Title
Negotiating Science through Policy: EarthCube, infrastructure and policy-relevant science.

Author
Slota, Stephen

Publication Date
2017

UNIVERSITY OF CALIFORNIA,
IRVINE

Negotiating Science through Policy:
EarthCube, infrastructure and policy-relevant science.

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Informatics

by

Stephen C. Slota

Dissertation Committee:
Professor Geoffrey C. Bowker, Chair
Professor Gary Olson
Associate Professor David Ribes

2017

DEDICATION

To

My friends and family,

My mentors, teachers, and colleagues.

To

All who helped and all who hindered

Thank you.

In exhumed mud squirms
a link of living chain
that brings forth clay and water
and, hidden from sight,
is mirror, scaffold
and pure text, endlessly rewritten.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ACKNOWLEDGMENTS

I would like to thank my chair and advisor, Geoffrey C. Bowker, without whose mentorship, advice, conversation and unwavering enthusiasm and support this thesis would have never been completed.

Thank you Gary Olson and David Ribes for your time in reading and discussing this work as members of my thesis committee. Your comments were insightful and challenging, and I am grateful for your perspective, your arguments, and your attention.

Thanks to Judith Gregory, John Seberger, Siri Jodha Sing Khalsa, Cory Knobel, and all the people with whom I discussed this research in process. Your advice, conversation and assistance is greatly appreciated.

Thanks to my family, both here and passed on, whose confidence in me was a great help when I lacked confidence in myself.

Thank you to the members of the BCube team for accommodating and accepting me as a participant observer.

CURRICULUM VITAE

Stephen C. Slota
412-973-6522
steveslota@gmail.com

EDUCATION

Ph.D. Defended November 14, 2017; University of California, Irvine; Irvine, California; Dept. of Informatics
MLIS, 2008; University of Pittsburgh; Pittsburgh, PA.
B.A. English, 2006; Messiah College; Harrisburg, PA.

PROFESSIONAL ACADEMIC EXPERIENCE

2017 - ; Post-doctoral researcher (Following Upcoming Graduation): "Collaborative Research: Institutionalizing the Data Sciences, a Sociotechnical Investigation of BDHubs," NSF Award (#1638932)

2014 - 2016; Adjunct Professor; Carlow University; Pittsburgh, PA
- NU725 Statistical Applications

2012 - ; Research Assistant; University of California, Irvine; Irvine, CA

2012 - 2016; Teaching Assistant; University of California, Irvine; Irvine, CA
- I&C Sci 139W: Critical Writing
    - Spring Quarter 2013
    - Summer Quarter I 2015
    - Summer Quarter II 2015
    - Fall Quarter 2015
    - Summer Quarter II 2016
- IN4MATX 261: Social Analysis of Computing: Fall Quarter 2013
- I&C Sci 3 LEC A: Internet, Technology and Society: Fall Quarter 2014
- I&C Sci 166 LEC A: Game Design: Spring Quarter 2015
- IN4MATX 163: Projects in Social Impacts of Technology: Spring Quarter 2016

2009-2011; Graduate Student Researcher; University of Pittsburgh; Pittsburgh, PA

2010; Teaching Fellow; University of Pittsburgh; Pittsburgh, PA
- LIS2186 Information Policy, Fall Semester 2010

2008-2009; Information Assistant; Carnegie Mellon University Archives; Pittsburgh, PA

FUNDING AND AWARDS

2016; NSF Award (#1638932) of $350,000: Co-Author (Graduate Student Assistant); "Collaborative Research: Institutionalizing the Data Sciences, a Sociotechnical Investigation of BDHubs," PIs: Geoffrey Bowker and David Ribes. Awarded August 1, 2016

2015; Newkirk Graduate Student Fellowship; $8,000

2013; NSF Award (#1343802) of $80,000; Funded member of "EarthCube Building Blocks: A Broker Framework for Next Generation Geoscience (BCube)" PIs: Siri Jodha Khalsa; Stefano Nativi, Ruth Duerr, Francoise Pearlman, Jay Pearlman;

PEER-REVIEWED WRITING

Stephen C. Slota and Geoffrey C. Bowker. "Chapter 18: How Infrastructures Matter," in Ulrike Felt, Rayvon Fouché, Clark A. Miller, and Laurel Smith-Doerr, eds. *Handbook of Science and Technology Studies, 4th Edition* The MIT Press, 2016

CONFERENCE PRESENTATIONS

Slota, S. 2017. "Bridging Science and Policy: Knowledge production, domain formation, and organization. 4S Annual Meeting 2017; Boston, MA

Slota, S., & Bowker, G. C. (2015). On the value of 'useless data': Infrastructures, biodiversity, and policy. *iConference 2015 Proceedings*. Available at http://hdl.handle.net/2142/73663

INVITED TALKS

"What does scholarship mean in the age of Wikileaks?" Digital Media @ Pitt Downloads: Critical Forums in Mediated Culture. Pittsburgh, PA. March 28, 2011

OTHER GRANTS AND PROPOSALS

2016; Co-Author; "Changing the Rules of the Game: Shaping Sustainability Standards"; PIs: Kaminsky, J. and Muench, S. Submitted to NSF Summer 2016. Resubmission, previously judged to be competitive.

2015; Co-Author; "RCN Proposal: The Transdiciplinary Field of Big Data, Ethics and Values" PI: Geoffrey Bowker; Submitted to NSF October 14, 2015. Resubmission planned for early 2018.

AFFILIATIONS

2017-; Post-Doctoral Researcher; DeLab; University of Washington; Seattle, WA
2014-2018; Senior Student Member; EVOKE Lab; University of California, Irvine; Irvine, CA
2012-2014; Student Director; EVOKE Lab; University of California, Irvine; Irvine, CA
2009-2011; Chairman/President; University of Pittsburgh Student Chapter of ASIS&T

OTHER PRESENTATIONS AND PANELS

Slota, S., Ramdeen, S. et al. "From Invention to Innovation in EarthCube with Examples for Disseminating Software Sharing and Science Reproducibility through Geoscience Papers of

the Future”
EarthCube All Hands Meeting 2015
Washington, D.C.; May 27-29, 2015

Slota S. “On the Social Dimensions of EarthCube Cyber-infrastructure”
EarthCube All-hands Meeting 2014
Washington, D.C.; June 23-26, 2014

“Leveling the Stage? Performance Curation and Embodied Knowledge as Knowledge Infrastructure”
Society for the Social Studies of Science (4S); Annual meeting 2013
San Diego, CA.; October 10, 2013

WORKSHOPS ORGANIZED

2012; Organizer/Participant; Values in the Design of Information Systems Workshop, University of California, Irvine; Irvine, CA

2010; Organizer; Values in the Design of Information Systems Workshop; New York, NY

ABSTRACT OF THE DISSERTATION

Negotiating Science through Policy:
EarthCube, infrastructure and policy-relevant science.

By

Stephen C. Slota

Doctor of Philosophy in Informatics

University of California, Irvine, 2017

Professor Geoffrey C. Bowker, Chair.

The NSF has supported early forms of scientific cyberinfrastructure from the 1960s. Since about 2000, however, new cyberinfrastructure (CI) initiatives have gathered momentum, guided by an increasingly comprehensive vision of CI as a principal agent of change for a new era of large-scale, distributed, data-intensive, collaborative science in virtually every domain. EarthCube, as a large-scale international CI project provides a fertile ground for the observation of the negotiation and policy work necessary to facilitate distant collaboration and the accumulation, provision and sharing of scientific knowledge and resources (data, tools, models). As a collaborative design process drawing from both the policy and scientific worlds, EarthCube is a site where the complex relationships and negotiations between governance and particular visions of novel science are particularly evident as well as being a microcosm of the co-constructive processes that produce relevant knowledge for both policy and science. This writing explores the relationship of policy work and the production of scientific knowledge, and provides an account of the complex interactions between regulatory, legislative, and organizational work with

knowledge production practices in the geosciences through the lens of research funding and cyberinfrastructure development both historically and ethnographically.

CHAPTER 1: INTRODUCTION

Science does not occur in a vacuum. In addition to accounting for the work of other scientists, historical knowledge, and the development of innovative technology techniques to enhance observational and analytic capacity, science writ large also responds to and engages with social and political goals, outcomes, or concerns. These social and political factors are vanishingly present in the discursive concept of scientific and social relevance, but become significantly more visible when the outcomes of scientific work directly affect legislation, regulatory regimes, and become enmeshed into monitoring and management processes. Even more visibly, science funding goals and decisions reflect a politically-directed notion of innovation, a particular conception of state of the art of a field of scientific inquiry, and, when knowledge infrastructure (Edwards et al, 2013) enters the conversation, the selection, adoption, and necessity of technology, standards and particular systems in pursuit of that science. Knowledge infrastructures, as defined by Edwards (2010) are "robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds." As we as humans seek to understand, manage, and predict the functioning of the world and its various systems, the direction of scientific knowledge production becomes not just an issue of supporting novelty and new learning, but also an issue of management and monitoring. And so, as scientists work to produce an understanding of the world, policymakers, legislators, and science funders simultaneously produce knowledge about that science in order to more effectively direct and manage its progress towards political, economic, or conservation outcomes. At the core of my discussion throughout this writing is the following question:

how does policy-level knowledge influence, account for, and respond to the production of scientific knowledge in terms of regulation, funding, and infrastructure development? Throughout this work I will be exploring the relationship between scientific knowledge production and the production of knowledge in support of legislation, regulation and management (which I loosely conflate to 'policy concerns'). To do this, I will trace the impingement of policy at a variety of scales on the production of scientific knowledge, with a particular focus on the development of shared knowledge, resources, technologies and techniques as supported and interpreted at some level by law. These scales include organizational policies on data access, particular acts of legislation in both modern and historic context, regimes for the management and monitoring of natural systems, and the process of science funding in the field of Earth Systems Science (ESS), or GeoScience. Fundamental to this analysis is the notion that tracing the design and development of knowledge infrastructure reveals and defines a particular relationship between policy (as realized in legislation, public statements, regulatory/monitoring frameworks, and organizational issues) and science. This is not a unique relationship, nor is it the only pathway to novel science and collaboration. It is characterized by processes of mutual fitting and co-production, (Jasanoff, 2004) and is iterative, responsive to change and cyclical.

The arguments I present here are jointly informed by ethnographic field work performed over a three-year period working with a 'building block' (called BCube) of a cyberinfrastructural project funded by the National Science Foundation (NSF) called EarthCube and documentary analysis of cases that illuminate particularities of the relationship between political goals and scientific knowledge production in a way that, while present in some ways in my field site, become significantly more analytically visible in historic context. In order to present an account of how scientific knowledge is reflected, accounted for, and acted upon in legislative and regulatory action, I examine the early history of Chesapeake Bay governance and management, and compare it to the more modern development of the Total Maximum Daily Load (TMDL) figure, which is a computational model-derived abstraction of the quantity of pollutants in a particular watershed over a period of time, and represents both a means of managing watershed cleanliness and an object of scientific inquiry in its own right. Based on the conclusions I draw from that modern and historic context, I look closely at the funding of technological, data, organizational and collaborative systems (collectively referred to as cyberinfrastructure (CI)) intended to support and enable the production of innovative science. In this analysis, I will show how organizational policy either supports or limits infrastructural change, and examine the role of science funding agencies (the NSF in particular) in selecting technologies and standards, co-producing a discourse on relevance, constituting and understanding a community of scientists, and directing the design and implementation of cyberinfrastructure through various mechanisms. In addition, I will present a close account of a project funded under the larger umbrella of EarthCube, the BCube middleware data broker, to examine how the development of a particular piece of infrastructural software both responded to and was limited by issues of organizational and governmental policy, and draw from that a perspective on evaluating the amenability of particular infrastructures to change that takes into account the multiplicity and heterogeneity of infrastructure. In this introductory chapter I will briefly present the

EarthCube and BCube projects, and discuss an analytic perspective on policy implementation that informs my discussion throughout the remainder of this work.

A Brief Description of EarthCube and BCube

The majority of this document is concerned with my observations and analysis made as a participant observer working in a 'building block' of the NSF-funded EarthCube infrastructure project. The EarthCube project was initiated in 2011 with the goal of supporting transformative geoscience through tool development, data access, and community building. The design objective of the EarthCube project was to build a data- and tool-sharing infrastructure that supported novel collaboration and innovative work in the broadly-defined field of Earth Systems Science, or Geoscience, which includes areas of research such as hydrology, oceaonography, climate and atmospheric studies, seismology, geology and more specialized areas of inquiry related to some earth-scale system. In addition to collaboration between different disciplines of science, EarthCube also sought to involve computer science and technology development. EarthCube was to achieve that goal through funding a series of relatively independent 'building block' projects that could be pieces of software, resource registries, or technologies, as well as data centers and research coordination networks that, coupled with follow-up grants for integrative activities and architectures, would eventually be linked together into a community-designed infrastructure that could support cross-disciplinary collaboration and the broad re-use of data.

The project I participated and studied most closely was a 'building block' grant awarded in 2011 and begun in 2012, called BCube. BCube was middleware data broker – a piece of software that operates 'in between' a data source and whatever analytic software an end-user might be employing. BCube, in short, was intended to provide a seamless many-to-many experience for its users through data standard and format translation, automated access to and synthesis of multiple data sources, and providing an automated registry of additional data sources. (Figure 1) This made BCube potentially infrastructural to the proposed EarthCube cyberinfrastructure in that it would ideally automate a variety of collaborative activities by linking data sets with different formats and standards, providing a means of collecting similar data from a variety of sources, manage metadata conflicts and differences, and potentially being able to synthesize, on the fly, data collected in very different ways – in particular gridded data from satellite images with point data from sensors.

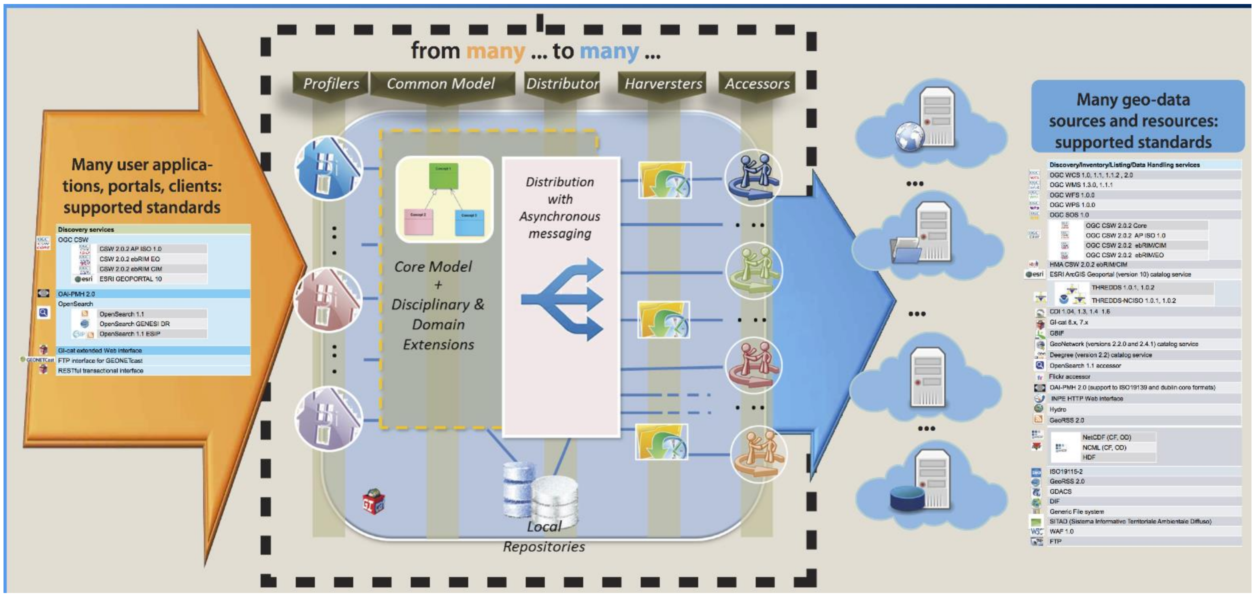


Figure 1: BCube 'Many-to-Many' model. (Khalsa and Nativi, 2012)

The BCube project was a collaboration of technology-focused computer scientists with a representative variety of collaborating domain scientists from hydrology, oceaonography, climate science, and polar sciences, as well as a single social scientist (me). BCube was intended to implement its broker into a test-bed in order to demonstrate its capability as a tool in the context of EarthCube's goals, as well as produce a specialized web crawler and to automate the process of gaining access to new data sources. Once implemented, new capabilities for and demonstrations of the broker would be produced on the basis of 'science scenarios,' or use cases that answer a particular scientific question. These scenarios would begin with individual domains, then begin to incorporate cross-domain questions. (Figure 2)

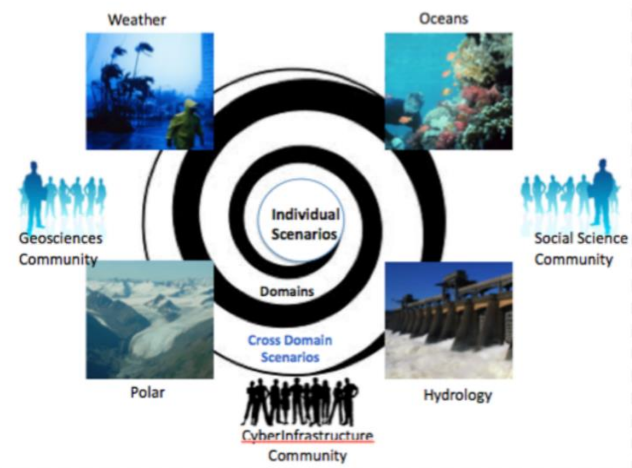


Figure 2: BCube Implementation Plan. (Khalsa and Nativi, 2012)

While I participated most closely with the BCube team, I also maintained participatory and observational roles in broader EarthCube governance and organizational activities. This provided me with a perspective on the development of infrastructure that spanned several scales of activity. From BCube I present an image of the ground-level of project work, its negotiation with organizational policy and response to funding concerns, and the processes of fitting, adjustment and iteration that any project goes through as new obstacles and opportunities are discovered. From EarthCube I present a broader conceptualization of the role of the NSF in supporting basic science through infrastructure development, and demonstrate the ways in which the design of the infrastructure, the constitution of its communities, and knowledge about the state of the art in Earth Systems Science was affected by and responsive to the concerns of the NSF as mediated by its program officers, as well as the knowledge produced about Earth Systems Science as part of the design process.

The EarthCube Journey

At the core of the motivation for EarthCube was a policy goal – enabling a better understanding of planet-level dynamics that can only be properly traced through interdisciplinary cooperation. Following a round of community input in the forms of solicited white papers, webinars, and design charrettes the NSF funded a series of end-user workshops, as well as a number of grants for building blocks, research collaboration networks (RCNs) and conceptual design awards. Through these building blocks, EarthCube took on a middle-out approach to the development of their infrastructure, with individual components seeking to provide a particular solution to a specific problem rather than initially organizing a top-down architecture. EarthCube, then, was planned as a project simultaneously in community design and NSF-funded leadership with the goal of engendering long-term enrollment and engagement on the part of participating researchers – in part evidenced by the stated intent of the NSF to fund the project through the year 2022. (Figure 3)

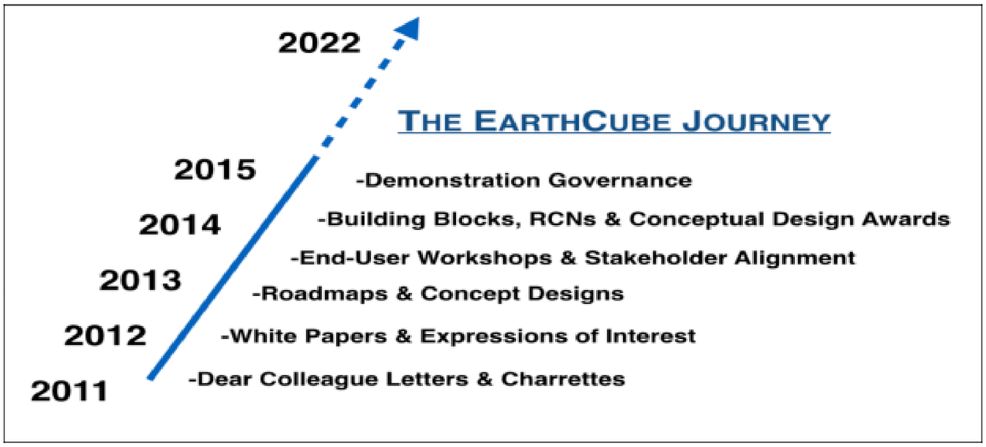


Figure 3: EarthCube Implementation Plan (EarthCube: Past, Present and Future, 2014)

After the initial building blocks were proposed and funded there also was funded a demonstration governance (interchangeably called test governance) committee whose goal was to create a transitionary governance structure towards community-led organization and leadership of the overall infrastructure. The test governance group collected both opinion and demographic data describing the initial makeup of EarthCube and led the PIs and the community at large through design activities and other work intended to promote engagement and develop a shared vision of the final structure of the infrastructure. This was largely seen as somewhat revolutionary – similar projects had no specific leadership or were designed from the top down by computer scientists with little domain engagement. (Ribes and Finholt, 2009; Jackson and Buyuktur, 2014; Bowker and Ribes, 2008; Baker et al., 2005) EarthCube, then, was not simply an exercise in tool development and research support, it was from its beginning an attempt at community-building through tool development. New collaboration and the potential for transformative science had been the mantra throughout the early days of EarthCube. The test governance group, by design, paved the way for a community-led governance structure, and took on the smaller role of administrative support as the EarthCube office.

The building block grants were a varied group of software development projects focused on providing particular solutions to problems as identified in the series of end-user workshops and design charrettes with input from both domain scientists and computer scientists. Building blocks were not selected or funded with a particular architecture in mind, but rather on their potential to provide worthwhile solutions to particular problems. This led, in some cases, to a sort of 'coopetition'- a portmanteau of cooperation and competition - among funded projects. Many building blocks sought to solve very similar problems identified in the end-user workshops, 'data discovery, access, and description' being one that was oft-repeated, through various non-complementary means.

This occasional overlap in terms of how projects might engage with their end-users and stakeholders was complicated by a general consensus that those building blocks that developed collaborations with other building blocks were more likely to receive additional funding later in the project. In addition to this, a series of architecture awards would eventually be funded with the goal of incorporating building block resources and completed projects into a single infrastructure. From the earliest days of a visible EarthCube community, the building blocks were simultaneously in competition with similar building blocks in terms of being incorporated into the final architecture while operating under a mandate for cooperation and being almost totally independent from each other. This occasionally led to some contentious meetings as project goals themselves needed to be negotiated and evaluated against potential collaborations (I explore this further in Chapter 6 below) – PIs were beholden to their original project descriptions, but their future collaborations in general to some extent required that their work deviate from their original plan. While PIs were in contact with each other from the drafting stage (as most of the eventual PIs took part in the end-user workshops), the building blocks themselves were written in relative vacuums. While there was an awareness that collaboration would be important in the future, it is still quite difficult to write a well-bounded and effectively managed proposal that takes into account the potential for work outside the scope of the proposal itself.

After what might be termed a phase of preparatory exercises (solicitation of letters, white papers, preparing end-user workshops), EarthCube as a visible project was centered around the work of the test governance committee. Developing structures of governance, management, and nascent collaboration characterized the early years of EarthCube as a whole, despite an apparent general discomfort with the notion of working on policy in the broader sense. While few would call themselves leaders or governing members of the community it was clear that each individual bore a commitment to certain modes of government and policy alignment that worked to inform the final governance structure, future mode of architecture development, and design priorities of the infrastructure itself. Policy impinges much more extremely at the moment of infrastructure: targeted financial investment designed to expand capacity is a statement on the relevance of certain science and entrenches certain methods and techniques while marginalizing and rendering discursively irrelevant others. (Knobel, 2010) There is a growing motivation among scientists in fields of new political relevance (Earth Systems Science, or ESS, primary among them) to be able to present consensus on certain issues that are seen as formative of the identity of the discipline. While the social and intellectual work of science thrives on differing levels of agreement there is a growing trend away from acknowledging that process and towards a unary, positivist presentation of science.

However, the building of new infrastructure – due to large-scale investment and institution-spanning development – is a result of policy action operating on an established co-construction of scientific relevance. At this basic level there is a tension in the design process between the desire to present a positivist, objective and apolitical narrative of the scientific process with the sociopolitical outcomes expected from the investment, and the knowledge produced within this tension have a tendency to be represented according to the expected direction of their travel. Insiders to the process of scientific knowledge production conceive of a necessity of translating their work towards a certain political goal – be it funding allocation, changes in public conception, or the actual implementation of recommended policies. This translation often takes the form of decontextualizing knowledge products from the conditions of their creation with the goal of making them more mobile, but has the effect of making that product less legitimate to insiders. The boundary between science and policy in this context, then, is the boundary of advisors and policy-makers – the assumption of the necessity of translation is the science-policy interface, rather than the basic epistemologies or ontologies relevant to knowledge production. This thesis is concerned with that which comes before, not in a historical sense, but in an emergent, interconnected hierarchy of work with loose, porous and shifting boundaries. As I introduce this topic, I will first discuss the concept of co-creation and, in particular, the formation of the concept of relevance as a bridge driving both scientific knowledge work and policy formation. My central thesis follows the interpenetration of knowledge-producing activities in infrastructure design, and works to trace how knowledge produced for and by policy actors shapes, negotiates with, and incorporates knowledge produced by scientific actors.

Knowledge bears with it the context of its creation, the epistemologies that inform its construction, and the sociotechnical systems that support its inscription. (Latour, 1974) Accounting for the production of scientific knowledge by necessity is also an account of the relevance of that particular bit of knowledge and the systems that support its storage, description, and sharing. And while knowledge carries with it the circumstances of its creation, the object of knowledge is also subject to re-contextualization and shifts in epistemology and ontology. Following the movement of a knowledge through shifting networks comprised of groups of actors who lack epistemological or ontological agreement in a significant way provides the basis for understanding boundary-spanning knowledge infrastructure. These sociotechnical systems that support the movement of knowledge across apparently incompatible contexts also provide the basis for (with some level of loss) translation of knowledge to differing epistemological and ontological contexts. An improved understanding of the translational processes of such boundary infrastructure eases the construction of relevance, aligns policy-level actions with effective scientific knowledge production, and unifies discourses on scientific relevance and social goals. Infrastructure, and particularly concerted infrastructural efforts, provide a basis for examining what may otherwise appear to be disconnected events as part of the same reality. As infrastructure resolves the tension of local and global, it can be seen both as a bridge and enabler of action at a variety of scales. (Star and Ruhleder, 1996) Examining infrastructure directly provides a narrative of change that is interested first and foremost with the invisible, with the marginalized, with that which silently contributes to differing conceptions of the possible, of the probable, and proposes a boundary circumscribing otherwise indistinct groups, regimes and organizations. Infrastructure is a relational proposition – it describes the relationship between certain activities, systems, and modes of work. When something is in the position of infrastructure, it does not need to be reconsidered at the point of action. To take a shower, for example, is to act through a variety of infrastructures - the most obvious of which is that infrastructure that moves clean water to the shower itself – none of which need be though about in order for the activity to take place. I do not need (unless I choose to) to lay pipe for water transmission, to establish standards for what constitutes clean water, to build filtration and treatment, or to provide a place for drainage in order to take a shower. It is in my relationship to these systems, however, that they are infrastructural – plumbing is not necessarily infrastructural to the plumber who works on it, to the sanitation and treatment workers, to those maintaining the reservoirs and pipelines – but in enabling my activity (showering) without needing to be reconsidered these systems are infrastructure from my perspective, at that time. This is not only a function of perception (though that provides an easy entrée to the relationship), but also of action and activity. Infrastructure is fundamentally active and temporal in this sense. In a vacuum, without the filter of activity, infrastructure is indistinguishable from other artifacts. So to infrastructure (as a verb) is to act in a certain way, to operate on a certain set of assumptions, and fundamentally to make an ontological statement. Plumbing as an infrastructure is different from the pipes that make it up, but ontologically as I take my shower they are one and the same – the pieces that comprise the system are subsumed into the object of the infrastructure itself. I don't talk about chunks of pavement, electric lighting, and painted symbols, I talk about roadways and enroll the entirety of those things that go into a roadway as a single, infrastructural object. This is an important consideration for progressing through this document, particularly as my discussion moves from large, relatively visible infrastructures to those infrastructures of governance, expertise and training, community, standards, and systems that are less immediately visible.

While the stated goal of EarthCube was a community-led architecture, the fact remains that the mechanism of individual architectural components (and conceptual design awards) was through standard NSF grant review. So while a community was formed around the concept of a geoscientific infrastructure, it was not exactly an open community. We can imagine, then, some logic to the selection of initial, second-round and architectural building blocks as attempts: not to select specific projects and individuals for participation in the community per se, but to fill in the gaps of a larger picture of knowledge infrastructure. I would call on a metaphor of creating a flower arrangement here – the arranger selects the best flowers according to their vision of the future arrangement. It is not a prescriptive task, but one that seeks to find beauty in the objects available – in similar way there was an intelligence and guiding vision to the selection of building blocks. Not a vision that defined the final architecture, but one that sought to arrange the available (both funded and unfunded) building blocks towards a vision of infrastructure. There existed from the beginning a conceptual vision of infrastructure held by those in a position to make decisions running in parallel to the various imaginaries of infrastructure as expressed by those participating in its development (which included attempts to understand exactly the rationale and goals of those in a decision-making position). Throughout the project there were consistent, negotiated design imaginaries where those making decisions and those who were participating in the project were working to understand each other and negotiate between their visions for the future of the infrastructure.

As my research is concerned with change and, to some extent, with the question of how systems, policies, practices, etc. become infrastructure, I am less interested in mapping those relationships that are relatively stable and more interested in working through the ideological, systemic, and theoretical *pivots* around which change occurs. In particular, I am looking at knowledge production, management and sharing as a central pivot around which the messy, ill-bounded and ill-distinguished worlds of science and policy revolve. As such, we will consider a series of pivots and their broader, infrastructural effects – the formation of the concept of relevance, the act and conception of translation and transfer of knowledge, and the social investment in the concept of scientific consensus will form a broad conceptual core guiding closer investigations.

Scientific knowledge is vital to the effective public management of common resources, informs economic and banking policy, and is used as a legitimizing argument in many forms of public discourse. The importance of science to policy makers and framers as a legitimizing force, and the ways in which scientific consensus has the ability to inform and affect environmental and resource regulation, funding and taxation politicizes scientific knowledge in a variety of ways. The basic practices present in knowledge production (uncertainty, peer review, model development) become sites of contested boundary between the science and policy informed by or about that science. The practice of making and adjusting policy towards a goal relies on its own human and technological infrastructure in practice, and the act of regulation itself produces knowledge, both technological and social.

Policy and science co-create, and the boundary between policy work and scientific work is at best a shifting, porous one: a permeable membrane for the movement of individuals, epistemologies and ontologies. For many of those I observed and spoke with over my investigation of EarthCube, policy was very much other: individuals avoided the use of certain phrases and terms that bore the weight of policy, there was worry and work done on the level of engagement with policy and the burgeoning role of Earth Systems Science as advocate of the scientific way of knowing. The struggle for scientific relevance was viewed separately as internal and external. Policy and regimes of regulation were externalized to the debate as something similar to the tides, or the growing progress of climate change – policy for those scientists I was observing was something to be known, to be understood as a relatively stable (though variable process) means by which scientific work may take place. Socio-political co-construction of scientific relevance (Jasanoff, 2004) has been assumed in order to discuss the ways in which that co-construction is subverted in defining group membership, providing space for competing discourses, and supporting marginal work. What may be termed as the world of policy: external, enabling and adversarial in turns to various scientific work takes on the form of a regulatory, personal and technological infrastructure of its own.

Perspective: Meso-level Policy Implementation and Policy Tractability

To some extent what is at stake in my analysis here is the issue of policy tractability, which I define as a way of accounting for the availability of an object to policy, the capacity of organizational or governmental policy to address that object, and the production of some understanding of that object such that the above can be effectively assessed. In approaching the Geosciences as a community in the EarthCube project the NSF first had to render that community in some way policy-tractable (I will discuss this process in more detail in Chapter 4) through surveys, voluntary calls for participation, and other 'scalar devices' as defined by Ribes. (2014) These techniques, as a group, occupy a space in between the macro level of policy implementation, which sets large-scale agendas and provides resources, and micro level implementation, which is the on-the-ground work necessary to enact those agendas with the resources provided. (Berman, 1978) Through both my primary and historic sites I saw some level of policy-related work that enabled the translation of broad social and political agendas into specific policies, regulations, funding calls, and other micro implementations. I have called this the meso-implementation of policy, as it does not define or organize on its own, but instead provides a means by which definition, organization and strategizing might take place.

Meso-level policy implementation is not just a mediation between broad agendas and on-the-ground implementation, though that is a part of it. I characterize it as a means of understanding the often-marginalized mechanisms by which policy is implemented through the use of knowledge-producing activities, rather than focusing solely on the observable dynamics of its implementation. Not only that, meso-level policy implementation works to coordinate, constitute, and organize particular communities around issues, motivations, methods, and topics of study. It is fairly reasonable to say that the community of Geoscientists as accounted for in the funding and policy around EarthCube does not constitute the entire population of individuals that might be part of that community. Geosciences as constituted within the funding regime of EarthCube is a community in its own right, bounded and defined by the mechanisms of meso-level policy implementation and knowledge production. This is not a one-way process where policymakers and funders dictate a group to the scientists themselves, but rather a coordinated activity where calls for participation, open workshops, solicitations of funding proposals, and dear colleague letters allowed the community to constitute itself in light of and in response to the image of that community initially presented and approached by the NSF in early stages of planning.

