any EarthCube members (since these are not yet defined). Getting the EarthCube "community" to buy into governance is a lot easier when they are the ones who determine this and own these decisions.

Without a membership model (and slight attention to this was given in this document) and actual members, it might be a cart-before-the horse issue to make certain decisions now. Perhaps the one

most important decision is the membership model. Who are the members of EarthCube? What are the benefits of belonging? What are the responsibilities? How does the membership authorize decision making, including decisions on the framework for governance?

Whatever decisions the EarthCube members make, they will find this document to be of real value.

Reply by Genevieve Pearthree on Friday, August 17, 2012

Hi Bruce,

Thanks for your comments and insights. I agree that a membership model wasn't discussed very much in the roadmap document, although there is a community engagement section in one of the appendices. The governance group has spent much more time thinking about community building and engagement since the roadmap was written; those changes will be reflected in the governance framework documents to be submitted to NSF and the EarthCube community on August 31, 2012.

Your earlier post spurred discussions among the governance group about what a membership model might mean for EarthCube, and how it might be incorporated into the community engagement program to be carried out over the next 6 months. People agreed that there should be participation policies defined at the beginning, although at this point, it is still being fleshed out and people felt like a membership model that was too formal or detailed at this point might limit participation by any and all interested individuals and organizations.

However, I think the crux of the issue is that people have different ideas about what a 'membership model' actually means. It seems like what you're thinking about is defining decision-making processes, determining who holds what responsibilities, and demonstrating to potential stakeholders the benefits they gain by participating, while other people might think about having to pay dues in order to be a member. I agree that building the EarthCube community is one of the most important issues that need to be addressed at this time, but that first, there needs to be agreement on what that is, and how it will be done.

On this note, one of the most important components of what the governance group will be doing between now and early 2013 will be to engage potential EarthCube stakeholders to both introduce EarthCube and gauge their needs and wants in terms of governance and the earthcube.org website functionalities. The goals of this engagement process will be to collect feedback that will be used in making EarthCube something that people actually use, and building the EarthCube community.

If you have time, it would be great if you can provide feedback and insights based on your experiences into how you might envision a membership model for EarthCube, and what steps should be taken right now to make that happen. Also, take a look at the documents included in

these posts, which will let you know what's been done so far, and see if anything is missing (they're not complete yet so please edit as you see fit).

1. Governance Framework Narrative
2. Governance Functions
3. Governance Guiding Principles
4. General Recommendations

Thanks for your comments and I look forward to hearing from you soon.

Genevieve

Reply by Bruce Caron on Sunday, August 19, 2012

HI Genevieve

I can see that (and how) a membership model is not so easy for EarthCube at this time. A lot of ambiguity still in this effort... But a decision to not determine this is also a defining moment for EarthCube. This is when the project/organization will become either a community/stakeholder/member-led organization, or it will become a community/stakeholder-serving organization, in which governance is really just management of a large funded project or projects. In a member-led organization, engagement will happen in part through governance, and with volunteer support. In a stakeholder-serving organization, engagement will need to be more centered on the value added by the funded efforts.

I would say that building EarthCube as a member/community-led organization is a more difficult and in several ways more interesting path to take (and a path the NSF normally does not take). To do this, a membership model is an early step to make. In this case, it's not an easy step.

A couple notes: the membership model may be a federated model, where there are different types of members. For example, one type could be currently funded NSF projects that are producing/stewarding earth data, e.g., DataOne. Another type could be earth-data producing research units at universities. A third type could be earth scientists at non-research universities. Each of these types could have a place at the EarthCube governance table. Given the goals of EarthCube, inclusivity seems to be the operating logic.

I'm on vacation this week... I'll check out the documents when I get back from the forests.

cheers

Bruce

Reply by Genevieve Pearthree on Monday, August 20, 2012

Bruce,

Those are all important things to consider and I agree that now is the time to make initial governance decisions regarding a membership-serving or membership-led model. These issues have been discussed among governance working group members, but they haven't been stated so explicitly, so I'm glad you're bringing them up. I will pass along our discussion to the steering committee members, and if you'll be back from vacation by this Friday, I'd like to invite you to join the weekly call to talk about EarthCube membership. And, if not this week, definitely next week if you have time.

The documents that are posted to the Ning site are a first step in developing the initial governance framework documents. Although the governance group is planning on submitting them to NSF on August 31st, they will be revised to reflect community input gathered throughout the next 6 months. Whatever input you have at this point, and in the future, will be very valuable.

Enjoy your vacation!

Genevieve

ADDITIONAL COMMUNITY INPUT

This feedback is regarding the evolution of Emerging Governance Concept: Three Tiers of Governance (see Section 2.4.2).