Meso-level implementation accounts for substantive work that supports and enables science funding but is rarely effectively represented in accounts of said funding. While among the people I work with there is a move towards better accounting for the constitutive work that precedes the funding of basic science there is still a bit of a gap in which really interesting design decisions are being made. The current vision of EarthCube was formed in its initial workshops and letters, refined and modified through and ongoing process of NSF engagement throughout the project's life. Staggered funding allowed for project-level adjustment and modification of the community in a very top-down way, but was not solely determinative. At the edges of funding are negotiations between those who form strategies and agendas within funding agencies and specific representatives or power players in the sciences that they approach. The dynamic is reaching out and reaching back - the NSF presents an area of interest formed in part through interactions between its officers and individuals with an interest in forming science policy agendas, then those interested in doing research work reach back. It is less a meeting in the middle than a very visible form of a mechanism by which the co-production of relevance and scientific knowledge as discussed by Jasanoff (2004) and others in enacted. At the meso-level agendas are configured and re-configured, user communities constituted and understood, knowledge produced about the state of the art in science as well as its gaps and a concept of innovative science is produced. I draw on this perspective heavily moving forward, and employ the notion that there exists a set of relatively stable mechanisms for rendering a community, domain, or other object tractable to some form of regulation, legislation or other policy action as central to my analytic perspective throughout the remainder of this document.

Moving Forward: EarthCube, BCube and Policy

While EarthCube and the broader landscape cyberinfrastructural development is the site of the research I am presenting here, my interest revolves much more closely around the nature of its relationship to policy. While I do not immediately, as above, accept either a clean interface between scientific knowledge production and policy, or the concept that this relationship can be fully explained by co-production, I did see a variety of interesting relationships, negotiations, and dynamics between the broad domains of science and policy. In fact, I tended to find policy and science in general working within the same general spaces. Objects of interest in science, at least among those I observed and read about, tended also to be objects of interest to policy, in one form or another. The production of relevance, then, is one of the primary areas where I see science and policy directly working alongside and from each other.

In a similar way, while observing the work of funded infrastructure development I discovered parallels between the knowledge production activities undertaken by scientists and those undertaken by policymakers and agency funders. In this, a fundamental part of the science funding process emerged – the work of both policy agents and particular scientists to constitute, describe, and account for a particular set of communities of research. Policymakers, through mechanisms like calls for proposals, dear colleague letters, the solicitation of white papers, and local engagements like workshops and design charrettes, were consistently engaged in knowledge-producing activities. While supporting work in particular domains, there is an evident effort on the part of policy professionals to understand the concerns of a particular domain of science, their technologies, methods, and outcomes, and to gain an understanding of the state of the art of that domain. In similar fashion, through self-selection and the submission of work, opinion and scholarship, scientists seeking funding – this is particularly evident in infrastructural projects – constitute themselves as a community and work together to represent particular pieces of knowledge as representative of that community.

My discussion of EarthCube as an infrastructure project begins with a description of exactly how policy engages in the work of basic science through infrastructure, and how knowledge produced in science becomes incorporated and acknowledged in policy outcomes. The Chesapeake Bay Watershed is a particularly revelatory case of how this happens. Fishing of oysters in the Chesapeake took place before it was well-understood how those oysters propagated and what conditions were required for them to thrive. It was through commissioned reports and the engagement of biologists in the area that knowledge of how to effectively govern the watershed towards the goal of growing oyster populations came about. In seeking to rehabilitate and support the livelihood of oyster fishers as the oyster beds slowly decreased in yield, policymakers produced knowledge not only about the state of science and biology relative to the study of oysters, but also produced a proto-infrastructure through which the knowledge produced by this science could be reflected and incorporated into policy. While the recommendations of scientists were not fully and immediately adopted into law, there is evidence from later commissions and ongoing regulation that the knowledge produced even in the earliest commissions became acknowledged in law. In similar fashion, the object of the Total Maximum Daily Load of a watershed presents not only an object of study and ongoing research for scientists in fields of hydrology and conservation, it is also a leveraging tool for policy that takes into account the current state of the art in technology and scientific knowledge. The modern TMDL is model-derived, produced from heterogeneous regimes of remote and local observation, experimentation, and data science. It is also the means by which the level of health of a particular watershed is gauged, and a tool for prioritizing conservation efforts. The TMDL is infrastructural to conservation policy, hydrologic work, and represents a significant convergence of policy knowledge construction about the state of science as well as a site of scientific knowledge production in its own right. Policymakers are fundamentally interested in the work of assessing, understanding and describing particular communities, their interests, epistemologies, and techniques in order to fund their work towards particular policy goals.

This fundamental interest was reflected in the early days of EarthCube's design and implementation. Significant work was done not only to assess the state of the art and the technological and data needs of the community of science addressed, but also to constitute a general community of geoscientists with the potential for cross-domain collaboration. Through end-user workshops and design charrettes, policymakers seeking to fund the development of a cyberinfrastructure for the geosciences produced a significant body of knowledge about the community of geoscience researchers, their resources, data sources, methods and techniques in producing knowledge for their field. Throughout the life of EarthCube the NSF remained engaged, modifying and adjusting the constitution of their community by providing new funding, building, assessing, and coordinating tool development, and drawing in a broad community of interested researchers and technologists.

Finally, the work done in the BCube building block was characterized by the negotiation not only of the infrastructure project as a whole but of the whole stack of existing infrastructures from which scientists were already working. In the NSF's attempts to fund new infrastructure, change to existing infrastructure, work practice, and social organization was necessary. New infrastructure does not slot in easily to existing infrastructure in most cases – there is a period of adjustment and significant work before a new system, technology, or organization can be infrastructural to a particular field of scientific inquiry. This process of becoming infrastructure reveals communities of researchers, complex interrelations of technologies, techniques, methods and resources leveraged towards knowledge production, and the difficulties of negotiating and implementing change to the extant infrastructural stack.

Throughout this work I expose the feedback loops, iterations, and knowledge-producing activities undertaken between the domains of science and policy that enable new work, support the movement of knowledge from the realm of science to policy and vice versa, and work to produce the concept of relevant or timely science. I work to account for the role of policy in both guiding and understanding science, and provide an account of infrastructure development from the perspective of knowledge production occurring from policy about science, from science about policy, and from both of these domains about the world at large. There is an increasing recognition that scientific knowledge production does not take place in a vacuum, but rather is influenced and affected by particular problems, the interests of industry, and policy goals. Not just at a selective or programmatic level, I show how scientific work orients itself towards the needs of policy, and the ways in which scientific knowledge production defines, expands, and fundamentally enables the action of policy at a variety of scales. With a better understanding of how knowledge production in policy constitutes, defines, and describes scientific domains, comes more effective ways to direct funding efforts, to conceive of the relationship between NSF programmatic work and the scientific work it supports, and how to negotiate the landscape of established, change-resistant infrastructure to do so.

CHAPTER 2: REVIEW OF RELEVANT LITERATURE

Infrastructure subtends our lives in so many ways. Without built infrastructure, such as roads and railways, we can neither reach work nor the seaside. Without an electricity grid or gas and oil pipelines, we would have the same problem. Without an information infrastructure (the Internet, mobile communications), we would have little to do once we got there. However, in traditional sociological and historical narrative, all this infrastructure is invisible – it fades into the background so that the *real* story can be told - the development of a social movement, the toppling of a power structure, the discovery of the standard model of particle physics. However, infrastructure is a concept that is very difficult to define. At first blush, it would seem that the highway system has very little in common with the Worldwide LHC Computing Grid. In what sense, then, is it useful to assert roads, pipes, electrical networks, the Internet and cyberinfrastructures fall under a single rubric? It is one of the goals of infrastructure studies – and thus this thesis - to address infrastructure as infrastructure despite the material heterogeneity of its physical appearances.

A central methodological move for infrastructure in STS is to flip the attention from the spectacle of the pageant of history to look at the role of infrastructure in its formation – a move that has been called infrastructural inversion. (Bowker, 1994) The core reason for making this move is the postulate that infrastructure matters. It is not just that the infrastructure is a neutral background which enables an infinite set of activities; rather infrastructure holds values, permits certain kinds of human and non-human relations while blocking others, shapes the very ways in which we think about the world. This is evident in

Veyne's (2013) proposition that it is impossible to trace the development of the concept of democracy over time because 'democracy' changes fundamentally with new infrastructural developments. Meeting in an agora or town square to determine matters of concern is fundamentally different from holding discussions through print media or following a 24-hour news cycle on electronic media. Or as Richard John (1995) has pointed out, we could not have the American state without the cheap circulation of newspapers nationally through the infrastructure of the Post Office, permitting the engagement of a national debate among residents of the otherwise remote States. Infrastructural inversion has been applied to work in health care (Jensen, 2008), water management in Thailand and the role of rice production (Morita, 2017), sociotechnical analyses of Wi-Fi, (Mackenzie, 2005) and in studies of policy and development. (Pelizza, 2016; Suarez-Villa, 1997; Korn and Voida, 2015; Hetherington and Campbell, 2014) Infrastructural inversion, however, seems most comfortably applied to the area of science studies for its revelations on knowledge production practices. (Mayernik et al., 2016; Georgiadou, Harvey and Miscione, 2009; Lee, Dourish and Mark, 2006)

Star and Ruhleder's (1996) well-known heterogeneous list stands in place of an essentialist description of infrastructure. I work from the definition that an infrastructural relationship is one between goals (getting to work, being able to see in order to read, sharing data in order to produce transformative science) and means (roads, the electricity network, cyberinfrastructure). Throughout much historical and sociological work, infrastructures have faded into the background – they have not been seen as bearers of political positions, moral values or cultural concerns. The key insight that science and technology studies has propagated is that this is just wrong: these all get woven into infrastructures in such a way as to affect what can be done in and understood about the world. It is this insight which permits us to travel freely between all kinds of infrastructure in order to uncover the ways in which this entanglement occurs.

One of the reasons for the traveling nature of infrastructure theory is the notion that there is no inner essence which makes a given thing an infrastructure. As Star and Ruhleder (1996) have argued so eloquently, one person's invisible infrastructure is another person's job, to be faced materially and directly every day. Infrastructure, as they argue, is inherently relational – a given system, technology or organization is *infrastructural to* a particular activity at a particular time. Infrastructure does not exist in a vacuum: much like Engestrom's (1990) argument about tools, the question is not whether or not a given thing *is in essence* an infrastructure, but to ask *when* it is an infrastructure. There is no system that is inherently infrastructural, there are only observed infrastructural relationships.

Materialities of Infrastructure

While we can find evidence of infrastructural relationships throughout history, it is a contentious task to pinpoint originating moments that lead to current conceptions of infrastructure. However, we can pinpoint a few moments in the study of built infrastructure that presage the STS study of infrastructure. Writing in the nineteenth century, the golden age of modern historiography, Karl Marx and Frederick Engels made a distinction between base and superstructure, (1970) which can be read as an early infrastructural turn. The base – that which underpinned all of social action – was the economic mode of production, be this slavery, feudalism, bourgeois or revolutionary. The superstructure comprised the content that was produced – culture, ideology, social institutions and so forth. In a telling phrase, he argued that human history was determined *in the last instance* by the base. (1970) One of the testaments to his influence is Lewis Mumford and George Copeland's magisterial *City in History* (1961), which drew attention to the role of civic infrastructure (sewers, roads, aquifers) in the development of urban life. Similar work in accounting for civic infrastructure in governance has been done by Graham and Marvin (2001) in describing how technological mobility produces the urban condition.

A clearer starting point for infrastructure within science studies is Thomas Hughes' *Networks of Power* (1993), which was a study of the development of electricity networks in the late nineteenth and early twentieth centuries in Germany, Great Britain and the United States. Hughes, coming from engineering and systems thinking, laid the basis for a general history of physical infrastructure, noting that there was a continuity in personnel and techniques between the canal builders of the late eighteenth century, then the architects of railway, telephone and electrical networks – it is simple (by way, for example, of Shannon and Weaver (1949)) to extend this to the rise of the Internet. The core insight which Hughes developed was that of the reverse salient. A reverse salient can be defined as any obstruction which prevents an infrastructural system from being developed. This might be a technical matter (could you find the right filament to make incandescent lights which would last) or a social one (fear of direct current because it was tied to the electric chair, so it was seen as dangerous to domestic hearths). From a systems perspective – and there is a strong cybernetic (Wiener, 1988; Wiener, 1961) element here – it made no difference where the reverse salients were occurring: they had to be addressed for the infrastructure itself to develop. This systems view found ready acceptance in Bruno Latour's *Science in Action.* (1987) It is arguable that the translation was facilitated by the origin of actor-network theory itself in cybernetics, however a more standard explanation is the semiotic one. Latour wanted no distinction between human and non-human actants; Hughes provided a framework which flattened problems with electrons and people onto a single analytic base. The infrastructure itself was not just technology: it was always already braided with social, cultural and political values. This is core to my analysis of EarthCube.

Latour also wrote an account of other infrastructural projects – *Aramis,* (1996) for example, explored the slow death of a personal rapid transit system in France. In *Aramis*, Latour introduces the concept of scientifiction, or a narrative account coupled with analysis in a sort of "hybrid genre." (pg. ix) In the central mystery of this tale, *Aramis* was killed, not by any particular actor, but by the broad failure of multiple groups to do the work sustaining and negotiating that technology. A similar work by Steve Jackson and Ayse Buyuktur, "Who killed WATERS," similarly accounts for the 'death' of a project by exploring the sociotechnical and political factors that did not effectively support it. (2014) We shall explore the slow death of EarchCube in this thesis.

Hughes' work can be seen as somewhat infrastructural to infrastructure studies; it can also be seen as part of the sociotechnical systems literature. STS was first developed through the work of the Tavistock Institute – initially through an analysis of coal mining work in England. (Trist and Bamforth, 1951) While this work did not have a large direct influence on science studies (Kaghan and Bowker, 2001), it did follow a parallel track, leading to some interesting entanglement between the two over the decades. Thus Rose's classic *Governing the Soul*, both tied to the radical psychoanalytic tradition of Tavistock and brought Foucault's analysis of power (1990) into play – the latter itself becoming variably central to science studies and infrastructure. A more direct filiation of Hughes' work is the rich tradition of large scale sociotechnical systems which drew heavily on the systems literature and on STS. A third, more recent tradition on sociotechnical has been pioneered by Steve Sawyer (2001) and Wayne Lutters (2000) among others that works through issues of valuation, prioritization and the market for funding infrastructural work.

Alongside Hughes' work as foundational to infrastructure studies must figure Langdon Winner's influential paper 'Do Artefacts Have Politics?" (1980) The argument most picked up from this work deployed the case of New York city planner Robert Moses to demonstrate that built infrastructure was itself politically active: if you build bridges that prevent buses from passing under expressways, then you can exclude the great unwashed from public beaches in ritzy neighborhoods. Although the case itself has been comprehensively disproven (Joerges, 1999; Woolgar, 1993), its message has continued to resonate; as indicated by Latour's dictum, adapting Clausewitz, that technology is politics by other means. (1988; cf Strum and Latour, 1987) Callon's article on two plans for electric cars in France being bets on two incompatible theories from political sociology is canonical here. (1986)

Responding to Changing Technology: Information Infrastructures

While early science studies concentrated in its separate ways on built infrastructure, a new concept was about to enter the lexicon to describe a fundamental shift in social and political organization: that of 'information infrastructure' – we now need to use the retronym of 'built infrastructure' to designate what 'infrastructure' used to mean, (cf Lakoff, 1993) much as we use the retronym 'snail mail' to designate the postal service. Of course, information infrastructures are built, too, so that much of the earlier work could be ported into the new space. However, the new formations did cause a change in analytic and methodological focus.

Where what is now called information infrastructure was once the domain of the library scientist (managing of books, questions of information retrieval) and the organizational theorist (file structures and so forth), but is now moving to the world of the technologist and the service provider. (Rubin, 1998) The nature of scholarly production and the profession of librarian has changed greatly over the past thirty years. The very infrastructure that had grown throughout the nineteenth and twentieth centuries (file folders, typewriters, carbon paper, hanging files, Xerox machines) has been largely displaced. (Yates, 1989) In step, the dominance of computer science across multiple disciplines has increased (echoing the earlier dominance of statistics with the development of governmentality (Foucault, 1991)).

The work of economist Paul David provides a double transition here. First, David argued classically that personal computers were like electric dynamos, in the sense that when both were introduced into the workplace there was a cultural lag between innovation and usability. One needed not just the infrastructural change (from gas to electric, from typewriters to pc's), one needed to develop the associated infrastructural imaginary. In both cases, as he pointed out, there was a 'productivity gap' of about 20 years before the new technology started to make a fundamental change. Second, David developed the theory of network externalities. (1990) If I am a member of a phone network with two others, it costs me nothing if three more users are added, but it affords me value. Where Hughes had argued the importance of load-bearing for the economics of built infrastructure, David argued the importance of network effects for the new economy. We will see this slow development of cyberinfrastructure despite the urgent rhetoric of EarthCube.

In many ways, the emergent focus on information infrastructures over built infrastructure was a recognition of the presence, importance and societal effects of the already-existing information infrastructures. Wartime projects, in addition to being secret, were often geographically distributed, and required significant coordination. As digital technologies were developed, they began to be seen as a solution to the problem of coordinating distributed work. These moves towards the digital, starting with ARPANET, as solutions to information management problems, are among the first instances of what came to be called Cyberinfrastructure. (Atkins, 2003) As the potential revolutionary nature of computerization and the internet began to be recognized there was significant effort towards using it in support of different kinds of science. In the 1980s we see the development and funding of collaboratories and digital libraries as infrastructural efforts attempting to capitalize on new technology.

Collaboratories were initially conceived as a "center without walls, in which the nation's researchers could perform their research without regard to physical location, interacting with colleagues, accessing instrumentation, sharing data and computational resources, [and] accessing information in digital libraries." (Wulf, 1989) The idea behind the collaboratory was to take what had previously been a locally isolated collaboration and allow for coordinated work across large distances (almost an identical story, sans venue, to that of ARPANET and the internet itself). The work done in collaboratories showed that, when they functioned, they were not just, "an elaborate collection of information and communications technologies," they were "a new networked organizational form that also include[d] social processes; collaboration techniques; formal and informal communication; and agreement on norms, principles, values, and rules." (Cogburn, 2003) Collaboraties were touted as an innovation inevitable part of the future of scientific collaboration (Kouzes, Myers and Wulf, 1996) and argued to represent new forms of scientific organization (Finholt and Olson, 1997; Finholt, 2003; Bos et al. 2007) Collaboratories were implemented in areas like microscopy, (Agarwal, Sachs and Johnston, 1998) higher education, (Sonnenwald and Li, 2003) biomedical research, (Schleyer et al., 2005; Craver and Gold, 2002) and history. (De Moor et al., 2008)

Roughly concurrent, though slightly following from, work on collaboratories was the concept of online information infrastructure for the sharing of documents and resources called digital libraries. (Borgman, 1999) Digital libraries were oriented much more closely to the documents they stored than to the collaborations they enabled, with research in digital library systems being significantly engaged in knowledge organization like citation management, (Lawrence, Giles and Bollacker, 1999) metadata and document importing, (Witten et al., 2002) the production of formal models for their development, (Schatz and Chen, 1996) and issues of how to scale from a traditional library service model to a digitally engaged one. (Schatz, 1997, Lynch and Garcia-Molina, 1996; Bates, 1998; Hodge, 2000) Work in digital libraries followed very closely from the concerns of traditional libraries, (Lesk, 1997) focusing on issues of document storage, retrieval and user engagement, (Hong, Tong and Wong, 2002) and even taking on issues related to preservation. (Hedstrom, 1997)

It was in the era of collaboratories that Star and Ruhleder published a seminal work in infrastructure studies, one that recognized and formalized many of the preceding concepts of infrastructure as well as heavily influencing future methodological work in infrastructure studies. “Steps Towards an Ecology of Infrastructure” (1996) was an ethnographic study of a group of biologists working with the Worm Community System (WCS), a digitized library of *C. Elegens* flatworm specimens and technologically-mediated path for collaboration among the biologists working with them. In many ways the WCS was an ideal site for the implementation of new computing infrastructure: there was already a social expectation of collaboration and well-established network of biologists sharing specimens (one of the traits making *C. Elegens* an excellent laboratory subject was its amenability to being transported by post). However, by most metrics the WCS was a failed project. It is in their accounting for that failure that Star and Ruhleder propose infrastructural issues as major factors influencing that outcome.

Infrastructure, as defined by Star and Ruhleder (1994), has a variety of different dimensions. It is embedded and transparent; infrastructure exists (metaphorically) within or underneath other social, technological and built worlds and does not need to be reconsidered at the moment of each task it enables. Infrastructure is learned as a part of membership in a given community and linked with the conventions of practice therein, and embodies some set of standards. It is built above an installed base, becomes visible upon breakdown, and is of a scale or scope that exceeds a single 'site' – however that might be conceived. The heterogeneous nature of this list was quite deliberate: it was less Borges' Chinese encyclopedia cited by Foucault (2002) than a recognition that infrastructure was integrally a social, organizational and physical phenomenon. This insight is core to my analysis.

Star and Ruhleder were not the only researchers thinking infrastructurally. The shift from collaboratories and digital libraries to cyberinfrastructure (CI) or e-science is characterized by the uptake of concepts and methods from science studies. Latour and Woolgar's work in *Laboratory Life* (1979) had already drawn attention to the substrate of intellectual production – the document as 'immutable mobile' which could circulate along scientific networks. We shall explore a particular version of this: the 'umbrella term' which is designed to move between science and policy. Star and Ruhleder's research helped to point out something that was becoming increasingly evident: infrastructure to support large-scale distributed scientific collaboration responds to factors other than the availability and ease of use of new communicative and sharing resources. That the WCS was underused was substantively less interesting than why and how it was.

As the collaboratory and the digital library gave way, in the wake of NSF programs, to the concept of cyberinfrastructure in the United States, there were two corresponding intellectual moves. The uptake agreement on method and a broadening of the theoretical space of infrastructure into issues of policy, temporality, design and values. A tale of CI can be one of confluence and a particular methodological turn. Scholars interested in studying the sociology of knowledge production, in the presence of a particular funding regime, (Gibbons et. al 1994) at a time where scientific work was becoming increasingly interdisciplinary and recognizing new, emergent roles in knowledge production like developing simulations and computer models of observed phenomena, (Galison, 1996) with the development of new digital communication technology adopted ethnomethodology as a technique for studying emergent scientific networks.

In an influential report for the National Science Foundation, Edwards et al. (2007) argued that one could learn a lot about CI from studying the lessons of built infrastructure. Thus, ‘path dependence’ is a key concept for the development of transportation or information networks – it is also crucial for CI. Similarly, the need to study standards (as brought out by Egyedi (2014)) works across both domains. Busch’s (2013) volume on standards develop these issues further, as does Lampland and Star’s collection. (2009)

From Infrastructure to Knowledge infrastructure.

Edwards et al. (2013) have argued that we need to move beyond the concept of information infrastructure to that of knowledge infrastructure in order to explore the ways in which knowledge work is changing in the twenty-first century. Studies of infrastructure at work seek to unearth marginalized work, recognize the invisibles and occlusions as they influence action at a variety of levels and propose ways in which these marginalia might be accounted for in future work. Knowledge infrastructures expand beyond instrumentation and work practice to account for the presence of political considerations, values and other invisibles that work to allow the exchange and proliferation of knowledge throughout a group. Instead of an either/or distinction between group membership and social isolation, researchers can bring to bear in their analysis a set of structural variables, such as the density of a network, how tightly it is bounded, and whether it is diversified or constricted in its size and heterogeneity, how narrowly specialized or broadly multiplex are its relationships, and how indirect connections and positions in social networks affect behavior.

Alfred Sohn-Rethel (1987) in a classic article suggests that the spatiotemporal framework of Galilean absolute space and time emerges from a reflection on the state of the economy – emergent capitalism created the commodity form, and within that form ideal was that capital and commodities should flow in a frictionless time and space. The move here is the argument that we think with and through our infrastructures. David Deutsch (1998) is one of a number who have maintained that our theories of the brain have been modeled after the height of technology in their time. Brains have been metaphorically characterized as hydraulic (grand dams being written by Freud into the fabric of our brains) to the telephone switchboard (in the 1920s) to the computer (today, with our binary neurons, firing ones or zeros into the substrate of the brain). Sadly, he also follows his forebears in his argument that this time we have got it right – the brain and the universe really *are* computers. When we take the products of our academic labor as independent entities which are removed from the complexity of socioeconomic life, we make precisely the mistake of assuming that our institutions could still/did ever exist as ivory towers. The knowledge infrastructure aligns with and meditates on the set of infrastructures within which it is embedded. This is why changing a knowledge infrastructure is about a lot more than developing new digital libraries and databases to allow us to deploy the methods that we know and trust, it is fundamentally about engaging with and understanding the social and the political. As we will see, the interaction of science and the political was core to EarthCube throughout its design and implementation.

With the rise of algorithmic reasoning through big data, it is no surprise that we get equally from the computer evangelist (Anderson, 2008) and the social theorist Latour (2007) a call for an end to theory, where theory is understood as classificatory reasoning. For Latour, the argument goes that we are moving beyond the world of Durkheim, whose sociology has reified categories such as 'society', 'gender', 'class' and 'race'. Rather than seeing these as fixed categories in the world, we can examine the variable collectivities that operate at any one instance in any one place. We only needed the categories as theoretical units when we did not have access to all the data. Anderson's argument is much the same, though dressed in different garb. He was drawing on algorithmic analysis of behavior though big data as superseding categories. Marketing firms no longer need to know what 'middle class women' want, if they have access to each of us individually. The map is indeed co-extensive with the territory. Nick Seaver (2012), studying music recommender systems, has shown in fine that his is how the algorithms which help shape our taste work. Natasha Dow-Schull (2012), through the extreme example of gambling in Las Vegas, has shown how the new form of social engineering can be precisely a Skinnerian black box – we really do not need to know what is going on 'inside' the brain of the gambler, we only need to be able to predict the behavior of the individual gambler.

There is a strong argument, then, that our deeply embedded knowledge infrastructure aligns well with our information and economic infrastructures. A corollary is that as our epochal shift in knowledge and information infrastructures is taking place across many levels simultaneously, we are not yet locked in through generative entrenchment (Wimsatt, 2001) to the perhaps troubling forms discussed thus far. There is certainly a move to develop a lockstep relationship between economy, management and knowledge. The ideology of inevitability is prosecuted most strongly at inflection points which might lead to new kinds of social, organizational and infrastructural arrangement. However, new forms of knowledge infrastructure, involving a new cognitive division of labor and new knowledge objects are proliferating in this rich space. New alignments can emerge from the complex sociotechnical spaces within which these are developed.

A new infrastructural force in our lives in the past ten years has been the development of the field of big data analytics. Dominique Boullier (2014) has explored this theoretically as the creation of a new kind of social fact. For him, while Durkheim (1884) gave us a society with reified categories (class, gender, ethnicity) together with a set of tools to uncover them, new social theory does away with the reification. (cf Latour, 2002 on Gabrielle Tarde) Crucially, so do the infrastructural tools for social action. As discussed above, when a fundamentally new information and communication infrastructure comes into being, society itself as a whole changes its very nature.

Infrastructure and Policy

As it relates to infrastructure studies, policy is something of a chimera. Even though policy concerns are presented as significant factors in the major infrastructure reports (Edwards et al., 2010, 2014) there is a conflation - or at the least a blurred distinction - between policy as law; organizational practice and rules; policy as embodied in standards, systems of classification and work practice; and policy as issues relating to governance of particular infrastructure projects. While this might make the concept of policy as related to infrastructure appear at first blush to be something of a misplaced concretism, the apparent heterogeneity of what might be termed ‘policy issues’ can be viewed as more of a product of Latour’s (among others) claim that “science is politics by other means.” (1988, 1984) For Latour, there is no simple distinction between science, values, society and political power - they all happen together, or at least at the same time. (cf. Jackson et al. 2013 and Callon and Latour, 1992) The apparent chimeric nature of specific policy topics in infrastructure studies reflects more the embodiment of values, practices and preferences in the material and organizational substrates from which infrastructure emerges.

The politic is most brazenly present in the infrastructure in the general need for financial support and upward accountability, and organizational structures combined with that political purpose has a powerful effect on design decisions and restrictions for the introduction of new infrastructural elements. (Sahay, Aanestad, and Monteiro, 2009) Increasing amounts of resources needed to expand infrastructural capacity or provide the basis for new infrastructural networks moves decision-making for large-scale research projects to ever larger groups and recognizes its impacts on more diverse social, academic, and industrial institutions. (Galison and Hevly, 1992) As such, representations of infrastructural results to policy-makers are both attended to and avoided in equal measure. Rip and Voss describe an entity called an 'umbrella term' that mediates between the work of science and the political and social understandings of that science. (Rip and Voss, 2013) Umbrella terms provide a basis for innovation by providing for a de facto understanding of emergent scientific work that does not need to be referenced back to individual research efforts. Like a snowball, these terms and numbers collect nuance and context as they move, only to be emptied in a moment as a new term gains traction, leaving only the nuance and context, the organizational realities and human connections.

It might be argued that the larger political effects of infrastructure, much like the infrastructure itself are easily relegated to the marginalia, rendered invisible by a focus on the subject discipline of a given infrastructure problem. Work done in support and not readily susceptible to a decontextualized measurement is marginalized, while still having profound effects on organization realities and local practice. (Bowker and Star, 2000)

Agreement on protocols, standards and even effective measures is a political process: the preferences, values and practices of one group are adopted and supported above others.

The concept of 'the standard' here is a strong one. A standard conceptually is able to represent a wide variety of design considerations, from physical attributes to classification systems (cf. the discussion of the International Classification of Diseases in Bowker and Star, 2000) to technical protocols, common pieces of software and file formats. While the mechanisms for how a given standard exerts power is specific to the nature of the standard itself, it remains that standards can influence infrastructural and organizational change. We shall see the central role of standards in the development of the BCube broker. Standards serve as a gateway between disparate sociotechnical systems, and their equalization of the design space is a vital component in how these systems interlink into networks (Jackson et al. 2007) and into the material world. (Busch 2013) Though standards have an "intuitive tension" with system flexibility, the establishment of standards in one area of a system tends to increase flexibility in others. (Egyedi 2014; Mulgan 1991) As standards simultaneously support the linking of systems into networks while catalyzing systemic change by stabilizing a set of design considerations, standards "achieve some small or large transformation of an existing social order." (Timmermans and Epstein 2010 p. 83)

When one builds a standard, it is built in such a way that one's own interests are promulgated. (Latour, 2007) From this perspective, adoption of a certain standard is a sort of political contest with the 'victor' achieving formalization of their own values and practices as agreed-upon or default policy. (Jasanoff 2007; Latour 1988) In other words, standards, protocols, and systems of classification expose interred socio-technical values as they make changes to regimes of decision making. (Edwards and Hecht 2010; Kranzberg 1986, Star and Lampland, 2009) Sensitive to this, DeNardis traces the negotiations and agreements that themselves were infrastructural to the interoperability characteristic of the modern internet with a special focus on the political effects of the standardization of protocols. (2009) Edwards, also, accounts for the political effects of legacy code, data regimes and the formal adoption and use of specific models in climate science. (2010) The standard can be pictured as an actor in its own right, catalyzing change and being changed through use – following the standard in use provides a basis for understanding the social effects of the network itself. (Latour, 2007)

The political, then, is not a distinct and separate feature of infrastructure. Instead, infrastructural work is fundamentally and pervasively political. Successful CI projects are those that result in the creation of a social reality for its participants where their individual work practice is shaped in relation to the attendant infrastructure. Organizational practice imports a set of values, ethics and implicit knowledge all its own. Even before work on standardization begins the negotiations and agreements necessary to allow heterogeneous groups to work together inevitably ensconces some agendas while marginalizing others. As participants and researchers are sensitized to the social effects of design decisions at the infrastructural scale a space for a more nuanced discussion of those decisions sensitive to their effects at the moment of design becomes possible – perhaps even inevitable.

Boundaries, Interfaces and Knots

The boundary between scientific knowledge work and political knowledge work is not firm nor impermeable, but rather is porous to both objects of knowledge and the individuals who create and manage that knowledge. In many cases, policy and scientific work done on the same subject uses much of the same data analyzed in the context of differing ontological and epistemological realities. As scientific outcomes are leveraged politically to legitimize certain policy outcomes there is an evaluation of the legitimacy of the knowledge itself and its relevance, just as in formal and informal structures of academic review. A basic structure linking scientific knowledge production and political action in the United States is the claim of rationality, be it scientific, legal or economic. (Brickman 1984) This reliance on the claim of rationality does not presuppose a strictly logical reasoning on the part of the individual making the claim, it is merely a means of establishing the legitimacy of a given policy decision in the nebulous public eye. In making such claims, political actors are participating in scientific discourse – that is, the discourse about the nature of science – without necessarily approaching the objects and actors of relevance to what would otherwise be considered the scientific community. Many of these methods by which knowledge is politically legitimized as a basis for action primarily make control and management possible, implying certainty where little exists. (Porter, 1996) Dissenting opinions are magnified as the importance of the rationality of a claim grows in political discourse, and uncertainty allows for the contestation of any such claim. (Jasanoff, 1994) The mobile outcomes of science (such as advice, publications, or orations) and the actors that propagate such outcomes work concurrently as scientific knowledge and political justification based on their relationship to the rest of the network. Many of the actors in regulatory policy networks, such as advisory councils and professional experts, do work that bears consequence simultaneously on political and scientific work, with the boundaries between such work formed primarily through differing relationships to and practices regarding the analysis of common sources of data.