"The governance diagram illustrates the challenges with integration of existing community resources. Each community resource has been developed independently, has unique access mechanisms, and deals with different types of data and different semantic ontologies.

However there are two places where interoperability is important:

- Research environment. This may be a laptop, or a cluster, or grid infrastructure (XSEDE, OSG, ESG).
- Collaboration environment. This may be a data grid, or a portal, or a workflow.

These environments correspond to levels in the governance diagram. Within the research environment, a researcher wants to be able to access the existing community resources, execute analyses (tightly coupled computations or loosely coupled workflows), and manage the research products. Within the collaboration environment the researcher wants to be able to share results, share workflows, re-execute workflows, track provenance, build collections, and publish results.

Multiple interoperability mechanisms are available for use within these environments:

- Brokers which translate between protocols, or semantics
- Standard web services (such as OGC) which provide uniform access across community resources
- Data grids which use brokers to translate between protocols, while providing unifying name spaces
- Portals, which link input data sets to applications that are run on a compute grid

It will be possible to build multiple versions of "glue" for integrating technologies. The XSEDE grid has had great success with Science Portals, which manage input data for scientific applications. DataONE, GeoBrain, and OGC provide standard web services. Data Grids such as iRODS are being used to manage collaboration environments and access to federal repositories.

An expectation is that each geoscience sub discipline will choose a different integration mechanism. An example is OOI, which is managing real-time sensor data using message bus technology. CUAHSI manages point data, and uses quite different technology. Both will have web service interfaces for access to their data, but the web services will be quite different. Either a portal or data grid could execute the multiple web services to extract data from both systems, and then manage execution of research analyses on the data."[132]

[132] Reagan Moore, (EarthCube Layered Architecture Concept Award Principal Investigator), personal communication, July 2, 2012.

APPENDIX 6: IMPORTANT DOCUMENTS AND WEBSITES

WORKING DOCUMENTS

The following is a list of working documents used throughout the development of this governance framework document. All are open to community editing and comment.

1. Governance Framework Document: https://docs.google.com/document/d/1UjOwPYBnI4uihlZsh57X1DK6MAknPa_wmlQK3UXMf5I/edit

2. General Governance Recommendations: https://docs.google.com/document/d/1OqDMwRndYSt8Tx2noJbJWizeHDQ9ssNAdoFnILg_90k/edit

3. Governance Guiding Principles https://docs.google.com/document/d/1HMSRwpTTPqSXnrspPdDd7LSU6bAvzHa0JxGm-AZkbzU/edit

4. Governance Functions https://docs.google.com/spreadsheet/ccc?key=0AhE8IKrxOMWndDZfTXBUUFE1QmhOT01XT2JrZ3BQSWc#gid=0

IMPORTANT DOCUMENTS AND WEBSITES

1. EarthCube Governance Roadmap, Version 2.0 http://earthcube.ning.com/group/governance/forum/topics/earthcube-governance-roadmap-final-version

2. EarthCube Community Group and Concept Team Draft Roadmaps http://earthcube.ning.com/page/draft-roadmap

3. June 2012 EarthCube Charrette Materials http://earthcube.ning.com/page/materials-june-2012

4. National Science Foundation EarthCube Website http://www.nsf.gov/geo/earthcube/

5. EarthCube Governance Website http://earthcube.ning.com/group/governance

6. EarthCube Workshops at the ESIP Federation Mid-Year Meeting, July 17-20, Madison, WI
    - EarthCube Governance Workshop Meeting Materials, July 17, 2012: https://docs.google.com/#folders/0BxE8IKrxOMWnOTZHczRCWUN0NVE

    - July 17th EarthCube Governance Workshop ESIP Commons Page: http://commons.esipfed.org/node/601

    - July 19th EarthCube Governance Presentation and Community Feedback: http://commons.esipfed.org/node/474

    - Other EarthCube ESIP Workshops: http://commons.esipfed.org/search/node/EarthCube

REFERENCES

Allison, M. Lee, Tim Ahern, David Arctur, Jim Bowring, Gary Crane, Cecelia DeLuca, Geoffrey Fox, Carroll Hood, Hannes Leetaru, Kerstin Lehnert, Chris MacDermaid, Mohan Ramamurthy, Erin Robinson, Ilya Zaslavsky, Genevieve Pearthree, Kim Patten, and Debra Winstead, "EarthCube Governance Roadmap: Documentation, Research, and Recommendations, Version 2.0," August, 2012

Committee on increasing National Resilience to Hazards and Disasters; Committee on Science, Engineering, and Public Policy; the National Academies, *Disaster Resilience: A National Imperative*, 2012.