Though policy makers and practicing scientists work from and on the same databases, their relationship to each other would seem to be one that is not closely bounded or even substantially visible. Knowledge is produced and intentionally disseminated as part of both scientific and policy work and their agendas are mutually influential. (Wynne and Irwin, 1996) Just as objects of knowledge, specific data sets and scientific outcomes move between different established conventions of evaluation and representation within scientific disciplines so to do they change moving outside of vocational knowledge-work. The management, analytic and interpretive work a scientist does on and through their recorded data is temporal and contextual, just as is the similar work of a politician. The boundary between scientific and political work can be seen to emerge more from standards of evaluation and the implicit claims of the utility of that knowledge than exist as a firm disciplinary divide. The economics of scientific funding and the development of scientific agendas at the level of national funding agencies adds further complexity to this relationship. While it is clear that there is a political reliance on the objects produced in scientific knowledge work, that knowledge work is evaluated, used, and legitimized contextually and relationally. “The credibility of regulatory science ultimately rests upon factors that have more to do with accountability in terms of democratic politics, than with the quality of science as assessed by scientific peers.” (Jasanoff, 2003) Scientific and political actors can move between these contexts in discourse, vocation and action, but as they move the context itself shifts and is reformed in that movement, bearing consequences for both scientific and political legitimacy. The extent to which science ought to influence and shape new policy is substantially contested in the United States, with scientific outcomes undergoing both deconstruction and reconstruction within political discourse. Good science, peer review, and science policy itself become leveraged towards a given conception of the basic nature of science and its best role in governance (Jasanoff, 1987) Science-policy interfaces (Van den Hove, 2007) have been proposed as a theoretical means for working through the complexity of knowledge production and political action. The idea of an interface and intersection, though, between constructs as demonstrably varied and constructed as policy and political relevance implies a directionality and separation of action that may not exist. In forming a scientific rationale for political action or a political rationale for the development of a scientific agenda scientists co-construct (Jasnaoff 2007) along with general public conceptions and policy makers the concepts of political relevance (including a set of epistemological and ontological assumptions as to the knowledge needs of policy work), useful science, and the boundaries between scientific work and policy work. The phrase 'the intersection of science and policy' assumes relatively independent vectors of action that happen at moments to interact – for example van den Hove describes science performed for the sake of curiosity (the past, pure science) and issue-driven inquiry (that science tainted by politics, the future of science driven by social reality).(Van den Hove, 2007) Rather than locating a personal motivation for scientific inquiry, I argue that we should explore not the motivation of the scientist in pursuing their work but the networks, ontological and epistemological claims that form and describe a specific agenda of inquiry. These can be traced in the observation of funding mandates, the formation of political and scientific importance of certain situations, personal motivations on the part of scientists, and other economic, politico-social and natural realities as they are constructed in political and scientific outcomes.

Both policy and science make a claim to the disciplinary relevance of certain situations by bringing their own ontological commitments to bear against that situation. A grove of trees that suddenly dies becomes a part of the political discourse on environmental management when it is suggested that acid rain (and therefore, human action) is the cause of that grove's decline. (Hajer, 1993) This is not a reconstruction of the situation as one of policy relevance, but instead a re-contextualization or reorganization, a drawing-in of the object and situation into the epistemological and ontological realities of working within policy. In much the same way, approaching a subject as a site of scientific relevance begins with the reorganization of that subject to make it amenable to a fit into extant epistemological and ontological practices. The sad state of the dying trees had to be contextualized in political discourse as both being within human control, of relevance (an oddly reflexive concept on its own) to a group, and amenable to political action. A good example of this in action is the process of mapping ecosystem services, which seeks to render the benefits humans obtain from nature amenable to valuation, prioritization and other policy activities. (Costanza et al., 2997; Maes et al., 2012)

High energy physics is characterized by its collaborative, distributed and social mode of work, and in many ways can be seen a success story among a veritable graveyard of underdeveloped CI projects. (Knorr-Cetina, 1999) The nature of high-energy physics research, with copious data produced at few sites and requiring substantial processing prior to analysis is particularly amenable to globally distributed work. However, few other fields of study have as much consensus on major questions to explore and the techniques needed to do so. In particular, the general acknowledgement of the necessity of support work in the high energy physics community provides a basis for organizational support that might not be possible in other areas. In the development of the GEON (Global Earth Observation Network) infrastructure it was noted that support work was often not rewarded by the home organization of many of the participating scholars, that there was little scholarly recognition for infrastructural work, and as such participation in the maintenance and development of that infrastructure waned relatively quickly. (Bowker and Ribes, 2008)

In the currently dominant project-focus of infrastructure research, issues of contribution, reward and participation become paramount to the account of a given infrastructural effort, and organizational issues tend to overshadow all others in describing a narrative of success, partial success, and failure. Policy is already present in the infrastructure, given the common need for financial support, and organizational structures combined with that political purpose has a powerful effect on design decisions and restrictions for the introduction of new infrastructural elements. (Sahay, Aanestad, and Monteiro, 2009) Policy, design considerations and technology are ‘knotted together’ in a way that allows a given account of the infrastructure to effectively occlude the others. (Jackson, et al. 2014) Organizational factors dominate the discussion, and, as with much reification of organizational attitudes, have a tendency to leave such projects deeply resistant to change. "There is a tension between the technical ideal of a stable, interoperable infrastructure for data sharing and reuse, and the reality of knowledge as evolving, socially and locally constructed, and often disputed." (Ure, et al. 2009) Infrastructural work can be seen operating along two cycles: those of amalgamation and fragmentation, where first work is done to draw together disparate localities of practice, metadata, reward, etc. through standardization, translation and tool development and the second is the adaptation of the outcomes of that drawing-together to suit local condition. (Hepsø, V., Monteiro, E., & Rolland, K. H., 2009)

Policy Studies, Technology, and Funded Science

In addition to how policy engages with infrastructure, there exists a number of modes though which science engages policy without direct reference to it. Sheila Jasanoff has argued that scientific knowledge is co-produced with policy concerns. (2004) She traces the movements of scientists into advisory roles as a means of policy formation (Jasanoff, 2009) and looks to the role of citizen participation. (Jasanoff, 2003) Jasanoff's approach to policy moves knowledge production from strictly the domain of science to a shared space somewhere between the scientist and the policymaker, acknowledging shared roles in the production not only of scientific knowledge but also practice and national policy. Hajer (2003) similarly discusses the *institutional void* of policy, holding that "there are no generally accepted rules and norms according to which policy making and politics is to be conducted." (pg. 175) Under Hajer, policy is reconsidered when it is needed, responding to the state of the world and the progress of discourse – it is dispersed, expansive, and has new relationships to science, citizen participation, and crosses spatiotemporal scales. (Marks, 1996) Hajer elsewhere characterizes the role of scientific knowledge as one of several discourse coalitions, (2005) who compete to provide narratives and policy frames that define not only the policy landscape but the physical one. (Fischer, 2003)

Sarewitz and Pielke (2006) describe the funding of science in terms of supply and demand, where the supply is scientifically produced, societal outcomes are the demand, and policy performs a mediating role between the two. While few others have taken up this particular interpretation of the relationship between science and societal demand, the need for 'usable science' informing policy and a close relationship to social good has produced a significant quantity of work, particularly in the field of conservation and environmental management. Dilling and Lemos (2009) propose that institutional arrangements could be revised to support the usability of information in climate science. In a similar vein, Cash, Borck and Patt (2006) argue for the conscious and intentional support of policy in defining scientific institutions. Takacs (1996) characterizes biodiversity science as being closely related to the social good of conservation to the point where it is difficult to define how they are bounded away from each other. Ecological indicators, while dependent on scientific models, are not solely the domain of science in no small part due to their complexity and the existence of management regimes and competing conceptions of social good. (Turnhout, Hisschemoller, and Eijsackers, 2005)

Current work done in policy studies consistently indicates that apparently simple arrangements of regulation, science and knowledge production are more complex the more closely they are looked at. What appears to emerge from one realm is present in multiple, and the boundaries between the apparently distinct groups of scientists, policymakers and the public are blurring, particularly when addressing areas of particular scientific relevance. Conservation agendas drive certain types of science, while the needs of policy are seen to be forming others. Policy frames do as much work to dictate the power of science in influencing regulation as the content of the knowledge itself. Overall, the dynamics of policy, particularly as it relates to science production, is a relatively open area amenable to a variety of interpretive schema and modes of analysis.

Infrastructure is oriented around its promises for the future, and presents a poetic view of the world based on desire, on the ideal, and on our goals for ourselves and the world around us. “Roads and railways are not just technical objects then but also operate on the level of fantasy and desire. They encode the dreams of individuals and societies and are the vehicles whereby those fantasies are transmitted and made emotionally real.” (Larkin, 2013, pg. 333) Infrastructure in this mode of enquiry bears a political address – it promises, through the development of technology, a certain future, a dream of the world as subtended by and supported through new infrastructure. A new infrastructural project then is fundamentally about the future and possibility. The material of failed or abandoned infrastructural products are a record of the dreams and imaginations of the past.

As prospective statements about the future, discussion of infrastructures are in many ways about their poetics. In the poetic mode of infrastructure the technical capability of the system is treated separately from the future that infrastructure is represented to enable.

“Infrastructures are the means by which a state proffers these representations to its citizens and asks them to take those representations as social facts.” (Larkin, 2013, pg. 335) The centrality of the material in anthropological infrastructure studies engenders a discussion of “embodied experience governed by the ways infrastructures produce the ambient conditions of everyday life: our sense of temperature, speed, florescence, and the ideas we have associated with these conditions.” (Larkin, 2013, pg. 335) Interactions with infrastructure govern not just the aesthetic experience of the world, they define imaginaries of what is possible and potentially possible, and are presented politically as a pathway to those potentials. The poetics of EarthCube revolve around its promises of novelty, of the goal of producing never-before-seen collaborations, modes of work, and scientific knowledge. EarthCube presents a pathway to a complete model of the dynamics of the earth, a means of understanding in full the carbon cycle, and the ability to predict and potentially even control changes in the earth as a totalizing system. The EarthCube is a four-dimensional representation of the earth – it promises a means to the understanding of the earth as a single, unified system through time. In the next chapter I will discuss and reflect upon my research approach to the site of my study, and present a theoretical rationale for the study and analytic methods I employed.

CHAPTER 3: Study Methods and Methodology

My object of study was the cyberinfrastructure project EarthCube and a constitutive, independent 'building block' project called BCube. In order to study these, I used mixed methods, so as to able to account for the variety of scales of action and discourse taking place. These sites, while objects of interest in their own right, were largely a means by which I might study a relationship that characterizes them – that of the relationship of policy and funding regimes to the production of knowledge in science. Rather than seeking to understand the finances and markets of science, my goal in this project was to understand how policy guides science, and how the knowledge produced through scientific investigations is present, represented and acted upon by policymakers in the context of funding and supporting basic science. Explicit formulations of this concept suffer from substantive complexity, as the 'policy sphere' is comprised of a large number of offices, institutions, individuals and infrastructures in its own right, and does not readily respond to being discussed singularly. As such, I have left 'policy' and 'science' as relatively open categories characterized mostly by self-representation and identity. Policy in this sense then becomes a general term for organizing, management, and regulatory activities performed from a place of authority relative to a particular community. Organizational policy is policy just as much as the law of the land, in this sense: they are abstractions of the role of humans in working through and with other humans with disparate abilities to act in the world. In similar vein, I left the category of 'science' somewhat open as well in order to acknowledge the heterogeneity of knowledge production activities undertaken by those who would identify themselves as scientists. These open categories allow movement between scales of regulation and research in a meaningful way, and enabled my comparison agency-funded cyberinfrastructural efforts with historical policy actions and the dynamics and interrelations of organizational policy.

BCube was funded under the larger EarthCube cyberinfrastructure program, and my engagement with BCube was both a site of close, direct observation, and a means of access to the larger EarthCube cyberinfrastructure project alongside the policy efforts that supported engaged it. I found my way to this project after having been recruited (alongside my advisor) by the BCube team to investigate social and cultural issues arising in design. Within BCube I was a participant observer, (Jorgens3n, 1989; Atkinson and Hammersley, 1994) and while there my role shifted somewhat throughout the project from a more design-theoretic to one of sensitizing and engagement with issues of sociality in infrastructure. The data I collected within and through BCube took several forms, and required a mix of methods in order to approach my particular problem. The data I collected was drawn from interviews, direct observation, and the collection and close reading of documents where I deployed discourse analysis.

Interviews

I initially conducted nine semi-structured interviews of BCube participants in order to uncover their approaches to their own science, the objects of interest to them within BCube, their understandings of EarthCube as a larger effort, and the role that interaction with policy at any of its scales played in the conduct of their research. Each interview was approximately one hour, and while I attempted the snowball sampling approach (Biernacki and Waldorf, 1981) I found access to further interviews outside of the BCube team a difficult task – I loosely speculate that a significant reason for this came from the general sense that EarthCube participating scientists felt a bit over-studied with the large numbers of workshops, social network analyses and other design activities they were engaged in throughout the initial planning and design phases of the project. During the all-hands meetings and at various conferences, primarily the AGU Winter ’14 and ’15 conferences, I also performed brief, unrecorded semi-formal interviews with a variety of scientists, presenters, and other social scientists working at and with the project. I estimate the number of such interactions at approximately 30 conversations ranging from 5 to 30 minutes, not including passive observation of interactions and negotiations among teams at the EarthCube all-hands meeting. I was able to record a portion of some of these conversations (12 total interactions), and took detailed notes shortly after the rest. In addition to this, I collected documentation, drafts of governance plans, and both internal and external communications in order to develop a more complete image not just of the process of design and governance but also the more idealistic representations of EarthCube developed and presented to potential external partners and known stakeholders.

My original intent was to take a grounded theory approach to the analysis of these interviews (Strauss and Corbin, 1994; Clarke, 2007) but the relatively small corpus of such data available to me proved a bit resistant to a fully grounded approach. Instead, I employed some grounded methods in my analysis of the interviews and documents I collected while taking more of a close-reading and discourse analytic approach. (Van Looy, 2003) I performed these interviews in the first four months of the BCube project, and in addition to the analysis they supported, the interviews also served to sensitize me to backgrounds and particular contexts of research from which participating team members emerged. In line with the constant comparative (Glaser, 1965; Dye et al., 2000) method I used these early interviews and attendant observation to characterize my ongoing analysis and guide my data collection. In post-interview and other general memos I attended to potential theory as new interviews took place, and worked to understand future work through that theory while refining it based on those observations. While my corpus of interviews was smaller than planned, they still served as vital windows into the attitudes and values of the project participants, and gave me a much better initial understanding of project roles and commitments.

Participant Observation

I engaged the work of BCube and EarthCube and multiple scales of activity. In addition to my role on BCube as a participant observer I also engaged with the formal and informal *scalar devices* (Ribes, 2014) employed by both the NSF and EarthCube to assess constitute and understand their community of researchers. During the four years I was engaged with BCube and EarthCube I took on formal roles in the (briefly existing) Coordinating Committee, whose role was to work between the NSF, an initially funded 'Test Governance' project with the task of developing an initial charter and governance structure for EarthCube and the funded projects themselves. Each funded project sent two members to the Coordinating Committee, and I was present as a participant observer at the initial Charter Workshop of EarthCube, where its first community-led governance program was proposed.

As the Coordinating Committee did not continue past the drafting of that first charter, I took on membership in the Science Committee and Engagement Team created in the EarthCube charter. The Science Committee was tasked with understanding the science drivers that would help define the eventual EarthCube architecture, while the Engagement Team worked to create non-funded partnerships with researchers, institutions, and data services throughout the life of the project. At these meetings, I was able to observe the self-constitutive activities of the EarthCube community, as well as their mechanisms of self-understanding in the form of surveys and the development of scientific use cases with specific research outcomes. On the BCube project I was a part of the XDT (Cross-Domain Team), which sought to guide the engagement of BCube with the larger EarthCube architecture as well as organizing participation and project activities for its membership. My time on the XDT helped me to understand not only the priorities of BCube's software development but also the diversity of smaller groups associated with the project and their attendant commitments. The XDT also gave me a window into the struggles, barriers and issues with the implementation and development of their middleware broker (which will be defined later) and the means by which they sought to correct those issues. The XDT, informally, served as project governance for the BCube team, and engaged with issues that required external negotiation to be resolved as well as coordinating the planned deliverables and schedules of work among project members. At these meetings I took transcriptive notes, focusing on recording *verbatim* particularly interesting turns of phrase precisely as perspicuous elements for analysis but otherwise capturing the content of the discussion. Following each meeting I produced reflective memos that informed my constant comparison as well as guided me toward document selection for further close reading.

Alongside my formal committee and project team memberships I pursued the scalar devices by which both EarthCube and BCube constituted and understood their respective communities. In practice, this was a combination of my presence on approximately fifty remote meetings of various BCube teams, forty remote meetings of the Science, Coordinating, and Engagement Committees, and my presence at three EarthCube co-located all-hands meetings and three BCube plenary sessions, two of which were co-located, one of which was remote. At the EarthCube all-hands meetings I paid special attention to coordinating activities and the ways in which particular funded projects sought to achieve the collaboration mandate that was part of the call for participation, and participated in governance and funding discussions wherever possible. I presented at two of the EarthCube All-hands meetings – both presentations revolved around the sensitization of EarthCube to sociotechnical issues in line with a mode of engagement as discussed by Ribes and Baker. (2007) Further, I attended conferences of interest to the BCube team and the larger EarthCube community: two meetings of the American Geophysical Union (AGU) annual meeting, and one of the AGU's policy meeting. Throughout all of these engagements I took ethnographic field notes and produced reflective memos (Lofland and Loflan, 2006), in addition to performing a series highly-informal, brief interviews and discussion with various participants that I reflected on in memos following that conversation. During the all-hands meetings and at various conferences, primarily the AGU Winter '14 and '15 conferences, I also performed brief, unrecorded semi-formal interviews with a variety of scientists, presenters, and other social scientists working at and with the project. This constituted approximately 30 conversations ranging from 5 to 30 minutes, not including passive observation of interactions and negotiations among teams at the EarthCube all-hands meeting. I was able to record a portion of some of these conversations (12 total interactions), and took detailed notes shortly after the rest. In addition to this, I collected documentation, drafts of governance plans, and both internal and external communications in order to develop a more complete image not just of the process of design and governance but also the more idealistic representations of EarthCube developed and presented to potential external partners and known stakeholders.

Document Analysis

The EarthCube and BCube projects produced extremely large quantities of documents ranging from posts on communal email list-serves, to white papers on various topics, peer-reviewed publications, slide decks for presentations, promotional materials and organizing document presenting aspects of project governance and its ongoing funding efforts. Committee meetings also produced minutes documents as well as various collaborative pieces of writing oriented towards particular goals. Key in these were the use cases prepared by the Science and Technology/Architecture (TAC) Committees, as well as road-map and community self-assessment documents. In addition, as orthogonal issues related to the scientific and political work undertaken within EarthCube/BCube or engaged by its members emerged, I pursued a closer reading that led to close readings and analyses of policy objects (the Total Maximum Daily Load case discussed in Chapter 1 and the comparative GBIF infrastructure discussed in Chapter 2 in particular). While I was not enable to engage in close reading and full discursive analysis for the 300+ documents present in EarthCube's repository, I used my observations, interviews, and other, more immediate documents to guide a closer reading of particular white papers, workshop results, and summative documents prepared for external reading.

My goal in document analysis was two-fold: the first was to understand the history and constitution of EarthCube as it happened prior to my observing presence, the second was to uncover the discourse on EarthCube, how it varied from funding agencies to participants, and the ways in which EarthCube and its activities were framed in those discourses. My close readings rarely stood alone – from document selection to analysis I was guided by the observations I made, the discussion I participated in, and the meetings I attended. In the constant comparison mode I used documents to refine my theories about my observations while directing my attention towards group dynamics, relationships, and negotiations that I might not have been previously attendant to.

Data Analysis and Relevant Theory

Lofland and Lofland (2006) encourage a naturalistic approach to beginning research, arguing that the methodological and ethical issues that arise from beginning research "where you are." (pg. 13) My participation in the design and configuration of the BCube middleware data broker provided both a set of contacts to serve as an initial interview population and the opportunity to observe the work of data and earth systems scientists collaborating on a project of policy relevance within an evolving organizational context. I had already begun to observe this group as they uncovered issues of governance, economic and social sustainability, and the setting and support of scientific agendas as they worked to build the relationships necessary for the success of a piece of middleware software. As I discovered new dimensions of these issues, and followed relevant actors throughout this design process, new populations for interview and observation (such as emerging governing committees) became apparent. (Rubin and Rubin, 2011)

Generalizable knowledge in this case does not exist in blanket statements about human nature or the nature of research and technology, but rather in the relation of marginalized and previously undescribed issues related to the development of large-scale technology, and the place of science and policy relative to each other. “The careful ethnographer [explores] the knowledge that undergirds the implicit claims that people make… [noticing] what people do in relation to other in order to produce specific, situated meanings.” (Emerson, Fretz and Shaw, 2011)

While the distributed nature of the project means there was unlikely ever to be one lab or specific site that I might visit for direct observation, it also means that substantial portions of discussion and interactions were captured as digital traces. (Østerlund et al., 2014) Most of the observable interactions on EarthCube took place over email, in recorded WebEx meetings, and over various mailing lists. Observation, in this case, largely revolved around the collection and analysis of these recorded interactions. Throughout this whole process I followed the oft-repeated slogan of actor-network theory (ANT): “follow the actors themselves.” (Latour, 2005, pg. 11) This is not limited to human actors within the CI’s network, but also the documents, instruments and technologies that may be seen acting politically or facilitating the movement of knowledge into the public sphere.

Successful science is “extremely social,” (Latour, 1987, pg. 62) and that sociality is not limited to sets of human actors, but rather extends to large networks of both human and non-human actors. This is particularly evident in science, where more and more the work of a scientist is in the deployment and interpretation of instrumentation rather than in direct observation of nature. "People map for us and for themselves the chains of associations that make-up their sociologics. The main characteristics of these chains is to be unpredictable--for the observer." (Latour, 1987 pg. 202) The full extent of the actors enrolled is not immediately obvious when approaching a site until such time as they become evident. This is particularly true in a project as distributed and large as EarthCube.

A major initial part of the research was to identify the actors enrolled and seeking out appropriate political actors to follow throughout the network in order to discover some portion of the wider network/knotwork (Engestrom, 2005) of relationships. Oftentimes these networks are not even fully evident without the introduction of conflict, or controversy – there are still "the set of elements that appear to be tied together when, and only when, we try to deny a claim or to shake an association." (Latour, 1987 pg. 201) It is in the resolution of the controversies and conflicts that the influence of often-hidden members of these networks become evident, and even then often only through the ways in which they transform other actors in meaning and use as their associations change within the network. “Since the settlement of a controversy is the cause of Society's stability, we cannot use Society to explain how and why a controversy has been settled. We should consider symmetrically the efforts to enroll human and non-human resources." (Latour, 1987, pg. 258) As such, only exploring the people related to the project abandons a major source of data – the archive itself.

A third, major part of my research follows Foucault's method of archaeology and discourse analysis –seeking the underlying discursive assumptions that lend themselves not just to the concept of true and false, but also work to separate the political from the scientific, the scientific from the non-scientific, and the underpins the logics that inform them. Archaeology in this sense is in no small part the deep exploration of the scientific and political historical archive – the recorded publications and orations and their complicated relationships with evident scientific and political practice – but also a means to discover the discourse that formed those documents. The historical archive is key to answering the following question: if, as according to Foucault, discourse forms and defines not just what is said but how what is said comes to have meaning, then how can we be sufficiently separate from the discourse as to analyze it without simply repeating its terms? Foucault's answer to this comes in the form of the historical archive (not limited, of course, to the institution of the archive, but rather comprising the recorded record). "[The archive] is the border of time that surrounds our presence, which overhangs it, and which indicates it in its otherness; it is that which, outside ourselves, delimits us." (Foucault, 1985, pg. 130) Archaeology, both historically and as an inquiry into current discourse, requires otherness/separation from the discourse being explored in order to discover the *statement* without being subject to it. Foucault's *statement* is not bound in rules of grammar, signification or meaning, but is the set of rules and assumptions needed for expression to be meaningful under a discourse. And it is in the archive that that separation is achieved, both in temporal separation and in the separation of the object of analysis from the discourse of its creation such that the discourse on that particular topic is not the discourse from which analysis begins.

> “I suspect one could find a kind of gradation between different types of discourse within most societies: discourse "uttered" in the course of the day and in casual meetings, and which disappears with the very act which gave rise to it; and those forms of discourse that lie at the origins of a certain number of new verbal acts, which are reiterated, transformed or discussed; in short, discourse which is spoken and remains spoken, indefinitely, beyond its formulation, and which remains to be spoken.” (Foucault, 1971, pg. 149)

The discourse, once uttered, remains spoken into the future. Discovering and ‘digging up’ evidence of that speech is archaeology that can reveal not just the logics underlying recorded expressions, but also the basic assumptions and discursive formations that form the basic structure of that logic and underpin the reasonableness and appropriateness of both the logic and the expression itself. In the study of science, policy, and political and scientific actors, these assumptions are not just limited to stated methodological, analytic and logical claims but also present in the public discussion on how such claims generate, support, or undermine meaningful speech outside of the scientific and political world. Discourse analysis as a method of research is deeply fragmented in definition, mode and practice. (Potter and Wetherell, 1987, Grant, Keenoy and Oswick, 1998, Keenoy and Oswick, 2004) Alvesson and Karreman warn that in its variety of uses, “discourse comes close to standing for everything, and thus nothing.” They argue for a grounded approach to discourse analysis, moving ‘up the ladder’ from discourse (the text) to Discourse (the society), with continual awareness of the social context of the speaker. Key to their analysis of the field of discourse analytics is the concept of moving between the micro and macro – moving up the ladder from discourses to Discourse. This begins an issue of division, where discourse is used to define and describe two very different things – often in similar language and when borrowing ideas from each other. Discourse can be easily divided into two categories: 'what is said' about a certain topic - speech here can be any form of expression and 'what forms speech.' The 'what forms speech' definition of discourse is found largely in the tradition of Foucault, where the study of discourse on a topic is constrained by the texts discussing that topic rather than embodied within them. Foucault's method of archaeology of text (Foucault, 1970) is designed to pull apart and dig into the language and circumstances of the text in order to reveal the discourse that forms it – the object of inquiry is the social condition that exerts power over what is reasonable to say as expressed in text, not the text itself. The other use of the term, that privileges the text, is an analysis of discourse that tries to generalize about what is said in order to reveal the ways that people express themselves about some given topic.

Potter holds that discourse analysis has an analytic commitment to studying texts and talk in social practices (Potter, 2004), a claim repeated in Potter and Wetherell's focus on the use of language in a social context over any concerns of the application of meaning, an individual point of view, or other more cognitive concerns. (Potter and Wetherell, 1987) Some theorists choose to include cognition and representations as forms of discourse (Clegg, 1989) or focus their analysis on access to discourse, rather than the content of the discourse itself. (Van Dijk, 1993) Other forms of discourse analysis focus largely on linguistic minutiae, focusing on the structure of language and the repetition of units of discourse. (Brown and Yule, 1983) Discourse analysis in this vein has been applied to the study of social and political policy, (Hastings, 1998; Collins, 1997; Fischer, 1993), How a given discourse is framed may affect whether an event, person, or issue is even an object of policy. In Hajer's study of the British government's response to the problem of acid rain, he shows how the framing of the discourse – the generation of a set of assumptions from which the problem might be approached – worked to define where there was a problem, whether that problem was an issue of policy, and eventually constrained the available solutions to that problem. (Haajer, 1993) These approaches to discourse analytic share something in common – a privileging of the language of the text being studied. Even those discourse analytic frameworks mentioned above that focus on the social context of language are 'locked in to' their texts – they speak of power in the world acting through the text itself.

This particular move is warned against by Foucault. "The production of discourse is at once controlled, selected, organized and redistributed by a certain number of procedures whose role is to ward off its power and dangers, to gain mastery over its chance events, to evade its ponderous, formidable materiality," (Foucault, 1981) The discourse, and the action of discourse in the world, is both an effect of a given power structure and an enacting of the power in the world. In this sense discourse both calls back to and works away from the moment of its establishment. Even while constituting a social reality it does not act through its own text, but through existing power relationships. According to Said, "…an understanding of how the will to exercise dominant control… has also discovered a way… to clothe… itself in the language of truth, discipline, rationality, utilitarian value, and knowledge. And this language in its naturalness, authority, professionalism, assertiveness and antitheoretical directness is what Foucault has called Discourse." (Said, 1983) This is a common misreading of Foucalt's writing on discourse – it is not in the naturalness of language that Discourse is evident, but in the hidden claims and divisions that inform and define the language itself. Discourse underpins in language in defining what is reasonable, sane and logical to be said. Discourse, according to Foucault, is able to exercise power not through its text and content but through its ability to limit the available space of rational response. It is "[forged of] three great systems of exclusion, the forbidden speech, the division of madness and the will to truth." (Foucault, 1981) In approaching a site like EarthCube, these exclusions are most evident in the sort of facts assumed to be true – they paradoxically more closely represent the discourse they enact the less-investigated they are. The statement 'it is a good idea to share data', for example, would have been entirely unexamined in the context of a project like EarthCube, despite its not necessarily being true – it is in this sort of assumption that the discourse it enacts becomes most immediately evident.

Power is the ability to act. To define and establish the discourse is to establish the criteria by which truth, appropriateness and sanity are established. To speak outside of the discourse is to speak as a madman, to speak without sense. As a pair of brief examples, there exists a discourse of science where to speak against the perceived consensus of that science is presented as being irrational, senseless, or fundamentally out of touch with reality: climate change denial and antivaccination movements. To engage with, to study the discourse is to go beyond the language of the discourse and through it. Foucault warns against privileging the language of the text in its analysis – to avoid commentary on an object of text in favor of an archaeology of the circumstances of that text. “Commentary exorcises the chance element of discourse by giving it its due; it allows us to say something other than the text itself, but on the condition that it is the text itself which is said, and in a sense completed.” (pg. 58) Commentary “… limits the chance-element in discourse by the play of an identity that would take the form of repetition and sameness.” (pg. 59) By Foucault’s estimation, privileging the language of the text allows one to only speak the text – it limits what can be said in the discourse and about the discourse to only what exists within the text. Based on the writings of Derek Hook, “One should approach discourse less as a language, or as textuality,” when informed by Foucault, “than as an active ‘occuring’, as something that implements power and action, and that also is power and action… discourse is the thing that is done.” (Hook, 2001) Discourse does not act only upon its own texts, nor does it act specifically within or from its texts – the discourse itself is an event in the world, the acting-out of the power to define what is true, what is known, what is sane, what is taboo. All this is resolved in a methodological avoidance of grounding discussion of a discourse too heavily within the text of that discourse. “The analyst of discourse, then, is predominantly concerned with exploiting the gaps or shortcomings of a given discourse.. these are the seams to be pulled, the joints and weaknesses to be relentlessly stressed.” (Hook, 2001) In investigating EarthCube I pulled at the notions of funding supporting relevant science rather than working to produce the concept of relevance, the apparent separation between the policymakers funding science and the comportment of that science, and the notion of novelty as a prime determiner of ‘good science’ in order to expose the discourse I was observing. While discourse seems to enact power in the world, it does so in a way that is not dependent on its own text or its own content – it becomes an actor in its own right, a point of focus and attention that bears material power that is not present in its own words.

As I worked with the BCube team I paid special attention, following from this, to structures of power – to the way in which participants accounted for and responded regimes of policy, the requirements of funders, and negotiations with 'peer' projects. My goal in this was a perspective that not only acknowledges the actors categories, (Collins, 2008) but also approaches them with a critical eye, evaluating who is within versus without, where particular modes of action are emphasized or avoided, and how the dialog of success and failure vary throughout project implementation. As I was approaching this through the lens of scalar devices, (Ribes, 2014) I paid special attention to the discursive work being done at points where groups interact at varying scales. The data I collected at these meetings were accounts of how the actors I was engaged with were not only representing their work, but representing the totality of the infrastructure project. I paid special attention to the mechanisms by which purported dynamics of action were expected to be achieved. In the case of BCube, their focus on agile software development methods (Beck et al., 2001; Martin, 2002) not only revealed expected leadership roles and expectations of personal responsibility but also the set of values and attributions of value that were espoused by the leadership of the project. In similar way, I interrogated other mechanisms by which projects interacted and represented themselves as doing so – whose language changed during interactions? Where did particular methods for assessing communities, coordinating action, or seeking funding persist or fail to do so? Did particular language propagate throughout connected groups or did they hold to their own formulations?

Another aspect of the analysis of discourse that informed my approach to this site (particularly EarthCube as a whole, though to some extent in my analysis of BCube as well) is the concept of policy framing. In short, policy frames are a discursive characterization of particular policy actions, especially as they are presented in formal communications and through the media. (Snow and Benford, 1988; Gamson and Lasch, 1983) Policy frames are a tool, methodologically, for interpreting how disparate groups work to develop, communicate and share particular understandings of reality. (Reese, 2007; Borah, 2011) According to Entman, “Frames have at least four locations in the communication process: the communicator, the text, the receiver, and the culture,” (Entman, 1993, pg. 52) and while Entman seeks to characterize the work of framing according to scientific paradigms in the mode of Lakatos (1970) and argues for a single paradigm for the interpretation of frames, this mode of approaching policy frames has been substantively challenged. D’angelo, (2002) argues for multiple paradigms of research work, while recently there has been a move towards a closer look at the concept of framing, rather than an investigation of the frames themselves. (Hulst and Yanow, 2016) It is in this mode, the investigation of the framing itself, that I approach my initial interpretations not only of the site itself but of the discourses of planning, design and negotiation informing the character and in many ways the shared reality of EarthCube. As a collaborative, multi-institutional product that spent a great many years in a largely imaginative space, EarthCube in many ways was the result of an act of policy framing in its own right. I paid special attention to the ways in which scientists discursively constructed their communities in light of the framings presented by the NSF, by institutions, and by other scientists, as well as the way in which these communities were presented, constructed and framed by others. Specifically, I looked for places in which conflicting framings of the nature, design, and role of EarthCube interacted or competed for attention and discursive uptake among the community.

My study of the individuals I worked and observed within the context of their daily activity was rooted much more closely in the principles of situated action. (Lave and Wenger, 1991) Situated action, according to Nardi, "deemphasizes study of more durable, stable phenomena that persist across situations... it is a one-time solution to a one-time problem, involving a personal improvisation that starts and stops with the dieter himself" (1996, pg. 26) Situated action emphasizes the immediate, the context in which action is taken in that moment, and treats plans more as representations, resources and other reasonings about actions, that should be differentiated from the work and action itself. (Suchman, 1987) Where Suchman, and others who more completely employ situated action approaches to studying the behaviors of individuals and groups, would de-emphasize the documentary, the prescriptive and the discursive in favor of close attention to the action itself, I would argue that these represent significant portions of the context out of which action arises, and the a close, discursive analysis of documentation, presentations, and plans informs an understanding of that action by representing important aspects of that context. I sought to understand, in my analysis, the way in which design documents, imaginaries, and various formalized plans influenced the observed behaviors of my studies population while observing their actions in the actual moment.

Conclusion/Reflection

Broadly distributed sociotechnical projects like EarthCube are exceedingly difficult to observe in their totality, and would represent significant work for even a very large team of researchers – which would be the ideal multi-sited ethnography approach in the vein of Marcus. (2009) Approaching the project through the lens of policy interactions and scalar devices allowed me to select objects and sites of interest, as well as an entrée to the theory I generated. Discursive frames gave me leverage into a policy landscape as well as a means of understanding how EarthCube and its governance changed over my time associated with the project. A combination of observation, interviews and document analysis worked well towards helping me characterize the site as well as my objects of interest. However, I found my role within the BCube team, my shifting responsibilities, and my level of engagement with that team to be somewhat less effective that I had hoped it would be.

As a participant observer, my desire to maintain access to the site led me to promise a bit more than I could effectively expect to deliver to the project team in terms of deliverables, and often distracted me from my observational activities. When I first arrived, I expected to be in a user-testing role and had prepared talk-aloud protocols as well as various observational metrics for evaluating the use and effectiveness of the broker. However, implementation of the broker into the testbed was a slow process, and a testable broker only really existed in the final months of the project. As the broker was designed for scientists in specific domains, it was difficult (in fact, a dead end) to find sufficiently capable researchers to test the software that were not already a part of the project team. In future projects, I would work to more carefully define and design my role to allow my primary focus to remain on the research work I am doing as part of the project, rather than so closely aligning my work with project-specific goals and concerns. While my inability to meet certain project deliverables was partially the result of factors outside of my control, there was still a splitting of focus from the research I was doing that I feel did not serve me well.

It is still unclear to me the extent to which I ought to have addressed the broader community of EarthCube researchers more directly. While focusing on BCube as a pathway into EarthCube as a whole allowed me to more closely observe a relatively small group of people, at the end of the day BCube faded from visibility in terms of the architecture as a whole. The general lack of policy and organizational support from EarthCube redirected the focus of brokering activities for a significant portion of the team to its implementation in the similar EuroGEOSS (eurogeoss.eu) program – at the end of BCube's life there were certain capacities still not implemented that had been successfully implemented in the EuroGEOSS instantiation. In the current presentation of EarthCube, the BCube middleware broker is not even listed as a completed or in-progress part of the overall catalog of tools that makes up the infrastructure. If I were to approach EarthCube as a site of study today, BCube would be closest to vaporware - a project that existed as white papers and presentations but is no longer accessible as software. In the mode of Latour's analysis in *Aramis,* BCube became more or less ontologically real throughout the life of the project on the basis of its ongoing work and lack of inclusion in the final architecture, but the associations, skills and techniques developed through its work persist beyond its funded life. (Latour, 1996) A broader engagement with a wider variety of individual projects might have helped me better understand the landscape of work being done on EarthCube and through its communities, but might also have further diffused my research focus to the point that it would be difficult to discuss the project at all. At the end of the day, my concern with maintaining effective access guided a large portion of the decisions I made – in future research I would seek to construct my access earlier in the life of the project to more closely focus on objects of interest as they emerge.

Throughout the following chapters I discuss not only my observations of BCube and EarthCube, but also attempting to address my larger question of the policy/science relationship through orthogonal cases that emerged from the work I was doing with EarthCube. I learned of the concept of the Total Maximum Daily Load (TMDL) of a watershed, and specifically the Chesapeake Bay Watershed as the largest TMDL in the nation through an interview, which prompted the close readings and document analysis that lead to the bulk of the first chapter. Similarly, my interest in infrastructure funding and investment led me to the comparative case on GBIF as presented in the second chapter, which gave me some insight into the pathways to moving from a speculative system and funded development project to an instantiated infrastructure supporting science. In this thesis I will discuss and work through the relationship of policy to science through the lens of EarthCube and BCube as it became visible in observation, through interviews, and was presented in documents.

CHAPTER 4: The Cybernetics of Science Policy: Feedback loops, regulatory regimes and management techniques.

There is a fairly fuzzy distinction between the knowledge production work done under the auspices of scientific inquiry as compared to similar work done under the umbrella of policy formation, framing and funding. Moving forward, it is important to remember that this is an actor's category mostly representing the discourse as I heard, read and observed throughout my time at EarthCube and Bcube. While to a certain extent these are valuable categories in terms of understanding the push-pull of negotiation between organizational entities and funded scientists in determining relevance, importance and impact of scientific work, they also serve to obscure the actions and relevance of the knowledge itself as it is produced. In this chapter, I look closely at two different interactions between policy and earth systems science to uncover the work and utility of scientific knowledge produced towards policy goals. First, I discuss conservation and hydrologic efforts relative to the management of the Chesapeake Bay Watershed, a hydrologic site with a relatively long history of policy-directed study and intervention. Secondly, I characterize EarthCube as a site of policy work, and account for large infrastructure projects as a site of investment in basic research. This chapter sets the terms for my discussions of EarthCube and BCube as scientific endeavors enabled by, related to, and interacting with regimes of policy and policy knowledge production. As a cyberinfrastructural effort, EarthCube's design, structure and modes of work are closely related to the policy goals from which it is produced. Throughout this chapter I show a relationship between science and policy that provides a basis for interpreting and understanding how the engagement of policy in infrastructural development supports basic science, constitutes particular scientific communities, and engages technology and regimes of scientific knowledge production. So far I have discussed science and its infrastructure bearing political weight, informed by policy, resulting in regimes of policy and new policy frames, deeply concerned with governance and the artifacts of policy, and achieving societal relevance through framing discourses similar to those that inform governmental policy. I have (somewhat critically) respected the stance of my research site and informants in maintaining that somewhere in the practice of science there is an apolitical seeking of knowledge whatever its potential social implications. In respect to my observations I have kept with the language and assertions of my subjects, their uneasy relationship with political activism, and bound parts of my conversation to the image of scientific outcomes flowing outward in some way from academia towards a sphere of policy and public action. In this chapter I modulate these characterizations of this dynamic and paint a different picture of scientific work – one that places science not only firmly in the dirt of politics, but one where the practices and infrastructures of science are fundamental to government even when its basic assertions are challenged by members of that government. This is concerned, in the mode of Serres, (1990) with humanity in general, and national governments in particular, for the first time in human history engaging in the work of planetary management.

Computing, both in the sense of performing computations and in the sense of advances in storage, collection, and distribution of data, has opened the door to understanding local change in terms of an understanding of the global. Model-driven simulations of planet-scale systems are increasingly able to evaluate and predict how localized anthropogenic change propagates throughout the world. In this narrative we are ever closer to pairing the apparent chaos of the global climate system down to a single, predictable, modeled world. In theory, a perfect model of the system results in perfect policy: understanding, predictably, how any change might propagate throughout a global system would allow policy-makers to weigh issues of human value against known environmental impact, and those conversations would revolve around the weighing of those values rather than the validity of the predictive model. The politics and diplomacy of the managed planet are enabled, informed, adjusted, and described through the work of scientific observation and analysis. While the reality of the situation is obviously more complex than this optimistic, positivistic narrative (Edwards, 2010; Latour and Woolgar, 1979), it does speak somewhat to a very hopeful view of the way that scientific work might benefit society, and the way that political action might be best informed by scientific knowledge.

In *States of Knowledge,* Sheila Jasanoff writes that "what we know about the world is intimately linked to our sense of what we can do about it, as well as the felt legitimacy of specific actors, instruments and courses of action." (2004, pg. 14) Following from this, she makes the claim that science and technology are inherently political goods, with substantive consequence not only for the location of power, but also the means by which that power is exercised. In Jasanoff's conception of the relationship of science and technology to policy and politics, two relatively distinct groups are working together to 'co-produce' not only scientific knowledge but also the means by which that knowledge is evaluated, expressed, and put to use. However, there remains an entire political ecology that, more than just responding to research and modeling, would not exist without it – even the tools and vocabulary for discussing that political space emerge from the work done studying earth-scale systems. From start to finish this work is inevitably political, and more specifically is the means by which global politics happen – the regimes of observation, analysis, modeling, and prediction arising from earth systems sciences are both the meat and grist of policies that, while informed by external frames and discourses, are solely interested in managing the results of that science.

In later sections of this chapter I will be discussing the Total Maximum Daily Load (TMDL) of a watershed as both a scientific, model-derived concept and a site of significant regulatory action. In order to address problems with surface water quality (a specific policy goal informed by particular industrial and aquaculture outcomes) policymakers needed not only a means of effective assessment but also predictive knowledge directing intervention efforts. This science, in Gibbon's Mode 2 conceptualization, (Gibbons, 2000) accounts for the sociopolitical context in which the research is occurring – problem formulation is tied particularly to social outcomes, and the relevance of that science emerges in part from market pressures (initially, in the case of the Chesapeake Bay area, the health and prevalence of oyster populations balanced against shipping and transport infrastructure). The needs of the market and knowledge deemed necessary for policy formation and framing here work not only to co-produce but significantly to form the nature of scientific inquiry into the watershed through a discourse on relevance, timeliness, and importance. Throughout this writing I am interested theoretically in blurring the lines between the political and the scientific, particularly when it comes to issues of knowledge production and infrastructure. (Edwards, et al. 2013) Latour's metaphor of the Janus-face of politics and nature as outlined in *Science in Action* is particularly compelling here. (Latour 1987) Latour characterizes his scientists as facing simultaneously towards nature as a source of truth while looking to conventions of practice, and general day to day actions in the actual production of that truth. I interpret a similar dynamic in the discursive production of scientific relevance. One face points towards an independence of science and inquiry as determined by the internal characteristics of the scientific field, while the other points to broader, application-focused concerns like resource availability, funding drives, and a broad discourse on the concepts of importance and the need for action in particular areas. However, this brings up a particular theoretical and definitional distinctions that I shall carry throughout this writing: those between dynamics and mechanisms. For present purposes, I define *dynamics* as observable activities, while *mechanisms* refer more specifically to the characteristics of those observable activities that in some way work towards the results of the observed dynamic. For example, Jasanoff (2009) writes about how scientific advisors influence policy through the recontextualization of particularly policy-oriented knowledge and semi-formal advisory roles tied to certain policy structures – a role she emphasizes the importance of by calling it the *fifth estate*. The presence of those scientists, their movement into advisory roles, and the resulting shifts in policy are the *dynamics* of the system she presents. White papers, presence at specific meetings, even telephone availability are examples of the *mechanisms* by which this *dynamic* functions. While I deal somewhat loosely with the overall dynamics of the relationship between science and policy, I am particularly interested in the mechanisms of knowledge production within that relationship that enable those dynamics. Presence on an advisory board does not inherently change policy action – there may be any number of ‘rubber stamp’ scientists or lame duck advisors. The mechanisms by which scientist-advocates might provoke policy change such as presentations at particular meetings, the preparation of statements (written or oral), recontextualization of research towards particular policy goals, and the active engagement of policymakers in understanding not only the state of the science but the constitution of its communities is more closely aligned with my research interest in this area. Of particular interest to me is the way in which policy goals work to construct, evaluate and support knowledge infrastructures (Edwards et al., 2013) to produce policy outcomes.

In this chapter I discuss a particular case of watershed management with a relatively long history in the United States – and one that bears a relatively unique political weight due to its proximity to the centers of American government – the Chesapeake Bay watershed. In addition to discussing the Watershed as an object of scientific/political significance that has produced a particular infrastructure of knowledge production and practice supporting policy action at a variety of scales I will be discussing the Chesapeake Bay Watershed’s conservation efforts as a field of work for those currently involved in EarthCube and describe how narratives of infrastructural design correlate with and respond to existing practices and techniques in the world.

Early History of the Chesapeake Bay and watershed as a policy/technological site.

Chesapeake Bay is a story of oysters and infrastructure. And not just infrastructure in the singular, but in the multiplicity of its interactions. While initial regulation of the

Chesapeake watershed was concerned with the availability of shipping lanes and the preservation of fishing, it also produced a knowledge infrastructure of people studying and understanding the Bay itself as well as the larger watershed feeding it. On March 28, 1785, there was the first agreement regulating access to the Chesapeake Bay made between Virginia and Maryland (both states bordering the Bay) for fishing, tolls, shipping, and the maintenance and support of water transportation infrastructure like lighthouses, buoys, and so on. (Rowland et al., 1888) This took place at a time in the history of the United States, where power distribution between the individual states and nation were under significant contestation – in this compact is also an agreement that currency exchanges between the states and for tolls be conducting in gold or silver by weight and at the same value. In this convention Pennsylvania even wrote a letter to express their concern with the potential of shipping tolls, tariffs or fees exceeding the cost of investment necessary to make the waterway navigable.

> "It is thought reasonable that... all articles of produce or merchandise, which may be conveyed to or from either of the said two states... shall pass throughout free from all duties or tolls whatsoever, other than such tolls as may be established and be necessary for reimbursing expenses incurred by the State... in clearing, or for defraying the expense of preserving the navigation of said rivers." (Rowland et al., 1888, pg. 422)

The Mount Vernon Convention, where the above agreement was made, provides a rich insight into the regulation and negotiation of transportation infrastructure prior to the prevalence of roadways, railways, or other land-based transportation. Pennsylvania had an interest in this agreement because the Potomac river, which feeds the bay, is one of the closest shipping lanes available. Water as transportation infrastructure here is a perspicuous case because, barring canal building, rivers, lakes, oceans, and so on cannot be moved from one place to another: it is simultaneously a negotiation between the states who lay claim to regulatory authority of the water and its attendant resources and the planetary system itself. Decisions made on the waterway affect broad communities, and regulation is responsive to the characteristics of the waterway, rather than the other way around. So we have here a multiplicity of infrastructures: the waterway itself serves as both transportation and agricultural infrastructure, bearing consequences on industries such as fishing, farming, shipping, and travel, while we also see negotiated regulatory agreements between newly-empowered states creating an infrastructure by which those waterways might be navigated (lighthouses, dredging, buoys) and regulated (the Act itself). In the 1800s we see the beginning of policy, as represented in legal action, reacting to shifts in technology. Technologies of the day were recognized within and informed the Mount Vernon Convention. Navigation technologies in the form of lighthouses and buoys, and shipping technology emphasized in Pennsylvania's interest in securing access to the waterway as the most efficient means of distant transportation available influenced both the content and the character of the negotiations. (Rowland et al., 1888) In the 1800s, however, a technological shift occurred that brought not only infrastructural concerns of waterway regulation to the forefront of negotiation, but also those of conservation and the protection of natural resources. In particular, the development of a device designed to trawl oyster beds while traveling along the Potomac river presented an immediate regulatory challenge to both the states of Virginia and Maryland. Dredges similar to the ones pictured below (fig. 1) were used to mine large reefs of oysters, and were quickly made illegal by both states, only to be re-legalized shortly after the end of the American Civil War. (Cronin, 1986) "For oysters, the coincidence of the importation of deep-water dredges, development of new technologies, high demand, and the discovery of large unknown beds resulted in a new important industry and changed the ecology of the Bay. The effects of poor management were also discovered" (Cronin, 1986, pg. 188) In addition to improved dredgers, canning technology made it not only feasible but profitable to harvest far more oysters than the local demand entailed, the development of railroad infrastructure expanded potential market size, and the demand for oyster shells also rose, prompting harvesters to break apart the oyster reefs themselves. Technologies supporting overharvesting, labor issues, and market demand all came together in the late 1800s and early 1900s to significantly deplete the population and viability of the oysters in the bay, bringing their conservation and regulation to the forefront of watershed management policy. (Kennedy and Mountford, 2001) During this time oyster harvests began to drop precipitously, from 14 million bushels in 1874 to 10.6 million bushels in 1879–1880. (Grave, 1912)

Infrastructure projects across the country both depleted resources in the watershed and provide increasing access. Policy interest, in the form of requested reports, increased attention from legislature, and social action to some extent produced a scientific interest in the area, with assessments of oyster population both of significant interest as well as up substantial contestation. A report was commissioned that recommended the privatization of oyster stock in conjunction with regulating farming and re-establishment of oyster beds, among other things. (Winslow, 1882) In this report, informed by the advocacy and research of commission member W.K. Brooks, there was proposed a significantly science-based approach to managing oyster stock, with regular inspections by the Oyster Police, as well as the opening and closing of particular oyster beds as determined by appointed experts in the field. These recommendations were, by and large, not enacted. "Despite years of advocacy, Brooks and his successors failed to persuade Maryland legislators to impose effective conservation measures on the state oyster fishery... Nor were they able to persuade politicians to encourage the oyster industry to accept intensive scientific management... The involvement of Brooks and other researchers in the Maryland oyster culture debate illustrates the weak role of scientific authority in influencing public policy making on a local level." (Keiner, 1998, pg. 384) While the states here had an historic policy interest in understanding, supporting and managing the Chesapeake Bay and its resources, it is clear that actual decisions made in regards to regulation were not strictly aligned with contemporary science. To phrase this a slightly different way, the knowledge production and decision supportive regimes for policymakers were not solely influenced by scientific inquiry in terms of policymaking. In terms of monitoring activities, however, researchers were trusted to report. Brooks' numbers on the current and declining population of oysters was apparently believed, but his ability to influence future policy, regulation and management was limited.

Keiner, as above, suspects that this limitation emerges from the inability to accurately predict future yields. However, Keiner's conflation of the lack of policy outcomes drawn directly from scientific recommendations with the weakness of science's ability to influence the production and enforcement of particular pieces of legislation fails to account for a core institutionalization of that knowledge. Conservation was clearly of concern, but also of concern to policymakers was the availability of common land for oyster harvesting, the ability to support local industry, and nascent concerns about empowering the Oyster Police to close and open particular beds essentially at will. Complete abdication of regulatory power over the watershed and its resources (something negotiated and maintained by the states from the earliest days of the United States) is somewhat different from a "weak role of scientific authority" – while conservation was at issue and was a goal of state policymakers as evidenced by the existence of commission in the first place, other factors and stakeholders outside of oyster populations were also influential on the eventual policy outcome. "There was much hyperbole in [Brooks'] writing but the gist, repeated by commentators then and later, was that political sensitivity to the wishes of oystermen (the result of the desire of politicians to ingratiate themselves with the oystermen for their votes) was contributing to the decline of the oyster industry in Maryland." (Kennedy and Breisch, 1983, pg. 160) However, knowledge produced by Brooks and other biologists about oysters was acknowledged and incorporated into legislation, and was the basis for regulation – "While the legislature ignored many of the Oyster Commission's recommendations, it did pass the Cull Law of 1890, which Grave considered to be the most efficient method ever devised for protecting natural oyster beds... It also set a minimum legal size of 21 inches for market oysters. Maryland was one of the first states to attempt the enforcement of such a law." (Kennedy and Breisch, 1983, pg. 160) While Keiner emphasizes the Maryland government's unwillingness to totally support the recommendations of Brooks (something he himself lamented in his 1891 publication of his report) (Brooks, 1891) as the weakness of scientific knowledge production's ability to immediately influence legislation, the passage of the Cull Law, and later the Haman Oyster Law and subsequent Maryland Oyster Survey and Board of Shell Fish Commission's report can serve to somewhat redefine that ‘weakness’ as more a product of incremental amelioration than a particular failure of policymakers to recognize scientific knowledge. (Grave, 1912, Yates, 1913) These reports provided an “tremendous accumulation of information, although incomplete in some aspects of the life history and biology of oysters, was undoubtedly sufficient for arresting the decline in production and for restoring at least some of the former economic strength of the industry, including the oyster packing industry.” (Kennedy and Breisch, 1983, pg. 161)

Throughout this story is a dynamic whose mechanisms I will be interrogating somewhat further in later sections, focusing more closely on more modern legislation. In this case both the transportation infrastructure and the management of oyster populations is somewhat characteristic of how policy as embodied in the actions of legislature accounts for and accommodates new knowledge, technology and social orders. First, a policy response to a particular problem expands the policy frame to account for and work with a given piece of knowledge or technology, such as regulating by first banning then legalizing oyster dredgers or accounting for the environmental, social, and market changes brought about by railroad, and later, roadway, development. Then, based on the representations of that knowledge or technology and understandings of how legislation might be able to provoke change the knowledge produced by science becomes incorporated into the landscape of that policy. Coupled with the initial policy access to the technology is the development of a monitoring regime – this can be either technical, human, or some combination that provides feedback not only on the efficacy of the particular policy action but also on the more general capability and response to the technology among the served populations. In the case of the early days of the management of oyster populations in the Chesapeake the monitoring regime was largely comprised of commissioned reports indicating the current state of the population as well as its yields for fishing. More interesting here is that the history of reports and legislative action begins to form something of an infrastructure that supports further regulation and management. The Oyster Commission report enabled future reports and commissions, and those recommendations that were followed began, over the next several decades, to build into a management regime that looked from a distance a lot like what was originally recommended by Brooks, with particular oyster beds being leased for fishing and ongoing monitoring activities both informing and informed by policy.

On the other hand, infrastructure has had significant impact on the region, from initial regulation of the watershed related to the negotiation of transport and shipping rights between states immediately following the revolutionary war, to the building of roadways, bridges and other transport. In many ways, the scientists working on the Chesapeake in the early days of its regulation and during the Oyster Wars benefitted from a ontologically limited view of their subject. Recommendations of scientists towards supporting the conservation of oyster stock did not account for the ethics, convictions, and practices of local oyster fisherman and other stakeholders with a vested interest in their availability, management and regulation. However, the mechanism of the commissioned report as both a monitoring tool and a site for recommendations of legislative action is characteristic of the management of the Chesapeake watershed, with periodic reports, commissions and subsequent policy adjustments following at a regular pace from the initial Oyster

Commission report. The incorporation of scientific knowledge into the policy process was not a total commitment to a single report, but a process of incremental amelioration, adjustment and action limited by the need to account for stakeholders beyond the subjects of the study and those performing it. These mechanisms, over time and with input from participating scientists, closely resembles the dynamic of translation described by Callon in his investigation of the scallop fisheries of St. Brieuc Bay. Callon, in describing the way in which fisherman, scallops and scientific community produce a discourse of certainty through the mechanism of translation (1986, pg. 19) presents a view of apparent scientific certainty producing power in its own right. In the case of the Chesapeake Bay in this discussion, the scientists speak separately from the fisherman, from the shippers, but at least make some claim to represent the oysters. While there is little visible exercise of power in this arrangement outside of that expended by the state in producing regulation, the process of commissions, reports, and ongoing research translates not only knowledge produced by scientists but also knowledge about scientists in an ongoing, iterative process. With this early history in place, I will skip ahead several decades to consider more modern, technologically driven science to demonstrate the way that legislation, regulation and framing of law not only accounts for newly produced knowledge through these longitudinal, complex loops of feedback, monitoring, and regulation but also the presence of new scientific methods and technology.

TMDLs and the 2001 NAS Report

While the early management of the Chesapeake Bay relied heavily on direct human observation of oyster yields and oyster beds (and the collection and analysis of the oysters themselves), the modern object of policy that I wish to discuss in some depth in this section is a means of observation through computation and prediction – the computer model. The total maximum daily load of a watershed (TMDL) is something that almost inherently cannot be measured without some sort of computational modeling. Watersheds are geographically distributed, heterogeneous with regards to government, nearby population, and environment, and would require significant labor to directly observe in the mode of the earlier commissions. In these sections, I am largely concerned with the way that policy work develops institutional knowledge and builds particular accounts of the capacity of new technologies to create that knowledge. While the early commission reports were based on population sampling and inference about the total population, modern regulation has a much closer relationship with particular monitoring, management and analytic technologies – a relationship that has placed the work of modern science much closer to the concerns of regulation than prior. Rather than a mostly-advisor, or report-back role, we see scientists working directly on the regulatory object of the TMDL (which I will discuss in more detail following this section) alongside policymakers. While policymakers regulate on the basis of the TMDL itself (something encoded into law), there remains a significant community of researchers seeking to improve the accuracy, predictive power, and capacity of the TMDL as a measure of watershed health.

The story I wish to emphasize relative to the history of the Chesapeake Bay watershed and, more specifically, around the concept of the TMDL as a regulatory and scientific tool revolves around the development of new knowledge among scientists and how it was adopted and implemented by management agencies. We see here, both historically and in recent accounts of the TMDL a simultaneous recognition of technological capacity alongside the scientific knowledge assessing their effects and uses. In the case of Brooks and the early history of Chesapeake legislation we saw acknowledgement and institutional recognition of the outcomes of science and effects of technological development on the yield of oyster beds, even where the recommendations of scientists were not formally encoded into regulation. (Kennedy and Breisch, 1983)

The TMDL is a model-derived, composite measure of the presence of various pollutants in a given defined watershed (all rivers, springs, and wetlands that drain into the ocean in the same place). Compared to the other numbers discussed in this section the TMDL is both the most complex and the object of the most long-term work. Also unlike the other measurements which leverage some policy frames, or seek to simplify very complex phenomena towards a particular policy outcome, the TMDL was implemented as a solution to a complex problem and a regulatory tool. As the other two numbers I will discuss in this section (global temperature and species loss counts in biodiversity) are used to motivate action much more than being the objects of action or site of work themselves, more time will spent on the special case of the TMDL as it relates to water conservation work in general, and policy- and science-driven conservation efforts of the Chesapeake Bay watershed in particular. The TMDL is a particularly interesting measure because it conceptually can fit and work as a variety of information-laden objects: a boundary object between the work of science and that of policy, an infrastructural object that works to support a variety of other activities while still being the site of work to significant communities, and as a singular, value-laden representation of the state of a complex system.

A measure of total maximum daily load of pollutants is written in to the 1972 Clean Water Act, which, among other things, granted the power to enforce water cleanliness standards to the EPA (though they did not choose to for some time) and requires that states both identify water objects that are at risk and identify the TMDL that is required for that body of water to be considered healthy. The TMDL was an attempted solution to a problem relatively particular to attempting to study, understand and regulate moving bodies of water (a problem likely as old as agriculture itself). While water cleanliness standards were identified in the Federal Water Pollution Control act of 1948, there was not a significant measure of either health or pollutant levels, and enforcement was largely focused on regulating 'point sources' of pollution – outflow pipes, drains, etc. that could be identified and monitored. (Barry, 1970) This failed to effectively account for non-point sources of pollution, like the slow leeching of pesticides and other agricultural byproducts into groundwater and general runoff from mining operations, urban areas, etc. Enforcement responsibility was given to the Army Corps of Engineers (and was largely focused on preventing disruption to shipping), and only applied to inter-state waterways.

"The goal of the TMDL is to eliminate an impairment, not meet a pollutant limit for its own sake. The TMDL itself does not establish new regulatory controls on sources of pollution, and it does not set discharge limits. Nor is it self-implementing." (Copeland 2014, pg. 1)

The construction of relevance in science

Behind the scenes in this history is a picture of the growth and development of relevance – and we can see this production of relevance in a couple of different ways. Most immediately the concept of relevance would seem to be a discursive framing. (Fisher, 2003) The oyster beds became important because of their role in local industry, but the conservation agendas initially formed in response to economic demands grew as additional commissions began to accumulate knowledge and recommendations. As the scientists and commissioned investigators engaged in the policy process through the accumulation of knowledge about the oyster beds, behaviors of oysters, and the effects of current harvesting practice there was a growing accumulation of discourse about the role of oyster beds in the local community. (Wennersten, 2007) However, there is more to this dynamic than the simple accumulation of discussion – there is also an element of the feedback loop in practice. Incremental legislation and slow-developing regimes of regulation and enforcement served not only to ameliorate some of the ecological effects of fishing, but also served as a sort of confirmation of the science being presented in commissioned reports. Increasing privatization of the oyster beds (something outright rejected when proposed by Brooks) occurred as part of this process of monitoring, adjustment and assessment.

In some sense, this feedback loop provides evidence that a particular issue is discursively relevant to both policy and science. The presence of ongoing monitoring regimes, investment in regulatory activities, and ongoing assessment/adjustment/evaluation reflects a consistent, ongoing interest in the effects of both knowledge production and the policy activities that it informs. Relevance here seems closely tied to ongoing innovation – new policies, new science, and new technologies assist in perpetuating regulatory feedback loops as well as prompting further innovation in that space. New technologies, techniques, models, or just research results 'refresh' the scientific and political relevance of the topic, occasionally prompting adjustment – a quick example of this is the existence of consensus-demonstrating documents being published out of earth system dynamics. (Oreskes, 2004; Doran and Zimmerman, 2009; Cook, et al. 2014) As scientific work continues on adjusting, verifying and collaborating on global climate models there appears to be a simultaneous desire to not only participate in the policymaking process but also to 'refresh' the discussion by presenting this ongoing work as evidence of its own validity. (Lewandosky, Gignac, and Vaughan, 2013) Demonstrating consensus is a particularly political act in this case insofar as it responds to the narratives questioning scientific results often used by opponents of certain environmental regulations. (Lindell and Brandt, 2000)

We saw in the early history of Chesapeake legislation that legislation, policy framing and regulation acknowledge science even when disputing particulars, and management is concerned with more than the best available knowledge and techniques. Innovation in this sense is a balance between the leading edge and infrastructural capacity, and the infrastructure itself is not necessarily the leading edge. I find evidence of this in the ongoing processes of reporting, adjustment, interpretation, analysis and presentation of science relevant to a certain policy outcome – the science itself is part of infrastructure leveraged by policymakers towards the production of both policy knowledge and regulation.

The rise of governance as a central attribute in infrastructure studies can be traced – to some extent – to a rise in the expectation of public accountability in publicly funded research. (Demerrit, 2000) As the speed at which basic science can be transformed into applications (particularly in data-intensive fields like GIS), there is a growing emphasis on presenting publicly funded research as a marketable investment. The cold-war mentality of competition and advancement of public science for its own sake is increasingly being replaced by a conception of basic research as part of the 'triple helix' (Eskowitz and Leydesdorff, 1997) of industry-government-university as partners in applying science. The triple-helix mode of science recognizes (in the mode of Third-Wave science studies) expertise and experience (Collins and Evans, 2002) as driving discovery even when it takes place outside of traditional academic and research institutions.

This bears a significant consequence on the category of 'relevant science', focusing on outcomes over process, funding research that can closely be tied to a research product rather than research already taking place, and encouraging researchers to approach funding and project selection as a market. This is a 'push-pull' mode of funding and determining relevant research, where researchers seeking funding are a supply-side 'push' responding to the industry/government 'pull' of demand for immediately applicable research outcomes. "Fetishizing the outcomes of research, in terms of new findings and results… favors those ways of practicing science most likely to generate immediate commercial and economic benefits and discounts other reasons for engaging in academic inquiry and conversation." (Demeritt, 2000, pg. 324) While Demeritt holds that this comprises a new social contract between funded science and the public, it is just as likely that this is a relocation of the 'pull', demand-side force of research funding. Rather than funding basic research in competition with the work being done by the Soviets, demand for research products has become more closely linked to economic outcomes – the gap between research and product is closer than it has been in the past.

Knowledge production in both science and law, according to Serres (1995), share a common interest in defining, dividing, and describing. Drawing on Herodotus' history of Egypt, Serres describes figures called the harpedonapti, who were responsible for marking the boundaries of tillable land following the annual flooding of the Nile river, which would destroy the prior markers of those boundaries. He claims this practice both as the practical origin of geometry and as the functional beginning of the law.

> "But, once again, since the flood erased the limits and markers of tillable fields, properties disappeared at the same time. Returning to the now chaotic terrain, the harpedonaptai redistribute them and thus give new birth to law, which had been erased. Law reappears at the same time as geometry; or rather, both are born along with the notion of limit, edge, and definition, with analytic thought. The definition of precise form implies properties: for geometry, those of the square or the parallelogram; for law, it implies the proprietor. Analytic thought takes root in the same word and the same operation, from which grow two branches, science and law." (Serres, 1995, pg. 52)

The production of knowledge begins with division, with establishing boundaries between what is claimed to be known compared to what is not so claimed, following with claims about what can and cannot be known. The production of knowledge acts upon its own origin, redrawing its boundaries and divisions as it operates. But more than that, knowledge production is an attachment – elsewhere described by Serres as parasitic (1988)– to the object being studied, described, ordered and the tools by which it is done. The harpedonaptai would work by attaching a cord to a stake, and using that cord as both boundary and instrument when surveying. This cord is significant to Serres, stating that the harpedonaptai's "mysterious title can be broken down into two words, a noun expressing the bond and a verb denoting his act of attaching it.“ (ibid.) That cord is both boundary and attachment – in surveying, dividing and describing the land the surveyors attach themselves to the object of study just as surely as they stake out its boundary, and be themselves characterized and described by the technologies of that practice. The act of stretching the cord is meaningless without a field upon which to work, and the geometries practiced by the harpedonaptai are sensible primarily in terms of that field. As encoded law grapples with the science it enables it reverts, in some small way, to the bounding work of the surveyor, seeking first to understand by dividing, describing, limiting and establishing an edge – and in this way constituting not only the field of study, but also bounding itself by the tools through which this was done.