EarthCube Data Discovery, Access and Mining Working Group, "A Community Roadmap for EarthCube Data: Discovery, Access, and Mining," draft roadmap released June 2012.

EarthCube Earth System Models Concept Award, "EarthCube Earth System Model Coupling Roadmap," draft roadmap released June 2012, 12.

EarthCube Layered Architecture Concept Team, "EarthCube Layered Architecture Roadmap," draft roadmap released June 2012.

EarthCube Web Services Concept Award, "Web Services Concept Group Roadmap," draft roadmap released June 2012.

EarthSystem Commodity Governance Project, "Governance," last modified 2012, http://earthsystemcog.org/projects/cog/governance_object

Gardler, Ross and Gabriel Hanganu, "Governance Models." Open Source Software Watch, last modified February 14, 2012. http://www.osswatch.ac.uk/resources/governanceModels.xml.

Gil, Yolanda, Aaron Braekel, Ewa Deelman, Ibrahim Demir, Christopher Duffy, Suresh Marru, Marlon Pierce, and Gary Weiner, "EarthCube Workflows Community Group Report," draft roadmap released June 2012.

Hood, Carroll, "Abstractions on EarthCube Governance," Virtual presentation to the EarthCube Governance Steering Committee, July 27 2012.

Hood, Carroll, "Developing an Optimal Governance Framework," White Paper Presented to the EarthCube Governance Steering Committee, 2012.

Khalsa, Siri Jodha, Mark Parsons, Ruth Duerr, Jay Pearlman, Francoise Pearlman, Steve Browdy, Stefano Nativi, Erin Robinson, and Ben Dominico, "Roadmap for Developing of Brokering as a Component of EarthCube," draft roadmap released June 2012.

Khalsa, SiriJodha, Mark Parsons, Ruth Duerr, Jay Pearlman, Francoise Pearlman, Steve Browdy, Stefano Nativi, Erin Robinson, and Ben Dominico, "Executive Summary for the Roadmap for Developing of Brokering as a Component of EarthCube," draft executive summary released June 2012.

Khalsa, SiriJodha, Jay Pearlman, Francoise Pearlman, Tim Ahern, Jennifer Arrigo, Kerstin Lehnert, Chris MacDermaid, Janet Fredericks, Don Middleton, Reagan Moore, Stefano Nativi, Scott Peckham, Krishna Sinha, Anne Wilson, and Ilya Zaslavsky, "Summary and Recommendations from EarthCube CA PI Meeting held at NSIDC, University of Colorado, Boulder Colorado, July 10 2012," 2010.

Khalsa, SiriJodha, Mark Parsons, Ruth Duerr, Jay Pearlman, Francoise Pearlman, Steve Browdy, Stefano Nativi, Erin Robinson, and Ben Dominico, "Brokering for EarthCube Communities: A Road Map," updated roadmap released August 2012.

Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers, *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute, 2011, 156p.

National Science Foundation, "EarthCube—Building the Cyberinfrastructure to Better Understand Our Complex and Changing Planet," 2011.

Sinha, Krishna, Gary Berg-Cross, Anne Thessen, Nancy Wiegand, Amit Sheth, Natasha Noy, Calvin Barnes, Leo Obrst, and Clinton Smyth, "Roadmap for Creating the Semantic/Ontologic Infrastructure for the Geosciences," draft roadmap released June 2012.

U.S. Geological Survey, "Science Strategy for Core Science Systems in the U.S. Geological Survey, 2013–2023," Public Review Release, USGS Open-File Report 2012-1093, 2012.

Vaughn, Shawn, "NASCIO Welcomes the Appointment of FirstNet Board Members," http://www.nascio.org/newsroom/pressrelease.cfm?id=131, last modified August 20, 2012.

Weill, Peter, and Jeanne Ross. *IT Governance: How Top Performers Manage IT Decision Rights for Superior Results*. Boston: Harvard Business School Press, 2004.

Zaslavsky, Ilya, Ilkay Altintas, David Arctur, Jerry Brotzge, Alva Couch, Philip Dibner, Ben Dominico, Rick Hooper, Kerstin Lehnert, Philip Murphy, Stefano Nativi, Beth Pale, Steve Richard, Karen Stocks, and David Valentine, "EarthCube Roadmap," draft roadmap released June 2012.

Zaslavsky, Ilya, Ilkay Altintas, David Arctur, Jerry Brotzge, Alva Couch, Philip Dibner, Ben Dominico, Rick Hooper, Kerstin Lehnert, Philip Murphy, Stefano Nativi, Beth Pale, Steve Richard, Karen Stocks, and David Valentine, "EarthCube Roadmap, Version 1.1," updated roadmap released August 2012.