Policy formation and enforcement has a tendency to be driven by broad issues rather than specific applications, and there is to some extent a mismatch between the general concerns of policymakers and the specific research and knowledge needed to address those concerns. (Sutherland et al. 2006) As such, relevance in research is decreasingly serial addresses of scientists towards policymakers and increasingly a two-way co-construction where specific research projects are aligned towards broad political goals. (Norton, 2005) Environmental science, for example, is not driven solely by neutral political goals, nor science for its own sake, but rather is heavily influenced by scientific representations of nature and the research culture and social goals of individual scientists in response to broad political goals and specific applications. (Demeritt, 2001) Zimmerer and Bassett hold that the environment from the perspective of political ecology is increasingly “a stage or arena in which struggles over resource access and control take place,” (2003, pg. 3) rather than a present actor in its own right. In similar way, the Earth Systems Scientists I interviewed and observed during my time working with EarthCube interacted with policy writ large as a site of research or a stage for competition rather than a partner in the conceptualization of relevance. Researchers approached calls for funding as opportunities for resource acquisition rather than the 'pull' of demand implied by the triple helix of research funding. Policy knowledge was framed as something of an issue of expertise, of 'game playing,' of knowing the terms, outcomes and individuals that result in the acquisition of funding. From their perspective, the 'others' of policy and industry were not partners in the construction of relevant scientific knowledge, but instead deposits of funding awaiting appropriate exploitation. Work done as funded research rarely aligned well with the goals of the call for proposals, and bore at best a nodding relationship with the outcomes and applications promised in their submitted proposals. While in some sense the triple helix is growing to dominate the expectations of research, the outcomes of research are still tailored to the social and organizational goals of the researchers and society at large – a concept called either the quadruple helix, or Mode 3 science. (Carayannis and Campbell, 2009) Making use of research towards resolving a political problem introduces issues of disciplinarity – often the 'most relevant' discipline to a problem is incapable of providing a solution to issues that stubbornly refuse to be solved within the expertise of that discipline. Issues such as land and water management with significant socio-political components are researched as purely physical/ecological phenomena, simultaneously privileging environmental modeling as the sole determiner of environmental management while ignoring or even replicating the social conditions that gave rise to the policy 'problem' in the first place. (Budds, 2009)

Relevance, then, is not ‘internal’ to science as it is characterized in the models of theory development and evolution presented by Kuhn. (1962) Relevance is a product of social alignments – in terms of the formation of conceptual relevance as well as in driving policy agendas and support. This is exactly the means by which science funders are interested in infrastructure as basic research. The technologies and methods are part of the policy process and actually act on behalf of governance – the science is in many ways the body of the government in the world, it sees, reacts, manipulates, and imagines on behalf of and in response to a larger political agenda. This happens deliberately in a relatively confined space like the Chesapeake watershed, but I propose is a part of projects at the national and theoretical scale. Earthcube management shows a similar process of policy involvement, relevance, support and response to certain technologies and methods in a particular scientific space, and at differing scales, as well. It is important to note here from a science of science policy stance that the essential value of basic research need not be quantified or even accounted a particular economic value, but that investment in basic research is a real-world enacting of governance goals that has non-obvious ties to policy outcomes that can’t easily respond to measurement. Not just ‘for the greater good’ but in pursuit of socially-defined, politically-motivated goals for humans and human behavior. This is evident in the debate around and management of the TMDL as both an object of regulation and scientific action.

Feedback Loops and the NAS Report

And so, having looked a bit into a prior assessment of the TMDL as directed by legislation, funding and regulation, we turn to another mechanism by which the feedback-loop policy dynamic is enacted: the active solicitation of evaluation from government agencies. In the year 2000, the EPA discontinued the use of the TMDL number in watershed management awaiting the result of a scientific assessment of the process following a period of intense scrutiny and evaluation in Congress. (Copeland, 2014) While the TMDL standards were initially enacted in 1972, they saw little actual enforcement on the part of the states due in part to a lack of a mandate by the EPA but also due to the sheer scale of the effort. Under the 1972 law, states were required to develop a list of threatened watersheds based on the TMDL of a particular nutrient or pollutant as well as a plan for managing that waterway. This represented a significant knowledge production activity as most waterways did not have ongoing monitoring, sensors, or active research of the sort that would generate this information. Few states developed full TMDLs, and many among those were simple lists of point sources of pollution and their relevant permits. (Houck, 1999) Throughout the 1990s the EPA attempted to expand the TMDL program, issuing requirements that states produce a list of threatened waterways in 1992, issuing guidance to states requiring the states to develop long-term plans for implementation of the TMDL in 1997, and a rule revision in 1999 that prompted significant debate. The 1999 rules would formalize “new requirement[s] for a more comprehensive list of impaired and threatened waterbodies; a new requirement that states, territories and authorized Indian tribes establish and submit schedules for establishing TMDLs; a new requirement that the listing methodologies be more specific, subject to public review, and submitted to EPA; clarification that TMDLs include 10 specific elements; a new requirement for a TMDL implementation plan” (Copeland, 2014) By the time the final rule was signed there had been 13 Congressional hearings, multiple budgetary actions delaying its implement, and variations in support across different presidential administrations. This eventually resulted in the withdrawal of the 1999 rules by the EPA in favor of maintaining the 1992 requirements. (Copeland, 2014) During this process of debate, however, there was a request from Congress that the National Academy of Science evaluate the scientific basis of the TMDL program and provide an assessment as to how it might be more effectively implemented. This report was hastened due to the perceived urgency of its content, and was completed in 2001.

The 2001 NAS report not only acknowledged the presence of scientific uncertainty and the need to interpret TMDLs through statistical models, but also included assessments on the availability of data as presented in a GAO report completed the prior year, which found that only 3 states had sufficient water quality data in order to properly follow the TMDL plan. (Malone, 2002) In addition, at the core of debate around the TMDL program comes from two sources: the Clean Water Act only provides for enforcement of point sources of pollution from particular effluents, while simultaneously requiring (as of the 1990s) that the TMDL measurements apply to ambient water quality, rather than water quality at particular sources. “The 303d focus on ambient water quality standards has returned the nation to a water quality program that was not considered implementable 35 years ago when there was a paucity of data and analytical tools for determining causes of impairment and assigning responsibility to various sources.” (National Research Council, 2001, pg. 16) This lead to a perceived injustice on the part of those responsible for certain points sources of pollution in the waterways – while the TMDL’s focus on ambient water accounts for non-point sources of pollution, the only remediation available to the states is in the regulation of points sources. (Malone, 2002) This, or so it is claimed, puts an unfair burden on those industries responsible for point sources of pollution while limiting the accountability of those responsible for non-point pollution. It also would appear to encourage a 'trading game' where certain points might trade outputs of particular pollutants in order to meet standards across the waterway. "Many waterbody stressors currently lie outside the CWA regulatory framework, where the only federal enforcement tool available is point source discharge limits... Perceptions of the inequity and the ineffectiveness of such a requirement may be manifested as technical critiques of the TMDL analysis itself." (National Research Council, 2001, pg. 100)

The growing requirements for measurement of ambient water quality and the focus on watersheds rather than particular waterways is a reflection of the incorporation of particular scientific techniques and technologies into the institutional knowledge employed by policymakers. While 35 years ago this plan would have been impossible, the introduction of sensor networks and growing acceptance of statistical models that can assess TMDLs across watersheds on the basis of sampling not only enables this particular implementation of a regulatory framework, but also becomes encoded into the understanding of how the waterways might be legislated. "Models are a required element of developing TMDLs because water quality standards are probabilistic in nature. However, although models can aid in the decision-making process, they do not eliminate the need for informed decision-making." (National Research Council, 2001) The NRC report distinguishes between simple mechanistic models, complex process models, and stochastic models, favoring the implementation of the simple mechanistic models due to their lower data requirement and closer relationship to empiric data. The particulars of the discussion around these models is interesting, but less important to my overall point in this section: policymakers making implementation decisions based not only on the presentation of scientific results but on an analysis of the scientific process, with reporting and attention paid to the availability and accessibility of differing forms of data (Malone, 2002) as well as recommendations and evaluation of the system on the basis of the cost of data acquisition and model development. (National Research Council, 2001)

In comparison to the early history of the Chesapeake Bay's regulation, we see a much more scientifically-literate approach to the evaluation of both the regulation itself and the particular mechanism of the TMDL's capacity for achieving that regulatory goal. *Prediction* is a consistently vital quantity in terms of how policymakers and enforcing bureaucrats accept, encode, and respond to scientific recommendations. The models described in the NAS report are predictive models, and while Brooks' recommendations were assumed to have failed on the basis of their lack of predictive accuracy, (Keiner, 1998) the NAS report acknowledges both scientific uncertainty and the presence of error in their models. Ideal model selection, in this report, is in part based on effectively representing uncertainty in addition to flexibility, low cost, consistency with available data, appropriate complexity, consistency with modern scientific theory, and a focus on the water quality standard. (National Research Council, 2001, pg. 72)

In addition to embodying the institutional knowledge production process within governance by acknowledging not only modern technology and scientific capacity, we see in the National Research Council report multiple representations of iterative, adaptive design enabled by feedback and monitoring loops. (Figure 4)

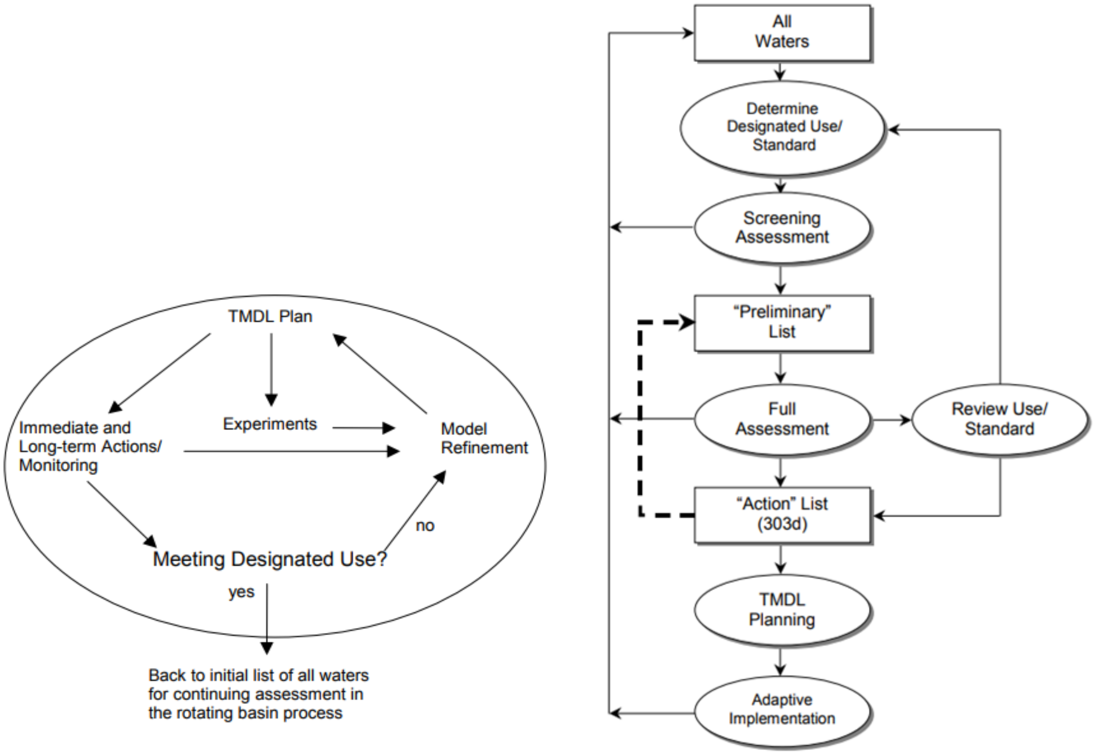


FIGURE ES-2 Adaptive implementation flowchart.

FIGURE ES-1 Framework for water quality management.

Figure 4: Flow charts describing the TMDL process. (National Research Council, 2001)

These loops include evaluation, adaptation, and cycles of feedback collection either through automated monitoring or direct evaluation of uses and standards. A similar depiction of the TMDL process from the EPA's website more directly represents the role of scientists in this policy process (Figure 5)

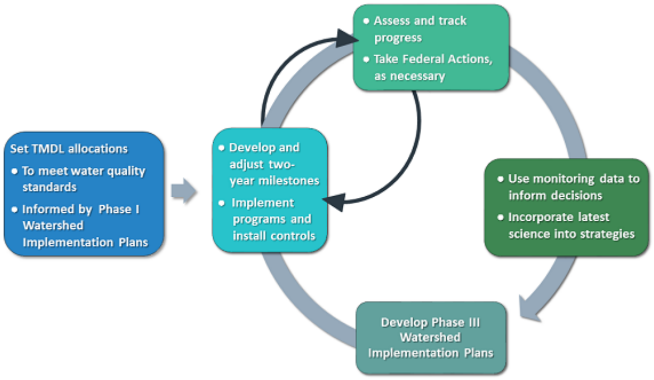


Figure 5: EPA Depiction of TMDL Process

Note that in the EPA's model of the process not only the recommendations and results of science but also its progress have a fundamental role. And while the phrase 'use monitoring data to inform decisions' is likely hiding a complexity of discourse, argument, stakeholder evaluation and policy framing, the necessity of responsiveness to 'latest science' should not be ignored. I would argue that this reflects an increasingly-vital role of active scientists in understanding the world for the benefit of policy production and modification, particularly when compared to early Chesapeake water management techniques. While the commissioned report was somewhat exterior to the policy process, with science at a remove, the NAS report (and even the public-facing accounts from the EPA) acknowledge a significant ongoing role of research and science in the active monitoring of the environment and forming a vital part of the regulatory feedback loop. In addition to understanding the world in the mode of Brooks' research on oyster propagation informing future laws and regulation, policymakers are increasingly producing assessments of the science itself. Where in the case of the Early Chesapeake implementations of regulation the assessment of the science was somewhat implicit (it is more an argument on Keiner's part that the lack of predictive power undermined Brooks than an apparent explicit statement), the combination of the GAO report on data availability with the NAS report on the effectiveness of modern science to support a particular regulatory object demonstrate an expansion of the knowledge production regimes among policymakers.

In my interview with a hydrologist attached to Bcube (and interested in EarthCube as a whole), we discussed for some time his role in the policy process, where he served in an advisory role on panels, as an 'on-call' resource for policymakers and lawyers when working through the regulation of the watershed, and as an operator of the instruments by which the watershed is understood and monitored. We spoke for some time during that conversation about the policy-oriented requirement that the TMDL be monitored through "proven methods, not black magic science." (interview, January 2012) The concept of the proven method versus black magic science is at the core of how different states interpret and apply the TMDL, and bear significant relevance on the process of modifying, changing, or reinterpreting the TMDL as a regulation. In order to grapple with non-point sources of pollution such as fertilizers used in adjacent farmland, groundwater contamination originating from local industry, or even natural phenomena, hydrologists employ a series of models that work together to predict and understand how that non-point contamination might be limited. The individual I spoke to was at the time working on the problem of denitrification of the stream: designing, testing and predicting the effectiveness of interventions like biofiltration, which uses a bioreactor – essentially a layer of microorganisms that can work to degrade or capture pollutants – to remove nitrates from the water. In order to perform that work he combined the use of 8-10 models, primarily selected on the basis of their resolution, time scales, and spatial reach, with data drawn real-time from internet-connected in-stream sensors, laboratory analysis of collected samples and occasional field experimentation to simulate flood events, runoff events, and other unusual circumstances not yet accounted for within his models. While his general methods were characteristic of the field, the particular needs of his interventions produced a heterogeneous set of methods specific to his lab and work.

This is part of the challenge of the concept of the proven method – being able, from the perspective of regulators and legislators, to conform to the mandate to employ proven methods in determining which rivers were at risk or in need of intervention according to their model-derived TMDL requires methodological and epistemological engagement with the science as it happens on the ground. In the absence of close, long-term engagement with scientists working in that field – E. said that in addition to his work on advisory panels he was usually called several times a month for clarification or additional advice - E.'s methods could very easily appear to be the 'black magic science' that lacks the weight of a proven method. In some ways it seems that the proof of a proven method is in part a function of position. In an interview with a different researcher approaching hydrology problems from the perspective of computer science, F., F. told me, somewhat jokingly, that "everybody trusts data except for the person who collected it, and nobody trusts a model except for the person who devised it." (interview performed 12/2013) Other than through comparison with other models and through broad adoption, there is little inherent to the method or individual that would indicate whether the methods used are proven methods or black magic. So much as I, when doing my ethnographic work, was somewhat reliant on the testimony of my informants, so too is the regulation and management of watershed pollution reliant on the testimony of their own particular engaged scientists. When the time comes to modify the TMDL model, or in the event of significant innovation in method, process, or technology, there exist pathways by which policymakers are able to produce knowledge about the state of the art in that science, and have already positioned themselves closely to advisors who serve to translate the particular jargon and epistemology of their field towards its evaluation.

Funding science becomes not just about understanding the world, but also about understanding the scientists themselves. This feeds directly into the concept of funding cyberinfrastructural projects like GEON and EarthCube – not only do these projects potentially provide the basis for science that might effectively inform the production of new policy, they also provide a window into the community of scientists. As I will demonstrate in the next chapter, the early EarthCube design process was heavily focused on constituting a community of scientists and understanding not only the state of their research but also their data and technology needs. There is a distinct move on the part of government funders and legislative offices to more closely engage not only with scientists themselves in the form of advisors, report writers, and so on, but also to engage with the concerns of those scientists and understand the barriers and enablers of their ongoing work. In this sense, funding infrastructure is the most logical way of achieving the goals of policy relative to science – it provides not only an understanding of scientific communities and the state of the art but also a means by which the availability, reliability, and progress of particular measures of the environment and society might be assessed and incorporated into policy outcomes.

Regulation as Cybernetics and EarthCube as Institutional Learning

In the above sections, environmental regulation is seen to engage not only with scientists but with science itself – the technologies, data and capabilities of science become encoded into law as greater understanding of the work becomes linked to a greater conception of what can be done in the world. (Jasanoff, 2004) This is not just a decontextualization of the knowledge for the purpose of incorporating it into a legislative or regulatory framework, it is a process of mutual adjustment by which the needs, capabilities and priorities discursively and materially present in policy work are fitted to a certain picture of the scientific state of the art. At neither end of this process is there an attempt to capture or present epistemological truth, rather it is finding a means by which the object of scientific inquiry may be rendered an object amenable simultaneously to scientific study and policy action. Oyster populations were rendered amenable to policy action as an understanding of how they reproduce and the ideal environment for their growth were more well-understood. Through the mechanism of directly engaging scientists in commissioned reports and requests for recommendations policymakers rendered the broader system of oyster health more amenable to policy action while simultaneously developing a justification for unpopular decisions that might have to be made. However, within that process of rendering amenable we see the production of a number of abstractions of the natural world, and more particularly, of the science performed. The NAS report discussed above relates the difference between mechanistic models and statistical models to reasoning about the height and weight of a population of students.

> "Unfortunately, the scale breaks after the first several children have been weighed. In order to proceed with the lesson (though on a somewhat different tack), a mechanistically inclined teacher might decide to use textbook data on the density of the human body, together with a variety of length measurements of each child (e.g.,

> waist, leg, and arm dimensions), to estimate body volumes as the sum of the volumes of body parts. The teacher may then obtain the weights of the students as the product of density and volume. A statistically inclined teacher, on the other hand, might simply use the data obtained for the first several children in a regression model of weight on height that could then be used to predict the weights of the other students based on their height. *The accuracy and utility of each of these two approaches depend on both the details of the input data and the calculation procedures*. If the mechanistic teacher has good information on tissue densities, for example, and has the time to make many length measurements, the results may be quite good. Conversely, the statistical approach may yield quite acceptable results at a fraction of the mechanistic effort if enough children had been weighed before the scale broke, and if those children were approximately representative of the whole class in terms of body build." (National Research Council, 2001, pg. 70, emphasis added)

This abstraction of the modeling process used on waterways provides both an analogous way of understanding the activities of scientists as well as the means by which they reach their conclusions and the data necessary to do so. Because while these policy outcomes to some extent demonstrate an engagement with the means of scientific knowledge production, it remains that policymakers in general are not trained in particular scientific applications, nor do they likely have the means to effectively evaluate those applications without additional information. The presence of reports is just one part of this – it is important to note that the knowledge production process in terms of policy production is not limited to scientific research out of the academy, but also the assertions of industry groups, NGOs and socially motivated individuals. The science behind oyster propagation was not in dialogue with competing science, but instead with the assertions of individual rights on the part of oyster fishermen and the interests of the emergent industry of canning and use of oyster shells in salable goods. In similar mode, the TMDL as a scientific object is not in contention, policy-wise, with other models of watershed health, but instead with the capacity and inclination of individual states to dedicate resources towards data collection, assessment of waterways, and the development of implementation plans.

It is tempting to deride broad abstractions of scientific concepts as fundamentally misleading, and in many cases such abstractions could fail to adequately account for the complexity of a particular situation. However, while the necessities of policy, and regulation in particular, have been shown to be consistently in the business of accumulating knowledge, be it about the lifecycle of the oyster or the technologies, data, and statistical techniques used to reason about and predict watershed health, what apparently emerges is more of a bounded rationality with respect to science. (Simon, 1979) Rather than having the expectation of perfect knowledge, models, or unambiguous science, policymakers seem more to rely on satisficing conditions for their decision-making in the conception of Herb Simon. In Simon's discussion of the behavior of consumers, he pointed out a significant gap in extant economic models – that rarely were consumers willing or ready to dedicate time to acquiring complete knowledge of a situation. Instead of expecting a consumer to always purchase the least expensive product, for example, it instead makes more sense to expect that consumer to purchase the least expensive version that they have recently seen. Neither policymakers nor individuals make decisions on the basis of perfect knowledge, and the production of useful abstractions of processes, debates, and other scientific dynamics serves to satisfice decision-making.

However, the evaluation of whether there is sufficient knowledge to produce a decision is not a one-off suggestion. Evidenced in the feedback loops depicted above and in the larger context of iterative investigation followed by assessment and adjustment is a process by which knowledge is fitted to a particular political agenda. Philosophically this is an almost cybernetic loop in the sense of the human use of human beings. (Wiener, 1988) Regulatory frameworks and monitoring cycles are built and undergo constant adjustment not just on the basis of political goals but also on the basis of the evolution of scientific knowledge. This new scientific knowledge is incorporated into policy frames or regulatory regimes as it becomes evident that existing knowledge is insufficient to meeting their current goals. New objects, techniques, and methods of study are introduced from a variety of places and adapted to local use. While the modern, model-derived, watershed-scale TMDL accounting for wide varieties of different pollutants and accounting for non-point sources of pollution may appear completely foreign to those who drafted the original legislation in the 1970s, it is the end result of a long process of assessment, adjustment, and mutual fitting of scientific capacity to the needs of policymakers.

Expansion of government is often accompanied by the expansion of knowledge production regimes – a classic example of this is Domesday Book, or Great Survey, which not only provided a demographic and value-oriented assessment of the country of England, but also enabled the assertion of Crown rights by William the Conqueror. (Poole, 1993) Not only that, the representation and accounting of value across the country provided a means for William to levy resources as needed – it represented a significant increase in the capacity of his government to act. (Poole, 1993) These knowledge producing regimes are somewhat self-contained loops –not generally the exploitation of knowledge produced elsewhere but instead knowledge structured and produced towards particular sociopolitical goals and realities. Sensitivity to issues of racial disharmony produces more categories of race such as the switch from Hispanic falling under ‘white’ in the US census to being its own racial category, the inclusion of mixed-race options and allowing for selection of multiple racial identities, and a bit more frighteningly in the Apartheid production of race categories as a tool of oppression and segregation. (Bowker and Star, 2000) These descriptive regimes not only allow new forms of governance and new objects which might be the subject of new policy but also exercise significant social power in their own right – Apartheid racial politics is fairly obvious but other categories and classifications of activity and work produce new categories of actor in their own right. (Bowker and Star, 2000)

The narrative of policy funding science claims to some extent or the other to be for the greater good, in the interest in establishing a competitive advantage relative to other countries, and so that there is a regime of subjects about which we know enough to render them amenable to governance. As a side product of the operation of science (and among the primary products of a significant portion of the policy process) is in the creation of a manageable, predictable, bounded subject about which policy can be written. Scientific work produces a discourse about nature that works simultaneously to enable scientific work and the work of management. This is the dynamic of co-production as discussed by Jasanoff – a mutual, intentional, and directed fitting process where scientists in policy-relevant work make a case for their work in part in an appeal the field of science and in part in its applicability to government operation.

This is part of why the NSF funds infrastructure as basic science, and why there is a policy interest in pushing the boundaries of the material capability of science. Not really why but how the planet is rendered governable is in part through investment in basic science. This works in three major ways (and a few minor ways): economic production through the spread of innovation (NASA is a big one, but computing technology is another), expansion of imaginative capacity in the sense where what we think we are capable of doing in the world is substantively a product of what we know about the world, (Jasanoff, 2004) and producing cybernetic-like regimes of monitoring and adjustment of relatively well-understood natural systems. Minor ways in which planetary systems (by which I mean the human-directed control of natural systems) is enabled through basic scientific investment include biome control (vaccination, disease prevention), evaluation of human responses to natural events either predicted or actively monitored, production of new cultivars and species through direct genetic intervention or interbreeding, and the ‘unnatural’ propagation of particular species more suited to a given set of human uses.

The Material and the Infrastructure: Relevance, Single Numbers, and other tools for Decontextualizing Knowledge

The end result of the decontextualizing efforts of translational work give us a particularly interesting tool for understanding how scientific concepts become political goods: what I call the discontext bit, which at times works to both obscure and reveal the condition of a certain politically relevant system. For example, atmospheric carbon has become a single-number ‘point of contact’ in popular discussion, news reporting and environmental policy. Similarly, so has the average global temperature. Alone (decontextualized), these numbers convey very little information about environmental health, but have come to stand in politically for substantial scientific work. Such abstract single numbers, though, move easily between the concerns of policy and the concerns of scientists – they are ontologically mobile, relevant to the conditions present in both spheres of action without substantial recontextualization. Such terms and numbers (and previously, but a bit more of a reach, the size of the hole in the ozone) move back into science, shedding political contextualization, in the form of a greater focus on atmospheric conditions as a fundable object of study and an object of political action. Rip and Voss describe a similar entity, called an 'umbrella term' that mediates between the work of science and the political and social understandings of that science. (Rip and Voss, 2013) Umbrella terms, according to Rip and Voss provide a basis for innovation by providing for a *de facto* understanding of emergent scientific work that does not need to be referenced back to individual research efforts.

This is akin to the deletion of modalities, (Latour and Woolgar, 1979) but more specifically oriented towards a particular purpose, rather than the natural consequence of representing the results of science. Nano-technology, for example, is an umbrella term referenced by Rip and Voss as a unifying mediator "through which scientific problems travel and get entangled in the constructions of 'relevant science.'" (Rip and Voss 2013) Umbrella terms, like the single number representations of an issue, provide for easier movement of concepts between policy-makers, scientists and the ever-nebulous public, establish a point of collection that unifies a variety of distinct efforts, and work as a foundation around which lasting organizational realities form.

Immutable mobiles like the umbrella term and the single number representation serve organizational purposes without accounting for the infrastructures, both technological and organizational that develop around the construction of political and scientific relevance in the discourse on a certain object. This is significant because infrastructure, by and large, is just too big to address briefly in a way that accounts for the depths of its ecological, social, and technical commitments – like the proverbial iceberg we focus on the tip to the exclusion of the depth (often to our detriment). These, alongside the notion of infrastructure as imaginary, work in similar ways: they allow the discussion and evaluation of a notion without needing the whole context and content, in short, they let us talk about things too complex to treat with briefly. They work to strategically occlude and expose aspects of discourse and advocacy, with a particular focus on portable numbers representing larger systems - discontext bits - as tools that enable regulation and advocacy. While the triple helix seems to introduce new categories of actors in research, in effect it blurs the line between insiders and outsiders in terms of basic research communities. Where earlier I discussed how knowledge work for the purpose of policy production and enforcement is closely linked to knowledge work in the sciences to the point of being essentially the same activity under differing epistemological regimes, there is a similar blurring of industry and academy in terms of basic knowledge production and market-driven relevance of science. The origins of watershed management, at least upon the Chesapeake, were couched in terms of the needs of industry. Arising from the desire to protect the oyster population of the Chesapeake Bay while still allowing for shipping traffic, knowledge about the Chesapeake watershed was simultaneously enacted upon the water itself as well as those with a stake in it. The government produces knowledge about the science it funds towards its goal of protecting the cleanliness of the water as well as producing knowledge about its understanding of the industry that affects it. In similar way the knowledge produced by scientists enabling that management is described through the terms of that resultant policy, and is often directed by the goals and requirements stated therein. The science that enables the management of that water engages the triple helix not only as participants in the process but also as objects to be understood. Industry learns about law to understand the limits of its activity, the law observes industry's operation to weigh, balance, and describe the uses of shared resources, while scientists describe its operation in terms of the needs of policy production, with an eye to the presence of industry as acting upon the water as part of its ecology. So not only do the three arms of the helix work upon the same object, they produce knowledge about each other in order to do so.

What is evident in this activity is not only the way that increasingly diverse communities become enrolled in the knowledge production process, though often at significant remove from research activities in that area, but also the way that particular regulatory goods might be motivated towards research. The TMDL is a strong example of the way a discontext bit might work to bridge varied interests and activities towards shared goals at a significant remove. Not only is the TMDL a means by which regulators and legislators might effectively assess water quality and prioritize rehabilitation efforts, it also provides a guide for industry when evaluating potential point sources of water pollution resulting from their activity. Finally, the TMDL has emerged not only as regulatory leverage but also as an object of study in its own right. Scientists develop models specifically to address the TMDL, (Shenk and Linker, 2013; Linker et al., 2013) to identify which models might effectively apply to TMDL uses, (Borah and Bera, 2004) to apply participatory methods to data collection in order to support said models, (Voinov and Gaddis, 2008) to effectively evaluate how best to employ this data towards implementation (Wainger et al., 2013; Wainger, 2012; Jones, 2014) and more generally working to improve the accuracy, completeness, or effectiveness of the TMDL for both modeling and policy compliance. (Borusk, Stow and Reckhow, 2002; Borah et al., 2006; Reckhow, 2003; Kang, et al., 2006; Shirmohammadi et al., 2006; Williams, et al. 2017) The citations presented here merely scratch the surface of work being done on the TMDL and its attendant applications, uses, attributes and analyses – not only is the TMDL a site of significant policy work, it remains a site of significant scientific attention as well. As a discontext bit, the simple numbers of the TMDL leverage, disclose, and abstract the operation of complex knowledge production practices from model development to sensor deployment to the establishment and evaluation of goals set for water quality and use.

So not only does the concept of the TMDL serve as a leveraging point for environmental conservation on the part of policymakers, as a means by which water quality might be assessed and rehabilitation prioritized, as a model-derived figure representing and enabling the better understanding of watersheds, and as the object of scientific work relevant to that understanding, it is also the enrollment point of various disparate actors into water quality management. While limitations on enforcement capacity may present some perceived unfairness in terms of financial responsibility to the TMDL, its existence as point of contact for overall water policy? allows it to move between contexts, applications, and areas of interest relatively freely. While each interested group may treat the TMDL somewhat differently, or contest particular outcomes, the TMDL has become infrastructural to the work of water conservation – while it has been reconsidered at points in the past and is still the site of significant work, the discontext bit of the TMDL need not be reconsidered at each moment of applications – its use in water quality assessment and rehabilitation might be improved, but is rarely discarded entirely.

Conclusion: governments and infrastructure.

Government is interested in knowledge infrastructure not just from the service provision perspective, but also from the perspective of being able to account for and grapple with the new social and technological realities it produces. Infrastructural changes and technology changes required a response in the Chesapeake Bay, and the process of legislating and monitoring watersheds beginning with the Chesapeake and moving outwards has produced an infrastructure of knowledge about those watersheds in its own right. Government is by its nature well-suited to the marginal work of infrastructure, but needs to continuously be sensitized to the cost of maintenance and repair. Ongoing work on the TMDL has been unavoidably expensive, with many states reporting a financial or material inability to meet its requirements. Accounting for maintenance and repair of very large or complex infrastructure is itself a significant analytic task – and one that is not commonly performed in the design process of new infrastructure.

Guiding infrastructural development is a tool for monitoring as well as designing – it hits multiple parts of the regulatory cycle simultaneously. As governance grows more reliant on scientific knowledge it also grows more complex, with basic science taking on new meanings as investment in infrastructure increases. In the following chapters, I will return to my field work on EarthCube, showing how these near-cyberntic cycles of assessment, modification, and adjustment have become a fundamental part of science investment in infrastructure and work to serve as means to understand and constitute scientific communities as well as provide for innovation in their results and expansion of their capacity.

CHAPTER 5: EarthCube Design, GBIF, and the Turn to Infrastructure

Infrastructures are pervasive but ill-bounded. Often the result of ad-hoc building and bootstrapping more than intentional, directed design, (Bowker, 2000) an account of any particular infrastructure by necessity enrolls accounts of similar, related, and interdependent infrastructures. A narrative of the development of landline telephone infrastructure, for example, also involves the preceding telegraph infrastructure as well as the power grid, highway system (which telephone lines often followed for development convenience), mining operations (nationwide telephony required a huge quantity of copper wiring), and the series of regulatory, funding and economic decisions that informed and enabled its monopolistic control by Bell until the 1980s. (Aufderheide, 1999) A similar account of the internet, which was initially implemented on the extant telephone infrastructure, would draw on all the prior infrastructures, the infrastructures of research funding, academia and particular history (cold war and nuclear fears) in addition to the specific sets of specialized knowledge and equipment that enabled early computing. Beyond this, new infrastructures begin in a landscape of not only extant physical infrastructures, but also infrastructures of policy, regulation and governance that have significant effects on not only the physical substrate of the infrastructure but also on its operation, adoption and use. This is often called the infrastructural stack – those sets of infrastructures built atop one another and their corresponding path dependencies (Barnes, Gartland and Stack, 2004) in behavior alongside the materiality of the infrastructure itself. Infrastructure is the inscription of our goals, values, and vision of the future. Broad societal change consistently (and interestingly) accompanies or is accompanies by similarly broad infrastructural change. (Van der Vleuten, 2004; McCarthy 1987) This is where the challenge of developing new infrastructure for scientific work lies: new infrastructure is accompanied by a retinue of changes in work practice, standards, organizational policy and even discourse. Modern simulation physics would be unrecognizable to a physicist from a few decades ago. (Galison 1992) The field researcher in earth systems now similarly has a counterpoint in the simulator, the modeler, and the system designer. Even social scientists work from a technological infrastructure that at one time was limited to the organizational structures of recognition and authority present in the academy. Our reality is diffuse, diverse and interconnected regimes of knowledge production and the possibility of rapid, technologically-driven change. Understanding how infrastructures form, not just as technological systems but as participants in the construction of sociopolitical reality provides a basis for managing, directing and provoking change.

A possible narrative of the development of modern infrastructure studies could trace its origin to the broad successes of distributed scientific collaboration during WWII and the recognition of the potential and impacts of basic research (especially in physics). The period of time following WWII is generally characterized as the point in time where big science began to emerge. (de Solla Price, 1986) Following the war, the scale of inquiry among theoretical and especially experimental physicists expanded substantially, and other areas of scientific study followed suit. (Galison 1992) Vannevar Bush's 1945 proposal, "Science: The Endless Frontier"—a proposal that contributed in part to the 1950 passage of the National Science Foundation Act—takes on the appearance, in hindsight, of a proto-cyberinfrastructural effort with the stated goal of capitalizing on existing scientific capital through investment in basic research and funding support. (Bush 1945) Bush's vision of interdisciplinarity and distant collaboration has become something of a clarion call for cyberinfrastructural development. Following the successes of ARPANET and the emergence of distant collaboration potential enabled by internet and other new technologies was a growing desire to capitalize on new technology in scientific space. Core pathways for achieving this in the 1980s and 1990s were the funding of collaboratories (Bos, et al. 2007; Kouzes, Myers and Wulf 1996; Olson and Olson, 2000) and digital libraries (Van House, Bishop and Buttenfield 2003; Bishop et al. 2000), both of which focused on employing new technologies to bridge distance in scientific work.

Collaboratories sought to bring scientists together using new technology, focusing on telecommunication resources and virtual presence of one form or another (Olson and Olson, 2000), where digital libraries were designed to bring together scientific resources, samples and tools accessible. Modern cyberinfrastructural efforts tend to combine both of these goals, seeking not only to develop shared tools, resources and databases but also to use new technology to create new collaborations and support geographically diverse collaborative work. Central to these efforts, though, is the notion that policy-supported infrastructure development can enable the development of novel science, be it through new collaboration, through making previously heterogeneous datasets and tools interoperable, or by building new communities that can identify new interdisciplinary avenues of investigation. The model of the Manhattan Project, where distributed teams of collaborating scientists and engineers resulted in (for better or worse) world-changing science and design, would seem to support that notion. However, the reality of infrastructure development as a means of creating that world-changing science is somewhat more complex, more closely tied to existing organizational and social factors, and consistently more difficult to sustain following the end of active support.

Infrastructure, as defined by Star and Ruhleder (1994) is embedded and transparent; infrastructure exists (metaphorically) within or underneath other social, technological and built worlds and does not need to be reconsidered at the moment of each task it enables. Infrastructure is learned as a part of membership and linked with the conventions of practice therein, and embodies some set of standards. It is built over top an installed base, becomes visible upon breakdown, and is of a scale or scope that exceeds a single 'site' – however that might be conceived. It is important to keep in mind, though, that infrastructure as defined here is relational – something is infrastructural related to a certain use or set of uses. Star reminds us that what from a certain perspective is infrastructure is from another the object of regular, daily work in maintenance and upkeep. Here, a terminological point is necessary. I may, throughout my writing here, reference a number of terms related to infrastructure work as used historically by various authors, funders, etc. such as cyberinfrastructure (usually the term applied to recent efforts in producing internet-enabled infrastructure supporting scientific work), digital libraries (referring to infrastructure supporting resource and information sharing) knowledge infrastructure or even collaboratories. I use these terms in context but semi-interchangeably: one of my basic assumptions throughout this writing is that infrastructure shares an important set of attributes regardless of the specific term, purpose or time in which it was implemented. While digital libraries and collaboratories are to some extent historically distinct from cyberinfrastructure, e-infrastructure, knowledge infrastructure, they are all science-supporting infrastructural efforts sharing the basic attributes of infrastructure as initially formulated so comprehensively by Star and Ruhleder. (1996)

Regardless of how it is described, the design of new infrastructure is as much a social goal as it is a technological – commonly success is evaluated just as much on the emergence of social and organizational configurations as technological ones. (Bowker et al., 2010) Specific scientific problems, or 'big questions' are often phrased at the moment of development to ensure some contribution to the scientific domain intended to be served by the infrastructure, (Bowker and Ribes, 2008) but ultimately effective new infrastructure is evaluated on the existence of an emergent social order. An infrastructural effort's ultimate goal is to develop a certain relationship among its intended user groups to the infrastructure itself – that is, intensive data services only become infrastructural to knowledge work if the social realities of those *knowledge workers* shift in response. The inertia of existing work practice must be overcome: in order for an infrastructure project to be successful the behavior of humans in the systems needs to change.

Infrastructure, and particularly concerted infrastructural efforts, provides a basis for examining what may otherwise appear to be disconnected events as part of the same reality. As infrastructure resolves the tension of local and global, it can be seen both as a bridge and enabler of action at a variety of scales. (Star and Ruhleder, 1996) Examining infrastructure directly provides a narrative of change that is interested first and foremost with the invisible, with the marginalized, with that which silently contributes to differing conceptions of the possible, of the probable, and proposes a boundary circumscribing otherwise indistinct groups, regimes and organizations. Infrastructure is a relational proposition – it describes the relationship between certain activities, systems, and modes of work. When something is in the position of infrastructure, it does not need to be reconsidered at the point of action. This is not only a function of perception (though that provides an easy entrée to the relationship), but also of action and activity. Infrastructure is fundamentally performative in this sense. In a vacuum, without the filter of activity, infrastructure is indistinguishable from other artifacts. So to infrastructure (as a verb) is to act in a certain way, to operate on a certain set of assumptions, and fundamentally to make an ontological statement. Plumbing as an infrastructure is different from the pipes that make it up, but ontologically as I take my shower they are one and the same – the pieces that comprise the system are subsumed into the object of the infrastructure itself. I don't talk about chunks of pavement, electric lighting, and painted symbols, I talk about roadways and enroll the entirety of those things that go into a roadway as a single, infrastructural object.

All this is to call attention to a significant fact: even those infrastructures that seem simple, those infrastructures that we can 'get our head around', so to speak, are still complexly interwoven knots of material, standards, cultures and various scales of governmental and organizational policy. (Engestrom, 2005) Fundamental to infrastructural work, when looking out from the interface of science and policy, is just how to effectively represent the extent and complexity of the infrastructure in support of its design, maintenance and governance. Often, decisions about infrastructure are made by those who do not work directly on that infrastructure – especially troubling considering the whole class of marginalized work endemic to infrastructural work (Star and Ruhleder 1996; Bowker and Star, 2000) even before one considers the under-theorized work of maintenance and repair. (Jackson et al., 2011) Significant work among the EarthCube building blocks, and in particular the test governance committee, was spent creating goal-focused, intentional representations of the infrastructure and its components intended to be received by particular audiences. Decisions made about infrastructure, due to the remove of decision-makers from the specialized, particular knowledge of infrastructure most visible to those working directly on it, are more accurately decisions made on the basis of representations of that infrastructure. Effective representations are major factors influencing the design of a system, both internally to those working on it and externally where relevant. Accounting for the perceived needs of policymakers and regulation inherently reacts as well as speaks back – an excellent example of the dynamic of co-production in action (Jasanoff, 2004) – there is a desire to not only fit design work to important policy considerations but also to be seen directing that work towards a certain conception of relevance. The work should not just fit in, but also be seen to fit in.

This chapter will deal with EarthCube primarily as an object of design with a specific set of scientific, organizational and social goals. Throughout the long-term design process of EarthCube the imaginaries of its capability, effectiveness, and general 'look and feel' have responded to community development, NSF funding decisions, and the daily work of governance and planning. Despite the goal of community-led, middle-out design of the EarthCube system, the material realities of the project demonstrate a much more top-down decision making process. I will explore some of the disconnect between identity and different levels of engagement with the infrastructure as it develops, and attempt to generalize the EarthCube infrastructural processes in terms of their negotiations and interpenetration with other extant infrastructures and infrastructure projects. From very early on EarthCube as an organization has grappled with an identity- and funding-disconnect between work with geoscientific outcomes and the development of tools and technologies intended to produce them. At the end of the chapter I introduce a comparative example to EarthCube, the Global Biodiversity Information Facility (GBIF) where technology and design decisions initially fell a bit short of their claims and goals but where continued use and investment was enabled through legislative action and directed engagement. While GBIF is not a prescriptive guide for establishing an infrastructural good, it presents a pathway to finding the infrastructural turn, where a system moves (for some population) from an object of work to the object that enables work.

EarthCube

EarthCube, to some extent, is a natural extension of the development track of scientific infrastructure, and is an evocative illustration of several important factors in the interaction between scientists and policymakers in spheres even outside earth systems science. First and foremost is the way in which personal and professional identity determines both the content and context of publications, statements about and visualizations of scientific work. In general it makes sense to have an imaginary (though a relatively weak one in practice) that there is an interface between the 'worlds' of science and policy – this interface is formed from the basic strata of personal and professional identity. Addressing those outside of one's group (regardless to any actual differences in education or training) is both an act of recontextualization and gamesmanship. Both policymakers and scientists here work at least somewhat in the realm of modifying human behavior – each group has a stake in the game and assumes the other is sufficiently different in motivation, knowledge and basic goals that there needs to be some element of strategy and translation to their interactions. It is an almost cliché observation that scientific work as presented to funders looks somewhat different from that work in practice, but that observation (again, regardless of its level of truth to a given individual project) is an important reminder of that element of strategy – at the level of scientific work there is some basic assumption that the goals of scientific work and the goals of policy work are not the same. (Hajer, 2003) Whether this is a slippage in method, epistemology or in basic values, it remains that at the interface of scientific and political goals there is some assumption of disconnect, of the need for translation (both of scientific work and policy work) and a resultant attenuation of knowledge from the context of its creation. This attenuation of knowledge applies to both the outcomes of scientific work, regardless of the format (presentations, papers, posters, monographs, etc.) and the larger discourse about the nature of the work itself.

EarthCube is an infrastructure project intended to provide a unified architecture for access to data, tools, and other resources for earth systems scientists. As an architectural project EarthCube is not primarily focused on the collection of new data (though a couple of funded projects did collect new data), instead funding independent project elements intended to lead to an overall architecture. Technical building blocks (BBs) intended to be pieces of an eventual architecture, integrative activities (IAs) that bring building blocks and other technologies together, conceptual architectural grants and research coordination networks (RCNs) all to some extent facilitate the reuse of existing databases, data streams, tools and expertise. EarthCube, like other historical infrastructure projects, is a joint effort of the government (as represented in the National Science Foundation) and interested scientists. The overall agenda of the project as well as decisions on the funding of individual building blocks, which include both a governance group and various architecture proposals for linking other building block projects, was set by program officers in the NSF with advice from participating scientists. The infrastructure itself, however, is intended to be the result of a 'middle-out' design process where useful tools, techniques, and systems were allowed to develop relatively independently before being considered as a part of a larger overall infrastructure. As a collaborative design process drawing from both the political and scientific worlds, EarthCube enrolls imaginaries of governance and future science as well as being a microcosm of the co-constructive processes of relevance, both political and scientific.

> "The EarthCube is envisioned as a system where "cyberinfrastructure enables the geosciences". The EarthCube will allow users to be conductors, using a palette of resources, processes, and communication to compose their work, reducing time spent resolving tedious problems. Facilitating data storage, processing, retrieval, and transformation are all key elements of the system-to-be. Common, time-consuming processes like re-projection and file format conversion are well-understood problems that can be (and have been) solved by CI and should not obstruct science, research, and education. Common problems do not have to be solved repeatedly. EarthCube will also be a collaboration hub, allowing users to discover and share resources, information, and contacts. This allows users to announce their resources, interests, and work to other users. EarthCube will support and enable interactions between the geosciences, from terminology and units of measure to data format and metadata" (Braeckel et al., 2011)

There are a couple of metaphors employed in the white paper quoted above that are important to keep in mind when creating an account for the overall design philosophy and decision-making of EarthCube leadership, especially from the perspective of project design and resource allocation. The first, that remains somewhat present still in the discourse of EarthCube, is this notion of the user as conductor. Like a symphonic conductor, the ideal EarthCube user (in this particular bit of discussion) takes a high-level view and works to coordinate the instrument groups (tools, systems, data sources) in the composition of a complex piece of scientific work. The conductor, here, creates a symphony of science by organizing and employing relatively independent and heterogeneous tools and resources, and is in some way removed from those tools, standing above them as the conductor stands above the orchestra pit. As the conductor gives guidance to professional musicians with significant personal knowledge and skill in their own right, so too would the designed-for user direct and employ other technological and scientific resources without necessarily working directly on them. The above quote also references a particular goal of EarthCube, and one that is tied closely to the role of infrastructure in other areas – the ideal user of EarthCube in this document is one that does less technologically-focused work with more and more diverse resources to produce complex, novel science and create new interdisciplinary collaborations.

However, this does not really fit well with the science I actually observed in process. During my time with BCube I spoke to scientists working with data drawn from oceanic sensor networks about their tools and data. Rather than working as conductors employing relatively independent resources, the accounts of those scientists revealed a much closer relationship to their tools, data and methods than that envisioned in the conductor metaphor. The researcher working with these sensors was able to tell me not only the short history of each ocean sensor they had, but also which ones needed to be accounted for in a particular way during early data transformation and employment in their model. One particular set of sediment collectors and sensors employed by an oceanographer attached to the BCUbe project, for example, had an opening and filter slightly larger than the others employed in that area. This meant that somewhat larger organisms were occasionally caught within the sensor to die (which naturally threw off measurements of the microorganisms), and this needed to be accounted for in the early phases of data cleaning. "Different groups make different choices," (interview conducted 2/28/2013) and there is very little transparency as to how these different groups might account for the bias introduced by this particular group of sensors. "Changes in instruments produce false trends" and often go uncurated when published, "the worst thing you can do is just go to NODC [National Oceanographic Data Center] and just grab data." (interview conducted 2/28/2013) While the sensor was still effective, the potential for this particular error had to be known and corrected for prior to the data's inclusion in the model for validation or development. Uncurated or undercurated data sources tended not to produce a significant account of how this data might be biased through its instrumentation, and "bad data is worse than no data… you get crap." (interview conducted 2/28/2013) In order to effectively curate data sources prior to their inclusion such that they could be just picked up and put to use there would have needed to be data specialists and curators attached to the project from its early stages, but "[We] can't afford afford data specialists… I think nobody could." (interview conducted 2/28/2013) This is a much different picture than that the scientific conductor leveraging and calling upon diverse data sources and scientific resources for the creation of a symphony of novel science. Many of the scientists I spoke to were much more a tinker than a conductor – directly and closely engaged with the detailed workings of the entire lifecycle of their scientific work from sensor purchasing and deployment to data cleaning to model development and verification. The notion of the scientist as a removed, high-level conductor was one that just did not seem to play out. In fact, it seemed to work in quite the opposite way, where potential users seemed wary of employing data and resources that they didn't know very well. And this in fact undermines the process of science itself - "Its becoming very easy to get access to data and abuse it." (interview conducted 2/28/2013) The conductor could not account for the potential bias of instrument changes within a dataset because they are simply too far from the process, too removed from the in-group that shares, develops and uses that data, and not necessarily using the same vocabulary, variables, or even practices on how, for example, to aggregate and divide plankton and chemicals into groups prior to their description within the data.

EarthCube has had a central focus on novel science since very early in its inception. There is even some of that metaphor in the above quote as it characterizes the ideal user as a creative worker, one who works from a palette to create something new. Novel science, in the discourse of EarthCube, was limited and to some extent determined by the necessity of repeating 'solved science' that only really needed to be done once and made available. New work and collaboration is impeded by repeated negotiation of standards, (terminology, units of measure, data formats, metadata), and it is important that scientists have tools to avoid the tedious work of data management, re-projection, and transformation. As we explore the particular building blocks and design decisions supported in the early years of

EarthCube this discourse will return over and over again. The tools to solve problems already exist and only need interested scientists attached to them in order for novel, CI-enabled science to occur. A preponderance of tools is better than a lack because the scientist simply directs and employs those resources without the need for management or close work on the resources themselves.

EarthCube sought to enroll both systems and individuals as well as creating new tools. In large part, systems (data systems, other aggregation services) are enrolled through interaction with a human that mediates access to the system, and who is required to do some work in order to make greater access possible (presumably systems are already accessible by their target communities – it is broader, coordinated, orchestrated access that EarthCube seeks to enable). This requires something of a 'policy mandate' – somebody needs to do some work in order to scale. Issues of contribution, reward and participation become paramount to the account of a given infrastructural effort, and organizational issues tend to overshadow all others in describing a narrative of success, partial success, and failure. Organizational factors dominate the discussion, and, like much reification of organizational attitudes, have a tendency to leave such projects deeply resistant to change. Infrastructural development tends to operate on two cycles: those of amalgamation and fragmentation, where first work is done to draw together disparate localities of practice, metadata, reward, etc. through standardization, translation and tool development and the second is the adaptation of the outcomes of that drawing-together to suit local condition. (Hepsø, V., Monteiro, E., & Rolland, K. H., 2009) EarthCube, through the majority of my time on site, was engaged in the work of amalgamation, bringing together interested researchers in Research Coordination Networks, developing and modifying new tools through Building Blocks, imagining and directing coordination of tools, data sources, and people through Integrative Activities, and managing Data Facilities.

To some extent, major moments of cultural shift are accompanied by (or preceded by) the emergence of new forms of infrastructure. It would be difficult to argue against the notion that the emergence of the internet represents a substantially different way of engaging with the world, both in terms of how the world is understood as well as in the way that objects of culture and expression move around the world. Without near-instantaneous communication or world-scale observation networks it would be nearly impossible to assess systemically even local climate phenomena. Peter Galison argues that atomics and general relativity have a very close relationship to the practical problem of how to properly synchronize remote clocks. (Galison, 2000) Infrastructure is the story of what happens while the 'real story' is taking place – be that story a scientific or cultural revolution, or just a new way of interacting with or understanding the world. We can situate the concept of infrastructure within a long historical trajectory: the road and aqueduct systems of the Roman Empire, the Silk Road caravan routes (and compare the latest Chinese Belt and Band initiative), oases and supply stations, or even the regime of knowledge practices, representations and devices that enabled oceanic navigation. However, for the purposes of this discussion we are looking at the concept of cyberinfrastructure which, among other things, is associated with a relatively specific cultural and historical position. Cyberinfrastructure (and the more nuanced concept of knowledge infrastructure) is a 'here and now' concept – closely linked to computing, electronic data, and situated in a culture of scientific development where the need for new infrastructure is not only commonly recognized but a focus of work for significant scholarly communities. (Edwards et al., 2013) If computing and communication technology was relatively stable in capacity and effect, we would not be worrying about its design as infrastructure to other important work. If that other important work was not already recognized as important work, with a significant community recognizing the value both of the outcome and the method, there would be no desire to expand its capacity. Galison (1992) notes that simulation physicists were once pariahs in a community that saw little need for advanced computing in their research work. The science of botany has been similarly invaded by data science in a variety of forms. (Bowker, 2000) Hand-drawn depictions of holotypes and the judgement of the researcher in the identification of plant specimens have been slowly taken over by digital depictions and genetic analysis. The botanist in the field with a paper notebook and colored pencils is an increasingly quaint notion.

The sort of cyberinfrastructural efforts I am speaking of here presupposes a relationship between science and government where there is a desire to encourage multi-sited, widely collaborative science enabled by advancing communicative and computing resources. There is an underlying notion that the progress of science is part of public good, (Callon, 1994) and therefore should be supported by the public (in the form of government disbursement of tax dollars) that almost inevitably lends a political shade to that progress. Beyond that, the agenda of government offices and representatives, be they bureaucratic employees of a particular agency or publicly voted legislators, bears significant weight in the determination of relevance of particular scientific efforts. In EarthCube, the direction of the NSF was particularly evident in the selection of funded projects and in the attentiveness and responsiveness of funded PIs. While the project was largely described as community-led and 'middle-out' in its design, there remained at least the spectre of NSF authority. As of the time of this writing, it has become quite evident that while the NSF has been collecting community input, its officers have the final say on the direction and design of EarthCube:

> "NSF is pushing hard to move forward to implement *something* so that they have something to show for their last 5 years of funding this effort, and therefore the time for community input in this round has passed, but there will probably be a feedback period after something has been implemented." (Minutes taken by Lynne Schreiber, Science Committee 10/2/2016 meeting)

This move reads as somewhat in conflict with the stated middle-out, community-led design model that EarthCube was supposed to instantiate. Organizational pressure within the NSF has overridden the will of the EarthCube community such that there is indeed a moment where the time for community input has passed, with the notion that at some indefinite point in the future participants in the EarthCube design process may someday be allowed to express their feedback. Rather than community-led development, the desire to demonstrate a positive, material outcome of the design process has led to a moment of enforced stability in the object and imaginary of the EarthCube system – now those researchers, technologists, data professionals and scientists that had been imaging the architecture of EarthCube in design charettes, in end-user workshops, in funded activities and group brainstorming, are bereft of their designed object, and told to work from this point of new, enforced stability. The fact that this initial material architecture (which will be discussed in more detail in a later chapter) has little resemblance to the architecture and infrastructure imagined by those participants is demonstrative of the nature of the relationship between the expected audience for the infrastructure and its funders – and is problematic to the notion of community-led development.

The EarthCube Design Process

There is, then, no single positive trajectory from 'invention' to 'innovation'– rather, the designed object changes in the process of development – it becomes very real sometimes, goes back to pretty unreal, bounces back to somewhat real, becomes definitely unreal, becomes almost real in any sequence.(Latour, 1996) Even when by the end of the book it is declared 'dead' (the conceit of the book is a murder mystery – who killed Aramis?), Latour questions whether it's 'really dead' and what real death would look like. Following the design process forward in real time is about recognizing how the social, the technical and the material get reconfirgured in the process of design. (cf. Stenner, 2007 on Whitehead's process ontology) In brief, the ontological unit of design analysis is not the design object itself but the sociotechnical 'design black box'. What you analyze is not the birth of an object conceived in the designer's head but a messy sociotechnical trajectory whose outcome (if it succeeds) is a new society with new institutions and a new object.

The language of performative closures can be used to describe points along the pathway of ontological variability. These can be defined as points in the design process where local maxima in the fitness landscape (using this term ecologically) are achieved – they are points of relative reality from which the next maximum will be achieved. This will never be a final closure: design is always a continuing process, particularly so in the infrastructural space. These local closures would then be considered as points along the way where you take stock and ask: “OK, what kind of a black-box do we have now?”. This can be externally forced (NSF asks me for a report on progress of a technical project, so I produce a demo – which can be conceived as a visualization of the finished blackbox) or internally imposed (let’s work out where we really are now and what are the next set of paths forward). Each act of performative closure is simultaneously a point of ontological reckoning – how real is our project now?

Looking at these points of ontological variability in the design process shows not just where iteration is occurring but how iteration conceptually changes designed objects. Seeds from previous iterations are able to open sometime in the future, or restructure the project retroactively as they become important. At each identified moment of performative closure there is both a conceptual unfolding as stabilizing certain characteristics of the design allows for variation in others and a folding-in of varied concepts into a stable point from which to move forward. These conceptual shifts are accompanied by ontologic shifts, stabilized and pivoting around moments of punctuated equilibrium and finally set in the performative closure of the project.

Performative Closure and Punctuated Closure

Drawing on a concept of a closed enzymatic system as relevant to evolutionary biology, Ger Wackers has developed the concept of performative closure as a tool for risk analysis in complex systems. (2009) Performative closure treats designed objects as complex systems rather than assemblies of operating components, and the maintenance of closure is a continuation of the design process as the given object moves between different contexts or responds to shifts in its system. “Selection does not act on individual mutations but on complex wholes,” (Kauffman, 1993) and designed objects ought to be considered in the context of the goals of their designers in addition to the conditions of its use through time. Performative closure, in short, is the ability of a complex system to continue operating according to a set of designed goals. Performative closure in this sense is not the ending of a project, but rather a given point of stasis or designed status quo. Helicopters, trains, etc. need continued maintenance and upgrades in use, but this is not a new closure to Wackers, the maintenance through time of these technologies is a performative closure of the goal that led to their design. The designed object is at a certain point in the process ‘closed’ but this does not mean it becomes unchanging, but rather that at that moment the closure operates as a pivot for homeostasis. (cf. Stuart Brand 1994 on How Buildings Learn) “Performative closure is (shorthand for) the achievement and/or maintenance of core task completion while maintaining functional system integrity.” (Wackers, 2009) For Wackers, performative closure is always tied to the vulnerability of and risks to functional system integrity: while systemic homeostasis is maintained performative closure is achieved, but if any piece of the system breaks down the object is no longer ‘closed’ to continued imaginative design tasks.

This ties to an actor-network theory argument that design does not end at the point of closure/invention. Rather than thinking of the design process as continuing unabated throughout the life of an object, Wackers acknowledges a moment of project completion, a performative closure, where the designed object is set into the world as a piece of a complex system that insures its continued operation. After this point, the object is subject to a new set of influences on its use, different techniques for its future growth, and new imaginaries for its uses all arising from the complex system in which the object operates. Extending the concept of performative closure to the imaginative design process itself, we see a number of moments of apparent or partial closure of the designed object that posit the complex systems necessary for it to maintain performative closure. Continuing, in the vein of Wackers, to draw on evolutionary metaphors in considering the design process, I call those moments of apparent, figurative or temporary closure instances of punctuated closure.

Generally accompanied by or resulting from periods of intense change within the design process (prior to some moment of performative closure), *punctuated closures are moments of new or newly stabilized homeostasis in the imaginary of the designed object*. I deploy the term 'punctuated' by analogy with evolutionary theory. The analogy is with Gould's (1982) analysis of evolutionary change as being marked by a series of punctuations of general equilibria, where new design possibilities abound: punctuated closures are points where a tactical closure is drawn around a given set of possibilities. These moments occur as consensus on given pieces of the system is achieved, work is stabilized into some smaller set of possible forms, or the design process is momentarily closed for demonstration or exhibition. At these points in time the agreements, intense work, or necessities of presenting the object externally stabilizes some portion of the complex system that will eventually sustain the operation of the designed object through time. In identifying moments of consensus, a punctuated closure proposes its own rhythm of the design process characterized by the collapse of various potential decisions into a single or limited set of possibilities and the subsequent opening of problem spaces taking that single component or limited possibility set as assumed. It is in moments of decision and obvious consensus that punctuated closures of the designed object are most apparent, but they are not limited to overt action.

The novelty of the EarthCube approach was its middle-out design model, in contrast to the top-down computer science/domain structure of the previous generation of cyberinfrastructures. Here tightly knit clusters of domain and computer scientists work together to produce separate 'building blocks', with the occasionally interchangeable blocks (for example, semantic web projects to produce modes for querying multiple heterogeneous databases) being progressively assembled under the guidance of a governance committee made up of domain, computer and social scientists. The temporality here is multiple: ongoing community relations and requirements gathering efforts are punctuated by funding cycles, group meetings, and workshops. Each building block operates on its own temporal scale within the context of the larger project - the smaller cycles of development within the building blocks are punctuated by moments of demonstration or other performances of continuous work. These moments of demonstration and performance are reflexive and summary - much like the close of field data collection for some scientists - and serve to imagine the problem-focused work as a functioning piece of a larger system.

This move is in opposition to the first round of cyberinfrastructure/eScience projects, which in general attempted to produce complete solutions from computer scientists not cognizant with the domain science working with a few domain scientists - with the hope that the broader community would later adopt the solution. The GEON project (Baru et al., 2009; Bowker and Ribes, 2008) is perceived by some within the NSF as the precursor to EarthCube which was 'before its time' is a notable case in point. After the initial round of funding was spent, the shell of an infrastructure had been created, but there were extremely few users. A second round of funding was awarded in order to garner users (at less than a tenth of the initial funding), and duly failed - the site today (nothing ever seems to die on the Internet) proudly announces upcoming events for July, 2011. (http://www.geongrid.org accessed 12 June, 2014)

GEON and related projects existed in a linear flow of time from conception by computer scientists and key informants to notional adoption by a community of users - moving forward in time entailed traversing social worlds. The same linearity was imposed on the object of study (the Earth) - not recognizing the multiple non-normalized temporalities for the Earth's history generated by communities of scientists from different discplines - temporalities often institutionally, organizationally and culturally engrained (Bowker, 2000) - it attempted to develop a single timeline through techniques of database interoperability alone. (The Chronos Project, 2007) This double temporality can be read as a double design temporality - the goal was to design an infrastructure which would enable us to design the earth (at the apogee of the Anthropocene) to suit human needs. Both design temporalities resonated with each other - a top-down design for scientific infrastructure would enable a top-down design for the future of the planet, and both would operate a simple model of linear time. In recent years, more attention has been afforded to infrastructural design from the bottom up (Twidale and Floyd, 2008) - called by Egyedi, Mehos and Vree (2012) 'inverse infrastructures'. Keith Kintigh (1984) attempted to develop such a design solution for archeology (a field with faced with major problems of reconciling data from different sites catalogued in different archeological traditions). Work on models of earth systems operates in a similar way: models of finer-grained natural phenomena are developed in the context of how they fit into a larger cycle. Studies of permafrost outgassing (a community of scientists well-represented within EarthCube projects), for example, performed on ice core samples develop representational models that fit into the more expansive carbon cycle models without necessarily having come from or been operated on the same temporal scale. Permafrost by definition is at least two years old, but may be as much as 1-3 million years old in areas where glaciation has consistently kept the soil frozen. (Janson and Tass, 2014) Analyzing a permafrost core sample requires resolving tensions emerging from widely differing temporal scales. The life cycle of a microbe, seasonal temperature and water cycles, and cycles of global climate change and glaciation all operate on the state of the core under analysis. Building blocks projects, while operating on their own development cycles at scales reasonable to the individual project are still fit to the larger development rhythm.

EarthCube projects fall into the rhythms of both grant and funding cycles of one to two years and several scheduled large-group demonstrations. The work of projects within EarthCube exhibits punctuated closure, where designers and researchers step away from the ongoing work in order to summarize and represent that work publicly since the last moment of closure. Closure "…generally refers to the emergence of some kind of order (socio-cognitive, technical, biological) in interactive (cooperative and competitive) and interdependent processes operating in a selective environment." (Wackers pg. 31) As most EarthCube participants are funded part-time (consistent with contemporary standard academic practice where one works on multiple projects simultaneously) their work and communication has a tendency to cluster into the moments leading up to these performances of closure. The overall architectural temporality is being a series of asynchronous development cycles punctuated by moments of reflection and momentary closure. In addition, at the moment of each engagement of the NSF with its funded community represents a newly-stabilized vision of the project informed by the prior instances but occasionally a significant deviation from the original plan.

EarthCube was planned as an infrastructural project and, typically, the cycle of "becoming infrastructure" takes anywhere upwards of 50 years. (Edwards et al, 2007; Bowker, 2008) As such, moments of design inflection operate at an extended temporal scale – the 'long now' of infrastructure development. (Ribes and Finholt, 2009) As the intent of EarthCube was to allow for community-led bottom-up development, design decisions as to the architecture of the infrastructure itself are expected to occur in a decentralized manner and over an extended period of time. However, one can identify moments of design inflection as punctuated closures in the growth of the scope of intended EarthCube projects as it moved from initial design charrettes intended to directly inform prototype development to the middle-out, building block centered design model eventually implemented. As the project is still ongoing and, potentially, will be for decades, these to relatively stable moments of community design are not moments of performative closure as identified by Wackers, but instead instances of punctuated closure as the vision of the social design of the EarthCube infrastructure stabilizes temporarily. These closures tended to occur at moments intended to address the scale of the project – what Ribes (2014) calls scalar devices.

GEO Vision and the Design Charettes – Changing conceptions of the EarthCube design process

Much like the process of designing the regulatory and observational apparatus for monitoring water quality described in the first chapter, the initial design and development of EarthCube was oriented around a series of scalar devices and knowledge production activities intended to assess the state of relevant communities and drive development and funding based on their technological and scientific needs. In the introduction I briefly summarized the process of initial assessment and design that occurred prior to initial funding, which included opportunities to both allow a community of geoscience to constitute itself relative to the policy-driven goal of infrastructure development as well as more directed interventions from the NSF intended to target and engage particular communities and sociotechnical issues. The NSF, in its goal of funding and supporting basic science, grapples particularly with issues of scale. Larger, more established research universities with grant support tend to be overrepresented in calls for proposals as well as dear colleague letters – in a demographic assessment of atmospheric science investment by the NSF it was noted that out of the 70 institutions with atmospheric science funding 40 were R1 (in the Carnegie Classification), 11 were R2, and "the remainder came from four master's schools (3 M1 and 1 M3), three baccalaureate colleges, seven small businesses, one Federally Funded Research and Development Center (FFRDC), two U.S. Government laboratories, one nonprofit, and one four-year engineering-focus school." (Avallone and Bauerle, 2017) However, despite this apparent institutional diversity, more than half of the PIs in Avallone and Bauerle's analysis came from just 11 institutions, all of them R1 schools. While to some extent it is a bit tautological to state that R1 Universities (so classified because of the quantity of research funding they receive) win more NSF funding, in recent funding of atmospheric science – a class of geoscience that was a part of EarthCube's intended community – there is an apparent concentration of funding and research support in relatively few institutions. This would seem to be the Matthew effect (Merton, 1968) in action: to them that hath shall be given. Those research universities that are already major players in the funding game, be it due to name recognition of the PIs, more experience with the grant-writing process, or just greater institutional support of research activity, are overrepresented in funding decisions, which represent a major means by which the NSF gauges and understands their research communities.

EarthCube shows significant evidence of attempts to grapple with the scale of the earth systems science community outside of the process of grant solicitation and awarding. As argued by Ribes, these moments of assessment, evaluation, and co-presence can be described as scalar devices – moments where a very large community can be more effectively represented, interpreted, or investigated ethnographically. (Ribes, 2014) This is a significant knowledge production activity on the part of the NSF as well as the presentation of a site for coordination among scientific actors. Voluntary participation in these activities, submission of letters, and eventually the preparation of grant proposals all serve simultaneously as moments where the community represents itself to the NSF even as it constitutes itself around that particular activity. Prior even to the workshops was a solicitation of white papers from potential PIs. Much like how the Atkins Report (Atkins, 2003) outlined an agenda and plan for cyberinfrastructure investment, the GEO Vision report (2009) articulated a set of particular challenges and infrastructural goals for the geosciences that directly informed and led to the EarthCube project. The GEO Vision report identified three major challenges in the geosciences: predicting and understanding behavior of Earth systems; reducing human vulnerability and sustaining life; and developing a geoscientific workforce capable of addressing the prior challenges. The GEO Vision report also emphasized the need for geoscientists to be able to effectively represent their research to policymakers and framers, the growing acceptance of the notion that atmospheric and hydrologic science models ought to be bridged or coupled together, and the need for advanced computation and data sharing resources. "Over the next decade, the geosciences community commits to developing a framework to understand and predict responses of the Earth as a system— from the space-atmosphere boundary to the core, including the influences of humans and ecosystems." (GEO Vision, 2009)

Following from the recommendations of the GEO Vision report the NSF published a document titled "EarthCube Guidance for the Community", which not only issued a call for participation in the early design process for EarthCube, but also outlined some of its more important design goals. Community-led design was emphasized, as well as the goal of building upon existing cyberinfrastructural investment by producing an integrative system. It was assumed that prior to participation in this project that interested communities would self-organize in order to have their interests effectively represented – the guidelines for applying to participate in the initial design charrette specifically stated that "Initial self-organizing efforts will occur prior to the EarthCube Charrette," with the goal of producing initial collaborations and making sure that the limited charrette space would most effectively represent the interested communities. (Earth Cube Guidance for the Community, 2011) The initial few months of EarthCube were specifically oriented around community self-organization, and an online forum was providedto encourage collaboration and the sharing of some initial ideas. Among the major outcomes of this initial design charrette was the concept of developing an initial landscape of community issues through the mechanism of 24 end-user workshops, including one workshop specifically dedicated to allowing potential PIs an opportunity to assess, discuss, and account for the results of prior workshops. The progress of EarthCube as a concept is closely linked to these moments of reflection, assessment, and community participation. The original guidance document did not foresee or specifically predict the end-user workshops, and actually expected a prototype infrastructure to be developed by December 2013. (Earth Cube Guidance for the Community, 2011) Prior to the design charrette there was little notion of the eventual design strategy present in either the GEO Vision report or the initial calls for participation – building blocks and integrative architecture grants were not mentioned, and there was an apparent expectation that there would develop a functional system demonstrating the EarthCube cyberinfrastructure before the period of heavy investment (2014-2022) would begin.

This initial design charrette would redirect both the funding efforts and the overall strategy for infrastructure development – following the initial charrette the NSF changed their goals from the development of an initial prototype to a widespread period of community assessment and feedback. Much like the regulation of water quality through the TMDL functioned as a process of mutual adjustment informed by knowledge production occurring both on the part of policymakers/funders and researchers the initial phases of community participation served both as organizing activity and as a feedback mechanism for the NSF's design strategy and funding plans. Through the design charrettes and initial end-user workshops the goal of EarthCube shifted as well, from the rapid development of a prototype infrastructure that would be worked upon by the broader community to the community-driven, 'middle-out' design model revolving around the funding of relatively independent building block projects, RCNs, and data facilities that would eventually have an architecture built around them. This feedback loop between communities and the NSF was fundamental to the design process of EarthCube, present in the earliest guideline documents and existing structurally in the pace of funding where each ongoing year additional building blocks, integrative activities, etc. would be funded alongside the evolution of the original building block projects. The ongoing calls for proposals served as a mechanism by which the NSF might potentially adjust the design of EarthCube, incorporate new communities, or address perceived gaps either in representation or technology.

The EarthCube community was assembled through the use of scalar devices (Ribes, 2014) – these moments of community assembly and engagement served not just as opportunities to speak to policy as embodied in science funders but also as inflection points for the membership of the eventual EarthCube community. "A specific goal of these workshops is to gather requirements on EarthCube science-drivers, data utilities, user-interfaces, modeling software, tools, and other needs so that EarthCube can be designed to help geoscientists more easily do the science they want and need to do." (Azriona Geological Survey, 2013) This is an important mechanism of co-production in the mode of Jasanoff, (2004) where an ongoing process of mutual fitting is embodied in planned community assessments, meetings of representative groups, and the solicitation of voluntary participation and opinion. The initial design charrettes, followed by the submission of white papers, development of community groups, and the eventual constitution of the end-user workshop communities each represent a moment of performative closure, where the design strategy of the infrastructure momentarily stabilizes around a certain vision not only of the eventual future system but also of the process by which that system might be achieved. As each round of funding introduced new building blocks, integrative activities, data centers or RCNs, the makeup of EarthCube as a funded community of researchers and technologists changes abruptly only to adopt its relatively stable new form until the next round of funding changed it once again.

Most recently, the NSF has funded a consulting group – the Xentity corporation – to guide the development of an architecture capable of spanning the completed building blocks, functional tools, and data produced through EarthCube funded projects. While this occurred after my period of research and observation concluded, it represents a new design mode and vision for EarthCube. Rather than focusing particularly on incorporating all building blocks, integrative activities, RCNs and data centers into a centralized infrastructure Xentity's solution architecture will prioritize a registry of existing EarthCube resources and a workbench allowing for certain types of interoperability. In addition, the employment of an industry consulting group (rather than an academic unit) to guide the development of a solution architecture for EarthCube represents a mode of collaboration fairly new to the project – that of industry engagement. This most recently-funded consulting group is a new punctuated closure for the EarthCube project. In the time since this solution architecture proposal was funded work among the governing committees (particularly the Technology and Architecture Committee and Science Committee) as well as the apparent brunt of community engagement has been reoriented towards its development and implementation, as well as the site of some significant criticism:

> "Of the EarthCube tools that have been built, it's possible that those constructed as back-end functions to assist with some aspect of data interoperability could be useful to include in the initial core platform... Generic sockets for translation through a small set of formats (e.g., JSON-LD); the ability to match data fields and convert units or scales between datasets, maybe using something like graph data structures as containers; and interfacing with a small set of standard APIs (e.g., REST) would be helpful... If constructing the data interoperability core functionality in a generic adaptable platform is beyond the scope of this contract, it is still problematic to instead construct a recommendation system and brand it as EarthCube core functionality when in reality it would still be another add-on to a missing core platform." (Comment Posted after 10/2/2016 meeting of the Science Committee)

It is still too recent to assess the effects and response to this new architecture. However, it is worth mentioning as yet another closure around which work is orienting, and shows institutional learning in the form of the NSF's funding of a form of collaboration not anticipated earlier in the project. However, this process of design, feedback and redesign is not without its tensions, and the particular mode of funding and intended middle-out design model bore significant consequence for funded projects as well as the EarthCube community as a whole.

The design strategy of EarthCube accommodated parallel development trends, with many of the building blocks working towards similar goals, but has several structures of implicit selection, where the choice between modes of solving a certain problem are chosen ‘behind the scenes’ of project governance through the NSF project cycle and shifting funding decisions. In not funding original research *per se,* the focus of EarthCube is on tool development, the formation of collaborations, and the re-use of existing data and data streams. However, the development of new tools, collaborations, groups, and interconnections requires that change propagate throughout the scientific community regardless of the durability of a given building block in terms of the EarthCube architecture.

> “as someone who has been involved in EC enterprise almost from the start, I have to say that Science has always, rightly or wrongly, been a placed [*sic]* behind the Technology. This is wholly because of the way NSF funds EC!... The idea that EC can develop a or several technologies that will transform the way 'we' do Geoscience is a non-starter, simply because of the diversity of topics/approaches involved. ” (Leadership Member of EarthCube, posted to Science Committee listserv 1/27/16)

The larger amount of time and effort on the project has so far been community organization, requirements collection, and research. Overall project governance was established as a building block of its own, tasked with bringing together representatives from the other building blocks to design a governance structure for the overall architecture and foster cross-award collaboration. The design process has been organized around moments of reflexive co-presence (workshops, meetings, webinars, etc.) and document publication. The initial focus on software solutions in the form of the building blocks awards determined the nature of possible work. The lack of funding for new data along with the requirement that building blocks address issues raised during end-user workshops resulted in a set of mediating (in a variety of ways that can be understood) positions for many of the building blocks. Many of the building blocks arising from the end-user workshops (with the caveat that in general new data will not be funded) claim similar goals through disparate and relatively incompatible means. At the end of the day, the movement of particular building blocks into the architecture and infrastructure would not necessarily be determined by the quality of their science, or the capacity of their technology, but rather by the willingness of the NSF to select that project for ongoing funding.

Designing Building Blocks: Tools versus Science

The hand of the NSF rested heavily on EarthCube conceptually and organizationally from the beginning of the project. Despite a stated goal of producing a middle-out design for infrastructure that could produce relevant and revolutionary science, EarthCube projects tended towards tool development with few geoscientific outcomes. As of 2016 EarthCube funded 51 proposals, each expected to contribute to the design, governance, community or functionality of the infrastructure. Projects were divided into technical Building Blocks, Research Coordination Networks (RCNs), Conceptual Designs, Integrative Activities, and Data Infrastructures. BCube, the Building Block project I was working on, was funded in the 2013 round of funding. Of the 51 proposals, nine were RCNs, twenty-five were Building Blocks, three were conceptual design awards, thirteen were integrative activities, and only one was a data infrastructure award. This was the result of several rounds of funding that began with nine building blocks, three RCNs, and two of the three conceptual design awards in 2013. 2014 saw the introduction of an additional 14 projects, eight of which were building blocks, and 2015 added 14 projects, none of which were building blocks. The 11 projects funded in 2016 included eight new building blocks (the funding of all prior building blocks running out around this time), two RCNs, and the lone data infrastructure award. What is important to note from these numbers is the way that the NSF was directing funding throughout the EarthCube project – of the 51 funded projects, 41 were primarily interested in technology development at one stage or another, and only nine directly sought to increase engagement and enroll collaborators (the RCNs). Throughout the life of the research project funding for community organization, outreach and enrollment was secondary to funding tool development and data integrative activities, and by a fairly wide margin. The tool-focused development had a number of effects on the project as a whole, and most particularly led to a (still-growing) sense of disconnect between the governance of the project through its own committees and the actuality of funding mandates through the NSF. The following extract was written during an email conversation about how to more directly engage the work of EarthCube with youth, particularly in the K-12 area.

> “This is the fundamental frustration I've had with EarthCube in general in the last few months. I was also at GSA, [Geological Society of America Annual meeting] and also spoke with people who were obviously interested in becoming engaged, but I had no idea what I should point them to, to get them started. They can sign up on the website, but as we all know the website isn't very engaging. I could point them to the EarthCube booth at GSA, but as far as I could tell, all they were doing was signing people up to the website and mailing lists. I could point them to the Test Governance monthly meetings, but based on the low participation in most of the groups, those aren't very engaging either (and probably hard to just jump right into without some sort of intro at this point).
>
> There needs to be something specific for people to *do*. Not just receive information.
>
> Meanwhile, in the most recent Science Committee webinars, the recurring them is that we aren't doing a good job of engaging people, AND there are projects like the semantic wiki and the use cases that crucially need input from people.
> It's like we're all talking past each other. They want to be engaged, we want to engage them, there are things they could do to become deeply engaged and yet we aren't telling them about those projects.” (Posted 11/6/15 to the Science Committee listserv)

The tool-design focus of EarthCube served to somewhat alienate casual participation, given the significant overhead necessary to participate in the design of a tool that is partway through its own project cycle and requires some substantial bit of specialized knowledge. Even those projects that might have been more accessible were beholden primarily to the funding cycle and NSF, and the more efficient means of working required that the group remained relatively closed – with no specific educational grants funded engagement with inexperienced or differently trained professionals was limited to those who were already actively engaged with a product already present in EarthCube.

And here is where the discourse on infrastructure most directly impinges upon the design of the infrastructure itself. At the meso- level, the design of EarthCube's architecture was about the future, and revolved around imaginaries of what infrastructure might exist to support the work of that nebulous, ill-defined but continuously referenced 'novel science.' When challenged in interviews and in more casual conversations, few of my informants were able to express the 'novel science' that might one day be enabled by the larger infrastructure of EarthCube, instead speaking of particular challenges within their own research communities. A tundra scientist I interviewed, for example, was concerned with access to and sharing of dark data from other researchers in the field. (interview conducted 01/2014) While there was significant work being done on tundra core samples in various institutions there was little sharing as each lab found and analyzed their own sample, stored their data locally (usually in the form of an excel sheet), and didn't necessarily conform to a shared standard of metadata or description. Those informants existing more closely in the computer science space were concerned with data standards (particularly among datasets that they might want to use for model verification), with tool design and development, and in many cases with the replicability of models created from streaming data sources. The immediacy of these issues stands in rather stark contrast to the lofty goals of an infrastructure project. In many ways, the middle from which the imaginaries of designing the architecture of EarthCube's infrastructure was itself missing the middle of their design process, little accounting for the steps between immediate problems and grand challenges. This was true of BCube itself, as well, in its scenario development process (which I discuss in more detail in a later chapter) – 'scaling up' into interdisciplinary space exposed not only the technological and organizational barriers one would expect when working with disparate and heterogeneous data sources but also exposed an issue of imagination. While the grand challenge was present and work was characterized as a move towards that challenge, the organizational, technological, and even ideological processes by which the immediate challenges might scale into that grand challenge were not consistently imagined among group members. The building blocks and other funded projects were imagined as comprising parts of an indeterminate whole – one that had a role, a goal, and a future but seemed inevitably further away with each step of progress.

Infrastructure is irreducibly a combination of material technologies, tools, competencies, and social groups. (Star and Ruhleder, 1996) While the community of EarthCube was focused around enrollment, outreach, and the development of a social reality around the design of a useful, revolutionary infrastructure, funding decisions made outside of the community repeatedly prioritized the material (in tool development, in data warehouse development, in catalog and discovery services) over the development of a community that needed and was capable over making use of that material. A road that nobody can drive on is infrastructural to nothing.

> "EC is currently facing much headwind because tech was put ahead of science. The way it should have been structured would have been to ask the community to define scientific challenges that couldn't be solved with existing tech/cyberinfrastructure, then pair up scientists and technologists to work TOGETHER to solve those problems (or at least, make headway in that direction). Instead, technologists have hijacked the enterprise and most use EC funds to advance their own pet projects, with no meaningful science outcomes in sight (in many cases, only involving scientists on the project so it gets rubberstamped)." (Posted 1/26/2016 to the Science Committee listserv)

While this is likely more an expression of the frustration of its author than the character of the EarthCube community as I observed, there existed a strong sense among those I spoke to that EarthCube was a technology design project with scientific consultation, rather than a project designed specifically around pushing science forward.

> "We are having a fundamental lack of communication within EarthCube and between EarthCube and the relevant funded groups, and between NSF and EarthCube. Since ultimately NSF reviewers are deciding what gets to be an "EarthCube" project, most people on funded projects are far more responsive to NSF guidance than EarthCube Governance" (Posted 11/6/15 to the Science Committee listserv)

A second notable effect (implicit in the above quote) was the feeling like it was difficult for anyone not directly funded through EarthCube to be able to participate in the larger imaginaries of the community. Funding was not only a means by which projects were supported, it was a second category of membership and participation in the infrastructure project. Funded projects were working on the infrastructure, whereas unfunded membership was participating in its organization.

The Global Biodiversity Information Facility (GBIF)

One of the major issues arising in this sort of study of infrastructure design comes from the fact that despite a growing and diverse field of study of new infrastructural projects and general innovation studies that there is no prescriptive method for designing successful infrastructure. While we often talk about the unique design context of infrastructure (particularly science-enabling infrastructure like EarthCube or GBIF) and the far-ranging outcomes of design and infrastructural work, it often seems to fall by the wayside that these shifting contexts in which design occurs also lead to unique endogenous failure conditions as well as unique endogenous success conditions. EarthCube is still strongly in the mode of measuring success through engagement – that is to say, the adoption, use and participation in EarthCube as a sociotechnical infrastructure is one of the primary imaginaries of success propagated through the ongoing work of the various subcommittees and in the steering committee. Both the Technology and Science Committees are working (together) on developing use cases to engage external scientists in the infrastructure, and adoption will be a major measure of success for the recently-funded architecture grants proposing the overall technical organization of the infrastructure. These measures of success stand somewhat in opposition to what might constitute success of its parts – in the usual run of modeling, data collection, and other scientific grants the coin of the realm is primarily the production of publications. The real-world impacts of the research work are measured not by continuous long-term engagement but instead by how amenable that publication is to citation and how well those citations support work that itself is cited. A model can be used by only a single researcher, once, and still be a very successful model on the basis of its outcomes in publication. A particular finding need not be further expounded upon to still have a significant impact in the field. A piece of software or device that results in significant citation can be successful from theviewpoint of academics regardless of whether or not it is well-used, maintained, or even shared. By these measures, even 'failed' or inactive infrastructure projects can be significant successes, resulting in publication, bearing impact on future work, and producing meaningful science, if only for a short time during their period of funding. EarthCube and other infrastructure projects that are similarly mixes of organizational, social and technological goals, are a category in their own right in the world of academic research and funding. Producing resources, encouraging reuse, and developing a platform all introduce the notion of innovation failure into existing scientific practice.

While innovation has always been a significant part of the scientific process, innovation failure – which is when the evaluation of scientific work is not on the basis of its intellectual merit but instead on its propagation as an invention, as seen in projects like GEON – seems fairly unique to science-supporting infrastructural projects. (Schumpeter, ; Bowker and Ribes, 2008) In the innovation-mode the work of scientists and researchers is subjected to competing regimes of evaluation that bear disparate conditions for failure and success of the same academic work. Data-scientific work like model development, broker development in the case of BCube, and the creation of resources supporting scientific work does not easily move into the publication space (though that condition is improving) and scientific work resulting in interesting and replicable findings would not really be expected to simultaneously result in a 'saleable', broadly usable product. Lessons learned in design often come at the cost of a particular iteration of that design which is discarded or substantively changed to incorporate new knowledge. Large-scale infrastructure projects, however, are more of a one-shot than would be ideal in an iterative design setting. From a market perspective, a new resource, infrastructure, service, etc. is 'sold' to researchers, institutions and scientists, whose use of/participation in the designed object becomes the value returned to the designers. The 'payment' in this case can be interpreted as the willingness to adapt and change existing research/institutional/data description practice (which is not a low barrier to navigate). New skills must be learned to assess and use new infrastructure, new devices produce new competencies, and infrastructure as a general phenomenon both requires and produces a regime of knowledge practice about that infrastructure. There is significant capital to be expended, then, in the enrollment of researchers and scientists in new infrastructure projects – it stands to reason that not every researcher is willing to change their practice for every new infrastructure project. The innovation-mode of science, then, finds itself eminently concerned with a form of marketing and market manipulation necessary order to enroll, engage with, and effectively present itself both to potential user communities and future stakeholder groups. This raises the stakes.

The sheer scale of resources necessary for infrastructure development place results in increasingly-large decision making groups and stakeholder populations. (Galison and Hevly, 1992) Public scrutiny is a natural consequence of such substantial projects, particularly when they are government funded. It can be difficult to parse the difference between a project exploring a variety of standards and policies and the notion that that project represents a funded mandate towards particular forms of interoperability. With aggregation from a variety of sources, metadata is unlikely to capture the full extent of information needed to make the data work. Particularly in the area of troubleshooting, it would be difficult to pinpoint where problems emerge from the data itself (flawed sensors, software bugs in the broker) without substantial documentation and understanding of the broker's function on the part of the users. Some will feel that they do not have enough from the metadata to make an informed decision, and others will have an insufficient sense of control. The next step in development is not in adding features to the broker. Extant features are disbelieved or ill-understood, and future development and demonstration at each stage that is evident to me requires the investment of some community of users – either data managers, EarthCube committees, or the wider scientific communities. Assuming EarthCube results in a success, buy in from any of these increases the likelihood of good placement of the broker – as middleware it does not work on its own.

However, this risk of innovation failure, the lack of adoption, or even an unsuitability of the technology to the goals of its creation is not an end point in and of itself. There is a moment, outlined somewhat in Bowker and Star (2000) when speaking of the adoption of the International Classification of Diseases, where there is a turn to infrastructure – where a system intended to serve an infrastructural role becomes less an object of work and more the site of it. The mechanisms and reasons for this turn are a bit opaque and may not be directly resulting from the material characteristics of the intended infrastructure. GBIF, my comparative example here, exemplifies how a relatively centralized data repository slowly turns to infrastructure despite what initially appear to significant flaws in its design. Despite these flaws, GBIF is evidently moving to infrastructure, supporting research, policy, and collaboration internationally.

GBIF as a comparative example

As the ability to meaningfully process increasingly large quantities of data has improved, the need for systems to support the aggregation and subsequent use of disparate smaller datasets is correspondingly greater. This need is compounded in efforts attempting to understand the dynamics of global-reaching systems for the purposes of defining and guiding national and local legislative and regulatory efforts. The GBIF (Global Biodiversity Information Facility) is just one such project among a larger group seeking to aggregate the smaller, focused, and disparate sources of information generated for the work of science. GBIF is simultaneously an effort to coordinate and aggregate digital species occurrence data and digitize natural history collections into a single global-scale resource for biodiversity work. The GBIF database is a piece of infrastructural design work, serving as a functional basis for expanding scientific and policy work on biodiversity and conservation without functioning as a sole source for either. Management and maintenance decisions made in the creation of the database render it unsuitable as a primary source for many categories of biodiversity science and unable to provide the analytic tools for evidence-based regulatory decisions. However, the value of efforts like GBIF is not in their present suitability for modern scientific and political tasks – there exist many databases and experts already who can fulfill those needs – but in providing a basis for future growth and in the development of new science and policy goals. GBIF as a social movement, as a political and scientific statement, and as the core of an infrastructure for global work in biodiversity science may well be more useful in guiding the course of global conservation policy than the data it contains. The value of an infrastructural effort like GBIF, instantiated

in its aggregation of wide swathes of species occurrence data, comes not from its immediate usefulness but from the creation of the social, political and technological conditions needed to support the development of such efforts at scale.
Critical reflection on the design of information systems and other artifacts shows that humans embody their values and morality, often unconsciously, in the things that they create. (Winner 1980; Latour 1992; Hughes 2004; Nissenbaum, 1998) These values may be intentionally *designed into* the physical state of the artifact or system (Flannagan, Howe and Nissenbaum, 2008; Friedman, Howe and Felten 2002) or be observed resultant from and of a myriad of social factors. (Pinch and Bijker, 1987) These values can produce bias (Friedman and Nissenbaum, 1996) or otherwise be seen to have and carry politics of their own. (Introna and Nissenbaum, 2000) Successful infrastructures serve those with a variety of values, but may prioritize certain values in their design. (Knobel and Bowker, 2011) For example, mobile technology that automatically reports your location through GPS to your friends and family values connectedness and intimacy above privacy. Though these value propositions are evident in the objects themselves, often they are the result of unconscious assumptions on the part of the designer, making it quite difficult to avoid their potential negative impacts on quality of life. (Introna and Nissenbaum, 2000)

The field of biodiversity science as described here is a particularly lucid example of how a single database (prosaically and in effect, the earth itself and all its living things) must be 'made amenable' to either scientific work or policy work. While biodiversity science, with its implicit focus on conservation as the end goal of the scientific work, resides very close to the legislation and regulation it informs, there is still evident the effort of and push towards rendering knowledge amenable to policy work. Bowker (2005) identifies two ways in which biodiversity science comes to relate itself to political action: implosion and particularity. Briefly, implosion is attempt to discover and apply value to various categories of living things, and particularity is the effort to exhaustively catalog and describe those living things. The biodiversity scientists in the modality of implosion of the database of the world work to render it amenable to valuation of the sort that would be useful to politicians, with all its attendant simplifications, obfuscations and hidden moral arguments. Given the scientist's existence in an arena of implicit knowledge, jargon and other restrictions to participations, the desire to see translational work as a simplification (shedding of context) rather than a context shift distances analysis of the data from the data itself, management of the data from its analysis, the mobile objects of knowledge from shared understandings implicit to its creation. Categorization then potentially become subservient to valuation, with disagreements over the ecologic role of a category of living things played out in its classification because that classification carries political value. The GBIF data portal is a particular (in this sense) effort of biodiversity science, and though it seeks to support the work of both scientists and policy-makers the metaphoric distance of the complete database from the data actually used in science is quite far. The global data portal renders amenable not the species occurrence data itself but the concept of global biodiversity mapping, the usefulness of having a global-scope resource, and the value of working globally on global systems.

Knowledge is produced and intentionally disseminated as part of both scientific and policy work and their agendas are mutually influential. (Irwin and Wynne, 1996) Just as objects of knowledge, specific data sets and scientific outcomes move between different established conventions of evaluation and representation within scientific disciplines so to do they change moving outside of vocational knowledge-work. The management, analytic and interpretive work a scientist does on and through their recorded data is temporal and contextual, just as is the similar work of a politician. “Knowledge is located in the nexus of participants, practices, artefacts and social arrangements,” and occurs and is validated within the context of a given community of practice, according to their own standards. (Van House, 2002, pg. 111) The boundary between scientific and political work can be seen to emerge more from standards of evaluation and the implicit claims of the utility of that knowledge than exist as a firm disciplinary divide. Data collected and publicized on the behalf of government is not fully in the territory of the sciences or of the politicald – its release implies a benefit to information movement across, without and into society as a whole. The claim of societal good bridges the discursive work of the politician and the scientist, and it would seem the discursive power of the organizations and terms that emerge from this work is not fully held by either. While the workers in science and policy may be vocationally separated the co-construction of knowledge, relevance and epistemological discourse cannot be quite so easily disentangled. Scientists can act as advocates for given policy regimes or political philosophies, politicians make take an active role in setting scientific agendas and supporting certain technologies over others. Biodiversity sciences can be seen as emerging from a belief in conservation on the part of biologists that is supported in, rather than developed from, scientific knowledge work. (Takacs, 1996) One of Takacs interview subjects, on the question of whether a species should be conserved, said,

> "the answer is always 'Yes!' with an exclamation point. Because it's obvious. And if you ask me to justify it, then I switch into a more cognitive consciousness and can start giving you reasons, economic reasons, aesthetic reasons. They're all dualistic, in a sense. But the feeling that underlies it is that 'yes!' ..." (Takacs, 1996, preface)

The practices present in the organization, structure and data practices of the GBIF data portal are formed in at least some part towards that conservationist goal – though there is little visible connection to specific policies, many of the design characteristics of the aggregated database support re-interpretation for policy goals often seemingly before scientific. The work of biodiversity science is closely entwined with the work of conservation biology, and both are difficult to fully separate from the set of values and political ideals that lead to national and international conservation movements.

The Database

The GBIF database itself is part of a cluster of similar efforts to develop wide-ranging, globally scaled data centers for the study of systems that operate at larger scales. The Long Term Ecological Research Network (http://www.lternet.edu/), the Group on Earth Observations Network (http://www.geongrid.org/), the Global Earth Observation System of Systems (http://www.earthobservations.org/geoss.php) and EarthCube (http://earthcube.org/) are all examples of similar efforts in unifying heterogeneous data sources on the scale of the earth itself, rather than limited to specific regions. GBIF itself is an international effort undertaken by a variety of labs, governments, and data collections to assemble all primary biodiversity data into a single database. As such, we explore the case of GBIF as part of a study of the general move towards global-scale, international, cooperative infrastructural efforts evidenced by the continued growth and preponderance of these projects. In its current state, the GBIF database includes over 15,000 data sets from more than 600 data publishers with over 500 million species occurrence records. ("What is GBIF?", 2014)

The database itself is heterogeneous in terms of sources with given data sets are unified by a controlled metadata structure and are automatically aggregated to a single storage site. Data can be generated at any of the participating sites, but is only included in the GBIF database when it is cleaned and organized in such a way to conform to the metadata and file format standards decided by the GBIF participating members. However, the distributed nature of data production opens up a certain set of problems that characterize many informatics ventures. A particular issue is that of taxonomy: "In general, obtaining reliable, updated taxonomic authority files is a major problem for most taxa. The large taxonomic information services… remain far from complete." (Soberon and Peterson, 2004, pg. 692) And even when completed, taxonomies are not static – the information facilities must be maintained as new consensus is reached and new debates emerge. Soberon and Peterson (2004) point to several other major limitations in biodiversity informatics, including improper identification of specimens, outdated taxonomy and faulty georeferencing in electronic collections. Data submitted to GBIF undergoes an automated form of cleaning in order to account for some of these problems – particularly that of faulty georeferencing. Georeference data is excluded if it is exactly on the equator (0 for latitude coordinate), exactly along the prime meridian (0 for longitudinal coordinate), or a certain distance away from the coastline. These data points are excluded to account for what is assumed to be the common user error in providing location data on specimen sightings – accidental exclusion of one or both points in the coordinate pair, and accidental reversal of latitude and longitude. The cleaning process is automated, and can only account for good-intended errors – other georeferencing errors can be much harder to correct, or even find, without deep exploration of the data as it is being submitted to the database. The distributed nature of GBIF is such that significant trust is placed in the scientific standards of the participant nodes – which is the primary method for accounting for the other potential problems (incorrectly identified specimens and outdated taxonomy) in biodiversity informatics as discussed by Soberon and Peterson.

Rigid metadata standards and the necessity of transforming the researcher's own database in order to allow its inclusion (often while changing the way the data had originally been organized) limits the number of smaller efforts – particularly those with limited staff – that can participate. Yesson, et al. (2007), in analyzing the coverage of the GBIF database, found that it was substantially accurate, though coverage was lacking in several significant areas. The database does not cover well biodiversity hotspots, its collection is dominated by a few very large data sources, and many biodiversity "hot spots" are not covered at all. Hot spots are places with significantly wider varieties and density of species representation and overall represent areas of particularly dense biodiversity and are often pointed out as areas in most need of conservation work. (Mittermeier, et al. 2005) Important to the study of biodiversity is in the analysis of global patterns of movement, migration and growth. It is because of this that there even exists the attempts at a global database like GBIF. However, the exclusion of hot spots and some portion of oceanic data (eliminated in some areas to avoid "data shadowing" caused by accidental transposition of coordinate pairs) can serve to limit the database to already well-known information. According to Yesson et al. "it is possible to retrieve large numbers of accurate data points, but without appropriate adjustment these will give a misleading view of biodiversity patterns." (Abstract, 2007) Establishing standards for metadata creation and assignment, establishing acceptable synonyms and namespaces, and relying on the scientific authority of participant nodes eliminates some significant sources of error, but also serves to exclude smaller scientific efforts, certain classes of citizen science (though not all – see Sullivan et. al, 2009), and scientific institutions that may not use a similar standard set or work methods.

More important is the claim that the GBIF database and data portal will eventually 'underpin policy,' and though there are a couple of efforts underway working to establish a better relationship between the biodiversity data and policy recommendations produced by scientists, there is little in the infrastructure itself to establish the relevance of certain data to particular regimes of regulation or policy outcomes. The GBIF database's production is decentralized with central control of metadata and formatting standards for the data produced by participant nodes. Data produced and submitted to the database is primarily from Western Europe, North America, and Australia with no data at all from China and minimal coverage of such biodiversity hot spots such as those in South America and Africa. (Yesson, et al. 2007; Collen, et. al 2008) Even as a proof of concept, a single global database that underpins policy formation, problem framing, and decision making can be a troubling notion. The designation of a global biodiversity information source implies a singularity to the database – it implies global coverage that is not necessarily global. The vast majority of species and global biodiversity is in microbial life and oceanic life, but these are not the species of political or conservationist concern. The GBIF database does not represent microbial life in a substantial way, and traces oceanic life only so far away from the coast of a given continent due to their data sanitation methods. Consistent with other movement towards including human action in studies of ecology, (Ribes, 2014) it is important to remember that biodiversity and conservation is at its core about the human ecology, and that is replicated in the data captured and presented.

The singular, global database has the tendency to undermine the conclusions of its scientists by oversimplifying the richness of scientific debate that generated that data. A significant portion of the GBIF metadata structure is the current (to the time of observation) taxonomy of the observed life form. The taxonomy of data points within the GBIF database is static: once data is submitted there appears to be little way to represent changes or new findings in taxonomy. Given its ultimate reliance on species and taxonomy as the primary descriptive metadata of its collections, the GBIF database is oddly silent about the work being done by biologists and geneticists who are challenging and debating the very notion of species as well as changing and making new discoveries in individual taxonomies. Once the observation is entered into the database, it is removed from the field of scientific debate that is occurring even among the scientists that generated that data. While this can be ameliorated somewhat by synonymizing and in the creation of namespaces, building these concepts into a database that is attempting to be global has the apparent effect of formalizing concepts and structures that are still far from consensus within their fields. In addition, namespaces, though various equivalents have been around for over 150 years, are very difficult to check against each other, and establishing species equivalence and distant locations represents substantial scholarly effort. (Bowker, 2005) While this might not be an explicitly political goal, the appearance of consensus that emerges lends itself it to more apparent credibility in pursuing species conservation. To a large extent, the taxanomic and other scientific work does not need to be represented in calls to action in protecting certain areas or species. "The credibility of regulatory science ultimately rests upon factors that have more to do with accountability in terms of democratic politics, than with the quality of science as assessed by scientific peers." (Jasanoff, 2003)

Gaps in GBIF's coverage can, to a certain extent, be explained by the availability of funding for certain scientific efforts over others. "But the U.S. has a lot more money and media than Thailand, so a species or subspecies of vertebrate in North America is likely to get more resources and attention than the possible extinction of the entire Family Craseonycteridae." (Kinman, 2002) In addition, the amount of work necessary to verify and clean the millions of species occurrence data points by hand would be non-trivial in terms of time and opportunity cost. And in fact there is little need for such complete coverage. The structure of the database itself alongside the goals of its creators reveals a reliance on a scientific advisory system (common in the EU, UK, Australia and America) in the creation of biodiversity policy. The central political theory that is implied in the creation and accretion of the GBIF database relies on a relationship between policy and science that is somewhat simplified – it holds to the "traditional view" presented by Almeida and Báscolo. (2006) The production of data (or scientific knowledge) is seen as an accumulation of useful facts – the greater the number the more useful – that policymakers are able to draw on when making decisions. In order for the GBIF database alone to underpin policy decisions policymakers must be (near-ideal) privileged rational actors capable of effectively interpreting scientific results as part of a series of strictly evidence-based rational decisions outside of economic, public or other political influences. In the field of health policy, there is (according to Habricht, et al., 1999) a three-fold requirement of adequacy, plausibility, and probability that ought to be met in allowing scientific evidence to inform policy decisions. The three-fold requirement is deeply contextual and relative – it seeks to establish a formal structure for evaluation because one is needed as a basis from which to work. Much like the GBIF database itself, inconsistencies in coverage and need for additional expertise is seen as insufficient reason to abandon the development of such an effort – it is a basis from which to work rather than an ultimate authority of its own. The data in the database is present for interpretation, it cannot represent knowledge in its own right except in the assumption of substantial knowledge and skills required for interpretation of that data. There is little in the structure or nature of the database or data portal to indicate any translational relationship of the data to a policymaker or debate – those tasks fall once again to diplomats of science.

Calling a database global implies consensus where consensus might not exist. Expecting that database to underpin policy decisions requires a significant, but in the case of GBIF, non-obvious translational step. The GBIF data portal reduces its findings to a single number: at a certain geographic point there are a certain number of species observations. A certain species (not individual, not colony, not family unit or nest) has been observed some number of times. While there exists data structures to report the time of observations, these are not required for entry in the database – some records contain no time of observation at all, and others include no georeference. Movement of species and change in species over time is an observation that is not built into the database itself – these things can only be interpreted through other scientific knowledge and an additional act of interpretation. A regulatory or legislative decision, such as whether to preserve one plot of land over another, requires that interpretative act to be performed, as well as significant amounts of other information that is not and (under current metadata standards) cannot be represented in the database. There is no place in GBIF for ecology: it does not render relationships between species, simply numbers of observations. While it may be said that this is outside the intended scope of the database, it does raise a question: if meaningful interpretations of the data require the scientists to have a sufficiency of local knowledge that is not represented within the database itself, what is the benefit of making the global connection? For policymakers and other scientists, the GBIF data can only be a starting point, a simplified visualization of what kind of biodiversity work is being done, not even the state of biodiversity globally or locally. While the data can show where a certain endangered species has been or might be seen, it cannot say what will happen if that area changes. It cannot say whether nearby changes will affect that area. It is solely descriptive, and does not represent change either over time or in scientific understanding. This lack is not an oversight, it is merely an acknowledgement of the human components of the infrastructure, the interpretations and expertise needed to allow the data to underpin policy.

Adoption of Standards and the Infrastructural Turn

Infrastructures, be they built spaces, standards or broad sociotechnical systems, are neither unary nor temporally stable. Infrastructural work can be seen operating along two cycles: those of amalgamation and fragmentation, where first work is done to draw together disparate localities of practice, metadata, reward, etc. through standardization, translation and tool development and the second is the adaptation of the outcomes of that drawing-together to suit local condition. (Hepsø, V., Monteiro, E., & Rolland, K. H., 2009) As a given database, system, or practice 'turns to' infrastructure, it exerts some social change contingent on how it is used in practice. Successful infrastructure projects are those that result in the creations of a social reality for its participants where their individual work practice is shaped in relation to the attributes of the intended infrastructure. Basic units of knowledge like classifications are easily exposed as sites of political action. Academic structures of recognition and reward apply value to certain activities while rendering others effectively invisible. Organizational practice imports a set of values, ethics and implicit knowledge all its own. Even before work on standardization begins the negotiations and agreements necessary to allow heterogeneous groups to work together inevitably ensconces some agendas while marginalizing others. As participants and researchers are sensitized to the social effects of design decisions at the infrastructural scale a space for a more nuanced discussion of those decisions sensitive to their effects at the moment of design becomes possible – perhaps even inevitable.

"I am increasingly convinced that the study of biodiversity is far and away the most important endeavor in the history of humanity, certainly until now, and very possibly into the future as well." (Pyle, 2010) Large-scale data aggregation efforts seek to develop pieces of the infrastructures that support both scientific work and policy work – they do not represent the end of the work but a tool towards its completion. While the GBIF data portal right now does not well represent either cutting-edge taxanomic work or the controversies and uncertainties that characterize scientific work it does provide a basis for future aggregation of data and representation of that work. Infrastructure work is subject to the 'long now' of development, where concerns for future development and maintenance are considered concurrently with daily work. (Ribes and Finholt, 2009) While the GBIF data portal may be seen as not immediately complete, its aggregation methods, partnership agreements, and overall goals exist as much as a basis for future development as they do to represent current scientific consensus. In fact, the presence and possibility of namespaces for translation (flawed though they may be) makes the database itself both responsive and somewhat resilient to shifts in taxanomic convention.

Infrastructure in this mode of enquiry bears a political address – it promises, through the development of technology, a certain future, a dream of the world as subtended by and supported through new infrastructure. A new infrastructural project then is fundamentally about the future and possibility. Efforts like GBIF cannot be effectively evaluated in the present tense, as they are not designed for the short term needs of their communities, but rather exist as physical, designed instances of the values, goals and predictions of their creators and contributing communities. Inconsistencies or inadequacies in coverage are not a failing of the data itself, nor of the database, but a clear indication that the database exists in the long present of development – and will likely never be 'done' or a closed topic. In fact, in reading GBIF as infrastructure as well as database there is an evident need to reconsider what failure and success mean, infrastructurally. In fact, it would be difficult to design any infrastructure that would not appear to be significantly flawed in the 'short now.' The significant scholarly work of observation, classification, data management, and technological development done in contribution to GBIF would not disappear, even if there was no longer an entity with that name.

The GBIF data portal exists evidently in its structure, design and organization in the future, not as a completed entity in and of itself but as a resource, map and object lesson for the development of such complete records and relationships. While the ICD was never a perfect tool, as it became infrastructural to global health policy the local users came to their own practice to its use. In much the same way, while there are tensions in the designed reality of GBIF as compared to the intent of its creators, the database has the potential to be infrastructural to the work of biodiversity science as scientists begin to adapt their local practice to making use of the database. In accounting for a wide scope of use, building the database so that it can respond to change, and establishing social investment in the infrastructure of the database through national contributions and Memoranda of Understanding (MOUs), GBIF has made the first move in becoming infrastructure. EarthCube is still in the midst of this process of becoming infrastructure, and, while it remains to be seen whether in fifty years there will be an infrastructure called EarthCube supporting geoscience, the sole determination of that potential will not be on the basis of technological capacity. In GBIF, it was not the amenability of database to standard modes of analysis, or its effectiveness in presenting the intricacies of its data that encouraged its slowly increasing use. It was, instead, the landscape of sociality and organizational policy into which it was introduced. The presence of ongoing, formal agreements supporting maintenance, upkeep and development of the site in the form of MOUs coupled with the willingness of biodiversity data centers to share their data has enabled the growth of GBIF entirely outside its technical capacities. In similar way, EarthCube's turn to infrastructure would more likely be the result of how the major sociotechnical concerns of the project, including issues of sustainability and ongoing funding, rather than specific tools, data, or technology it implements. GBIF not fitting well with extant workflow was somewhat inevitable in infrastructure development, but concerted effort to make use of the database in a variety of areas supported its ongoing work. Similarly, EarthCube's persistence as infrastructure may well be more closely tied to the commitment of its membership and policy support than its specific capacity.

Conclusion

> "We need to highlight the socio-technical accomplishments, which are great."
> (Recorded during the 12/6/16 meeting of the Science Committee)

The EarthCube design process represents not only an approach to infrastructure informed by past infrastructure projects, its enrollment of communities, solicitation of input and guidance from researchers, reveal some of the mechanisms of the ongoing relationship between funded science and the policy funding that science. Projects the size of EarthCube are significant sites of knowledge production, not only for the scientific fields it intends to support but also for policymakers attempting to assess, engage, and fund that community. Throughout the EarthCube design process policymakers and scientists co-produce not only the infrastructure itself but also work together to constitute the community to be served and gauge the state of the field. While EarthCube aims towards infrastructure, it has not yet made the infrastructural turn evidenced in the uptake of GBIF as both a resource and object of work for scientists in the field of biodiversity science. Tensions in the design process, flaws in the architecture as it emerges, and initial uptake of the infrastructure are not the sole determiners of success or failure of the infrastructure as innovation, despite its orientation towards community development and use. Perhaps more important than even the technology of infrastructure are the communities, practices, agreements and standards it accretes. While individual components or technologies of an infrastructure may be ill-suited to current practiced, deeply flawed in some way, or limited in application, the ability of a designed system to turn to infrastructure seems to be more related to the negotiations, fittings and adjustments and that go into its development rather than its formal attributes. EarthCube is as much a dialogue between a community of researchers and the government entity funding their work as it is a technical system. The formal funding of enterprise governance, the early design focus on assessing and exploring different related communities and (as I will be talking about in more detail in the next chapter) the overall focus on understanding the workflows and data practices of engaged researchers in the form of use cases shows an emphasis on producing knowledge about the behavior of scientists that is pervasive to whole project. In the case of EarthCube and somewhat in GBIF, infrastructure development is as much about producing knowledge about and from a particular community for policymakers as it is about developing useful technology, standards, or resources. Funded science starts with monitoring and institutional memory, and large-scale ongoing projects like EarthCube work to produce knowledge about science just as much as they produce scientific knowledge. In the realm of infrastructure policy stands very close to science, seeking to understand even as it enables, and the practice of knowledge production among policymakers has a tendency to produce and constitute communities even as it does so. In the next chapter, I discuss in detail the design process of the BCube middleware broker, with special attention to the concerns of organizational policy and funder's wishes, and their impact on the project.

CHAPTER 6: BCube, Middleware, Innovation and Reciprocal Subtension

I have deliberately not defined the separation between the realms of science and policy: rather than relying on rather than relying upon a notion of science and policy that is exclusive to particular groups or part of the identity of a particular actor I use these terms to point towards an orientation rather than limiting it to a particular conception of a group. And while I've earlier explored way in which these lines being blurred (particularly when looking at their activities outside of motivations, epistemology and ontology) by shared interests in knowledge production, conservation, participation in management regimes, and concerned with issues of monitoring and assessment, there remains a power relationship between the two groups particularly in the space of funded research. Even as the community of interested scientists constitutes themselves in response to the knowledge-producing activities of the NSF, there is still a moment in which some projects are selected and some are not. As funding, professional success, and in some cases the livelihood of the scientists may hinge on the decisions made by the NSF, their expectations and agenda carries significant weight. While in previous chapters I have largely been exploring the relationship between policy knowledge production and scientific knowledge production from the perspective of legislative and regulatory interest, in this chapter I take a closer look EarthCube as a whole, and the particular project in which I was embedded, The BCube Middleware Broker, negotiated a relatively well-established organizational and governmental policy landscape in the form of ongoing NSF engagement with the EarthCube cyberinfrastructure project.

This section is largely based in an account of my observation and participation in the BCube Middleware Broker Building Block of EarthCube. During my time working with the BCube team I found my role shifting several times, from what was essentially expected to be a user-testing and design assistance role to a more analytic/sensitizing one as various coordination and control issues arose that limited the development and implementation both of the broker itself and of the written Science Scenarios that were intended to drive its development. Though the BCube team was planning for an agile development scenario, the change in licensing of the brokering middleware from the expected open source availability to what more closely approximated a ticket system reduced the ability of team members to contribute to software development, slowed the implementation process, and changed the group of scientific scenarios chosen for development. In addition, policy-oriented concerns such as organizational policy on access and authorization within data centers, the existence of parallel building blocks muddying the waters of how to engage with EarthCube from the perspective of unfunded projects, and issues of community engagement and time commitment among funded members consistently slowed or abated development and implementation of the broker.

Middleware refers to a class of software objects that are intended to operate in between two other software objects to support their interoperability. Usually providing an "abstraction layer between low-level computational services and domain-specific applications," (Bietz, Paine and Lee 2013, pg. 1527) middleware in cyberinfrastructure most often is implemented as a collaboration between domain scientists and technology developers, and provides a basis for the mediation of standards between data centers or computational resources and various analytic applications. Middleware is almost universally networked, and focused on resolving issues of distance, in data format and selection as well as in access. (Bernstein, 1996) As related in Beitz, Paine and Lee (2013), academic middleware development in academic settings has its own particular set of issues that differ widely from the hierarchical, authoritative structures of corporate development. In academic middleware development, “small and relatively autonomous teams,” propose work that is largely funded through grants, and their “metrics of success had more to do with providing novel solutions (or enabling transformative science) than profit.” (Bietz, Paine and Lee, 2013, pg. 1529) Not heavily discussed in Bietz et al.’s analysis is the selective and evaluative role that policy, in the form of government funding agencies, holds in this design process. In the particular case of BCube, there is a complexifying factor in that the overall architecture and ‘look and feel’ of the infrastructure in which their middleware would be implemented was very much undefined throughout their project’s lifetime. EarthCube, with its planned ‘middle-out’ design model and large number of relatively independent funded building blocks working in response to a shared set of user concerns drawn from end-user workshops, presented the BCube middleware broker as largely one among many solutions to potential issues of interoperability, data access, discovery, or standardization issues. “If we take a sociotechnical approach, we see that it is not just a matter of interoperable software, but that there is a meshing of work practices that must also occur simultaneously.” (Bietz, Paine and Lee, 2013, pg. 1535) As throughout the life of BCube it was unclear exactly what role the broker would have in the planned EarthCube cyberinfrastructure. There were significant barriers to entry in the forms of adjustment of work practice. Implementation, as we will discuss in a bit more detail later, had some difficulty in broaching the existing ‘infrastructural stack’ through which participating domain scientists and technologists were already working.

Reverse Salients and Innovation

Originally formulated by Hughes in *Networks of Power,* the concept of reverse salient is inspired by its use in military studies: the reverse salient is the part of a battle line that lags behind the main force and is often the weakest part of the formation. In technology studies, the reverse salient is a component of a system – usually an older technology - that restricts the pace of development within that system. Hughes’ example is that of Edison and his work on the lightbulb: the non-durable experimental filaments available in 1875 were a reverse salient of the incandescent lightbulb, the point at which development was lagging in a way that was so obvious it pre-supposes the formation of a critical problem. Reverse salients are to some extent a tool of historical analysis, but they can serve to invert the characterization of traditional narratives of innovation. When focusing on the reverse salient, innovation becomes to a certain extent assumed – something will always be advancing somewhere. If innovation is inevitable the question changes from “how do we innovate?” to “what stabilizes our technical landscape?” and “where ought innovation to occur?”

New and developing cyberinfrastructure can be seen as a network of reverse salients which might potentially serve as a grounds for the emergence of innovation. With a focus on assembling extant, successfully working groups into a larger collaborative space, scientific infrastructure projects fall out of the usual flow of innovation. The infrastructure intentionally lags behind the innovation it hopes eventually to support; infrastructure is designed in the ‘long now’ and simultaneously attempts to fit in with present work practice while building toward future collaborations, technologies and uses. (Ribes and Finholt, 2009) While there is occasionally systemic novelty in a new bit of scientific infrastructure, it is assumed that innovation will take place in the margins of the infrastructure rather than at its core. What is innovative for infrastructure is in its ability generate agreement on methods, tools, and technologies. Scientists, data professionals, and technologists are expected to be doing novel work, developing innovative technology and techniques regardless of the presence of new infrastructure projects. Largely individualized and local groups are sought out as potential collaborators in large-scale projects, and leveraged for their innovative capacity. Scientific infrastructural projects assume that innovation has, is and will happen regardless of the presence of that infrastructure – they seek to expand communities, improve re-use and communication among distributed groups.

By assuming innovation will happen somewhere, eventually there is the possibility the focus of innovation studies can be moved from studies of how to make a single project succeed to improving our understanding of how best to respond to the emergence of innovations from an organizational, social and technological perspective. The reverse salient can be used as a site of problem development in its role as a signifier of components that will readily respond to investment as well as a tool for evaluating an innovation as it is in development. Infrastructure studies often refers to the concept of the inertia of the installed base – a classic example is the persistence of the QWERTY keyboard layout despite a general lack of evidence that it reflects the most efficient or effective key organization for typing. (Star and Ruhleder 1996) The reverse salient problematizes the inertia of the installed base, and prompts the researcher to investigate just how that inertia limits the advance of ‘forward salient’ technologies.

The bulk of this chapter will be an account of the work I observed being done towards implementation and expansion of the capacity of the BCube middleware broker with particular focus on how that broker worked to negotiate not only between software systems but also the communities, interests, and activities of expected collaborators. Effective implementation of the BCube middleware was never a question of its capability as a piece of software, it was both enabled and limited by a set of organizational and funding realities. Organizationally and in terms of work practice, the BCube project was somewhat characterized by diffusion – most project members were funded at small portions of their time, work was undertaken in a variety of different areas that assumed eventual implementation as part of the EarthCube infrastructure without a clear organizational pathway to its adoptions, and attempts to be an ‘everything for everyone’ solution to a set of problems that, despite being formulated in a similar way, differed in key ways across its potential different application domains. At the end of the day, the story of BCube was not one of software development but of the assessment, interpretation and outreaching of the broker itself as well as the project as a whole into an already-established landscape of work practice, standards, and competing techniques for resolving similar problems. In many ways, the issues faced by BCube in its implementation were similar to those explored by Bietz et al. in their discussion of moving middleware from a particular application to a more broadly-applicable system. (Bietz, Paine and Lee, 2013) However, while Bietz et al. mention difficulties arising from the need for change in an infrastructural system to propagate across an infrastructural system, their account does not reveal something I saw that was vital to that dynamic – the hierarchies of the infrastructural stack.

In the mode of Barnes, Gartland, and Stack (2004), another way to look at the concept of the inertia of the installed base is through the concept of path dependence, where previous knowledge trajectories are more likely to be re-implemented even in situations where a more novel solution might work better. Path dependence is a short way of saying 'history matters' (though its implications are much more interesting than that) and provides some insight into the positive feedback mechanisms through which certain behaviors, solutions, and modes of work persist beyond their period of initial establishment and effectiveness. (Barnes, Gartland and Stack, 2004) New infrastructure projects are not simply inserted into an ecosystem of relatively equivalent infrastructural entities – certain data sources, pieces of software, standards, or even workflows can be seen to sit 'above' or 'below' others in terms of resistance to change and/or the cascading effects that change in that entity would cause. In general, be it interpreted through path dependency or the inertia of the installed base, infrastructure resists substantial change. It becomes a function of potential benefit, policy and organizational support or mandates, and substantive novelty.

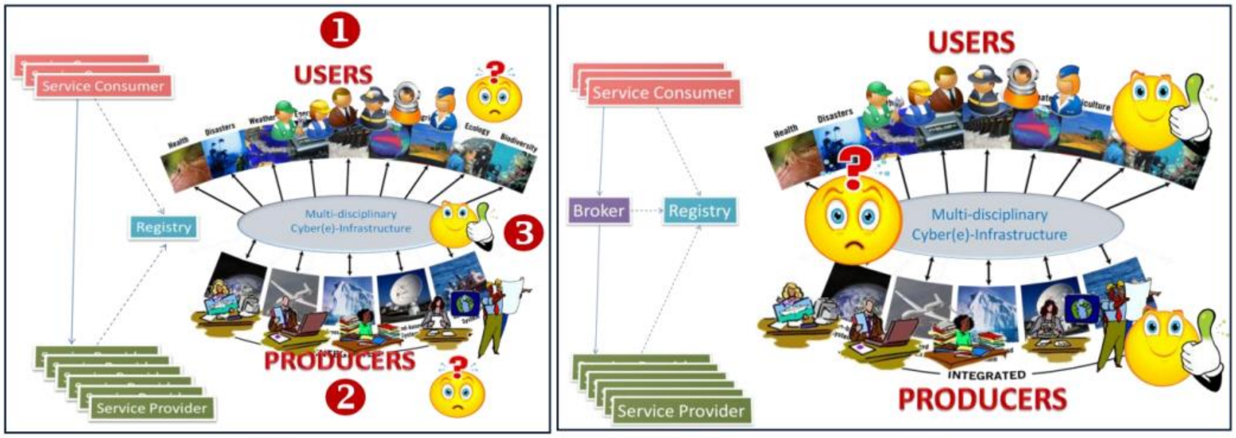
Throughout my time participating in and observing the work of BCube I saw efforts on the part of the BCube team to negotiate not only organizational policy in the form of data access methods, authorization procedures, and the attempt to account for a landscape of data-collecting activities with limited self-representations but also a larger policy space represented by the funding mandates and engagement by the NSF in the ongoing process of

EarthCube development. While not fully accounted for in the dynamic of the inertia of the installed base, there was significant tension in the development and implementation of the broker that has precluded its inclusion in the public-facing “EarthCube Tools Inventory” as prepared by the EarthCube office and presented on its website. (EarthCube Tools Inventory, 2017) These tensions reflect concerns that largely have little to do with the technological capacity of the broker, nor the effectiveness of team members in designing and implementing software solutions to particular problems, but instead the policy and infrastructural landscape into which it was intended to be implemented. BCube, which as middleware sits between the community of users and the service and data providers, had some difficulty on both levels of engagement – potential users not already engaged in the project had little incentive to adjust their workflows to incorporate brokering; similarly, data and service providers lacked sufficient incentive to modify their own access practices in order to accommodate brokering. Throughout this chapter I characterize the work and goals of the BCube project team; examine work-orienting concepts like novelty, innovation and agility; explore the broader landscape of building block projects relative to broker development; evaluate self-constitutive activities like use case development; and finally, propose a concept – reciprocal subtension – which could help in evaluating the amenability of an infrastructural system to support substantive change through an examination of the propagation of standards and extant levels of policy support. What I found in BCube, in part, and EarthCube as a larger entity, was a growing ‘expansive middle’ of development, where each new implementation, piece of software, or engaged community revealed mediating, enabling, or negotiating work. While the BCube broker was characterized in early meetings as sitting between users and service providers, work on the broker revealed ‘new middles’ between the broker and its intended end points that themselves required significant work.

Designing the broker, defining the project space.

Information brokering via middleware, as it was claimed by the BCube design team, would help resolve a significant number of these issues by reducing the need for constant coordination between data/information producers and the users of said data. Users could, through a broker, adopt new standards, work in changing data formats, or include social technologies (like web 2.0 annotation, for example) without significant commitment from data producers – it would only require a new function of the broker in order to effectively translate between the needs of the users and the established practices of the service providers. (Figure 6) Under the brokering plan, the broker is a third entity in the user/service architecture that would permit variability in standards, formats, and other data characteristics to “realize all the necessary mediation, adaptation, distribution, semantic mapping, and even quality checks required to address the complexity of the cyberinfrastructure.” (Nativi, et al. 2011) The broker, as a whole would be designed to

> “Keep the existing capacities as autonomous as possible by interconnecting and mediating standard and non-standard capacities... Supplement but not supplant systems mandates and governance arrangements... Assure a low entry barrier for both resource users and producers... Be flexible enough to accommodate existing and future information systems and information technologies.... Build incrementally on existing infrastructures (information systems) and incorporate heterogeneous resources by introducing distribution and mediation functionalities to interconnect heterogeneous resources” (Nativi et al. 2011)



(a) (b)

Figure 6: Implementation/Maintenance load for current CI v. the Brokering approach.

Even at the moment of project initialization, the broker was a functioning piece of software with a history of successful implementation – brokering had been incorporated into projects like EuroGEOSS, GeoViQua, was funded by GEOWOW, and was under development for UncertWeb and had demonstrated effectiveness in providing an access point to the THREDDS data server via ISO/OGC resources. (Nativi et al. 2011; Nativi et al., 2011a) One of the major claims in this initial white paper was the reduction of complexity for both user groups and service providers – the broker would hopefully help to simplify data access across a variety of particular needs. However, simply mediating data exchange did not comprise the entirety of the scope of the project: as the broker had already been to some extent implemented in software significant work was also planned for creating a data crawler and registry of data sources, preparing an auto-configurator to effectively access those sources, and preparing and demonstrating a number of scientific use cases mediated by the broker at varying levels of complexity. In my observation, the bulk of coordinated group activity was directed by these scenarios, with software development work occurring largely independently (though with coordination of its own in the form of a Configuration Management Board).

Back, once more, to design

I was a participant observer in the EarthCube infrastructure project for approximately three years, with engagement in the liaison and science committees ongoing. My primary site of research during that time was as a member of the BCube building block team – an intended component of the overall infrastructure centered around the development and implementation of an automated data broker and discovery engine. The BCube broker was novel in its intended inclusion of a web crawler and automated configuration system – moving beyond the standard middleware premise of fully invisible operation the BCube broker was to have a web crawler and auto configurator that would allow it to access and broker data from new sources without needing a work ticket or any particular labor on the part of either data provider or user. While, for various reasons that will be discussed elsewhere, at the end of the project the BCube broker was not fully implemented, the team was able to get the broker, auto-configurator and web crawler working, even if only over limited data sets.

The design and implementation plan of the broker revolved around the development and implementation of a series of use cases, called science scenarios. The science scenarios were in three stages: those requiring only a single source of data and relevant primarily to just one 'domain' of science as represented among BCube membership and partners, those incorporating a second domain of scientific practice, and finally those scenarios that would be relevant to many domains of earth systems science simultaneously. Each scenario was derived from an interview process where a partner researcher was interviewed twice by members of the BCube team – once by a data scientist specializing in the technology of the broker, and once by a domain scientist working in a similar field. The goal of doing each interview twice was itself twofold: to bridge inconsistencies in data and technical literacy between interviewed scientists and those who would be implementing the broker as well as providing a better understanding of the relevant science to the technical implementation team. It was hoped that the scenarios would be more completely implemented if potential gaps in knowledge between the more technically focused and more domain focused members of the team were identified early in the scenario development process. As might have been expected, the outcomes of this process were a bit more complex than the hoped-for knowledge sharing, but resulted in detailed and relevant science scenarios whose implementation was limited primarily by issues external to the broker's development. BCube also developed and ran a web crawler to collect metadata and configuration details from potentially thousands of individual data sources (with sufficient XML documentation) and an auto-configurator to allow these crawled data sources to be described and accessed by the broker without direct intervention. As such, the BCube broker was intended to aid not only in the completion of particular research tasks requiring specialized data access and transformation, but also to aid in discovery and evaluation of new data sources, eventually being able to incorporate and access various forms of 'dark data' - Excel sheets, spreadsheets and other loosely-structured data that generally requires substantial amounts of work. As work on the project went on, a member of the BCube team also did some experimentation with using some functions of the broker to create more reproducible models by storing data access parameters so that future models could be made with identical data even in more actively changing data-sets. The team also created a template, or development kit, for building accessors to data sets with particular or difficult access policies in order to facilitate automated access of data sets that, for example, require separate logins for each access and retrieval, limited the number of connections available, or otherwise were middleware-resistant.

As a member of the BCube team my task initially was to perform user testing and evaluation of various implementations of the broker towards particular use cases – which we called science scenarios – in practice. However, due to difficulties in implementation that will be discussed elsewhere, there was never a valid, testable user group that could be observed using the broker. In fact, only a couple of the science scenarios were able to implemented at all, and those occasionally in fairly limited ways. As my role in the BCube team moved away from direct user testing (due to a lack of opportunity) I took on more of a sensitizing and evaluating role, providing feedback on science scenario interviews and protocols as well as presenting on social and organizational issues and representing the team as a member of several EarthCube governing committees. During the three years I was a member of the BCube team I attended and presented at two EarthCube all-hands meetings and served as a member of the Liaison and Science committees, and delivered feedback on governance proposals to the EarthCube Test Governance committee. Again, in these roles my goals were joint – both to serve as a sensitizing influence on social concerns and to observe the processes of governance and decision making as they occurred.

EarthCube, like many other infrastructure projects, is broadly distributed without a central office or site of work. This presented a particular set of methodological challenges that, while hardly unique, needed a set of workarounds and solutions specific to the distributed nature of the site. As such, I followed the ‘scalar devices,’ (Ribes, 2014) though at the time I didn’t conceive of them as such. I pur;sued those moments where people came together to discuss individual work, to coordinate future work, and to assess, understand the community as a whole. I attended and recorded meetings both within and without my ‘home group’ of BCube, looking towards the larger community of EarthCube through its work. As there was no physical site to observe design work or negotiations, the majority of my observations were at some level of remote. EarthCube, in its early design phases, mostly took shape in three ways: the preparation and circulation of documents by individuals and teams, face-to-face conceptual and end-user workshops, and in WebEx meetings of several committees made up of representation from funded projects mostly guided by the Demonstration Governance building block. While researching EarthCube I participated first in the Coordinating Committee, which was a temporary leadership group established by the Demonstration Governance building block. As part of the coordinating committee I participated in the drafting and revision of the original EarthCube charter, as well as observed the initial set of negotiations around the particular structure the organization of EarthCube would take. As the coordinating committee (which had only a couple formal meetings, and was also briefly called the liaison committee) gave way to the current committee structure of EarthCube I took up membership in the Science Committee and Liaison Committee. I chose to join the Science Committee both due to the needs of

BCube and due to the similarity of initial set of goals to the activities I was observing within BCube.

BCube's development plan, and a significant part of the mission of the science committee, revolved around the creation of use cases to effectively guide development and organize work. This issue of the coordination of development was not particular to BCube, but seemed to pervade EarthCube both at larger scales and smaller. Among the 57 initial participants in BCube, only 19 were funded. Among the 19 funded participants the majority were funded for some fraction of their time – many in the range of 10-15%. This was not unusual among the EarthCube participants I spoke to. There were very few participants in EarthCube as a whole or in the projects I observed in action for whom the infrastructure was the center of their work. Most dedicated some small portion of their time to the infrastructure, and many were similarly committed to a variety of projects, each of which they were responsible for in small increments of their total time. As such, the issue of coordinating work was complicated by the need to navigate the other professional commitments of project members, and solicit voluntary work from those who were totally unfunded but still interested. This issue of juggling multiple commitments is discussed in Freidberg and Crozier. (1980) Additionally, in BCube, modifications to the software backbone of the broker that was being implemented and tested as an infrastructural component of EarthCube were performed by a team of researchers in Italy through a ticketing system. Changes to the broker needed to be prioritized and specific – there was little room for open experimentation or modification, and some significant turnaround time between the identification of a potential software modification and its actual implementation. And so the science scenarios became central to the coordination of team activities and the prioritization of the limited available work time of funded members and the software development team. While there was some room for individual members to experiment with the broker, major modifications needed to be routed through Italy and took some time – modification of the brokering software was both valuable and restricted, and so became the subject of some burden of planning efforts. The development plan for the BCube broker was to implement specific, representative science scenarios at increasing levels of interdisciplinarity in order to demonstrate its effectiveness and potentially serve as a portion of the testbed for the overall EarthCube architecture, and so those science scenarios were a site of significant organizational work across the active team.

'Agility' as a broader organizational concept

Among the BCube cube there were a number of simultaneous, relatively independent but interrelated projects. One group was working on developing and producing science scenarios that could be used to drive development and determine priorities, another was implementing a web crawler intended to discover data sources not previously indexed or accessible to the broker, and a related group was developing an auto-configurator that would have allowed the broker to access crawled data without needing the direct intervention of a team member or coder. In addition to all of these, work was being done on developing and implementing a test bed where users might be able to use and evaluate brokered resources in real time while providing monitoring information on efficiency and effectiveness in data transfer. There was evident in this organization and the focus on agility the expectation that despite this being a shared venture that most work would be done relatively independently. The core of agile design is proposed as:

> “We are uncovering better ways of developing software by doing it and helping others do it. Through this work we have come to value:
> •Individuals and interactions over processes and tools
> •Working software over comprehensive documentation
> •Customer collaboration over contract negotiation
> •Responding to change over following a plan
> That is, while there is value in the items on the right, we value the items on the left more.” (Beck, et al. 2001)

While some aspects of this design philosophy revolve around the assumption of particular problems with software development prior to instantiation, the agile method largely is concerned with providing a justification for change in processes and limiting the amount of time spent among project members in negotiating bureaucracy – “comprehensive documentation,” “contract negotiation,” “processes and tools,” and “following a plan” all seem to speak to freedom from management requirements in favor of support design that changes rapidly in a relatively unknown and unregulated space. Agile development is also test-driven, revolves around iteration, is informed by ‘user stories’ rather than explicit use cases, and work is oriented around regular ‘scrums’ where leadership endeavors to resolve issues encountered by its members, and seeks to eliminate entrenchment in certain fomalisms in favor of responding rapidly to a changing landscape. (Martin, 2002) Agile development has become something of a bandwagon, in the mode of Fujimura (1988) – a trend that has moved out of its origin in corporate software development and into a variety of disparate spaces.

Agility bears some interesting parallels to the concept of the broker and middleware itself – this takes us back to my characteristic metaphor of EarthCube as focused around the 'expansive middle'. Middleware and brokering is modular, enabling software intended in some ways to delegate to software what otherwise would be done by a human broker or data provider. Not only does the BCube middleware support modularity by being relatively independent to the front-end data access software that a given researcher or group is using, it also allows for format and to some extent meta-data neutral access to a variety of data sources. While most scenarios were not fully implemented, work was done to establish and describe equivalence between disparate data fields in different formats; to be able to reconfigure data on the basis of a variety of standards and data formats; to allow for the rapid translation of gridded data to point data (a task that I was told otherwise is about six months work for a grad student); and to implement new data connections as the web crawler ran. Leadership, in accordance with the single-responsibility principles of agile development, worked in mediating and coordinating roles where individual developers largely self-directed their efforts and techniques.

It is in BCube's implementation of agile development that we find a further feedback loop of the sort I have been discussing throughout this document – one that reveals how the project was expected to be guided as well as how the input of engaged groups might be incorporated into software design. Figure 7, below, was presented at the kickoff meeting of BCube, and shows that not only were the technical requirements of the broker expected to respond to input from geoscientific and CIS communities, but also the broader vision for the broker itself. Also of note in this diagram is a relationship that did change during my time on the project – the expectation that science scenarios emerge solely from the input of geoscience communities. While initial scenarios and data landscapes were presented at the kickoff meeting by a variety of geoscientists, as the process of developing and implmenting scenarios went on the technologists of the group became more engaged.

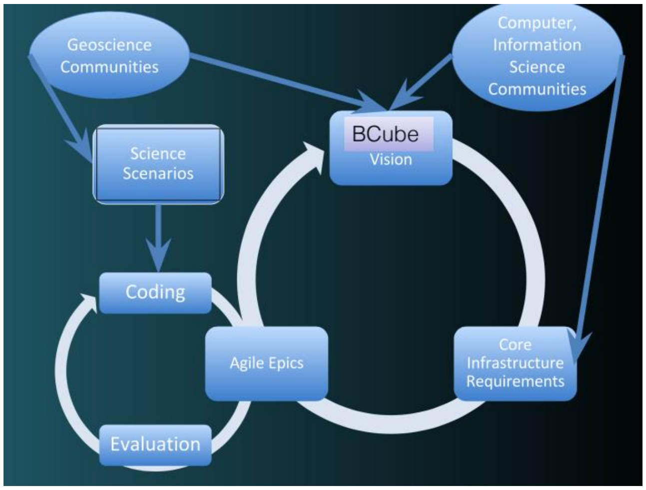


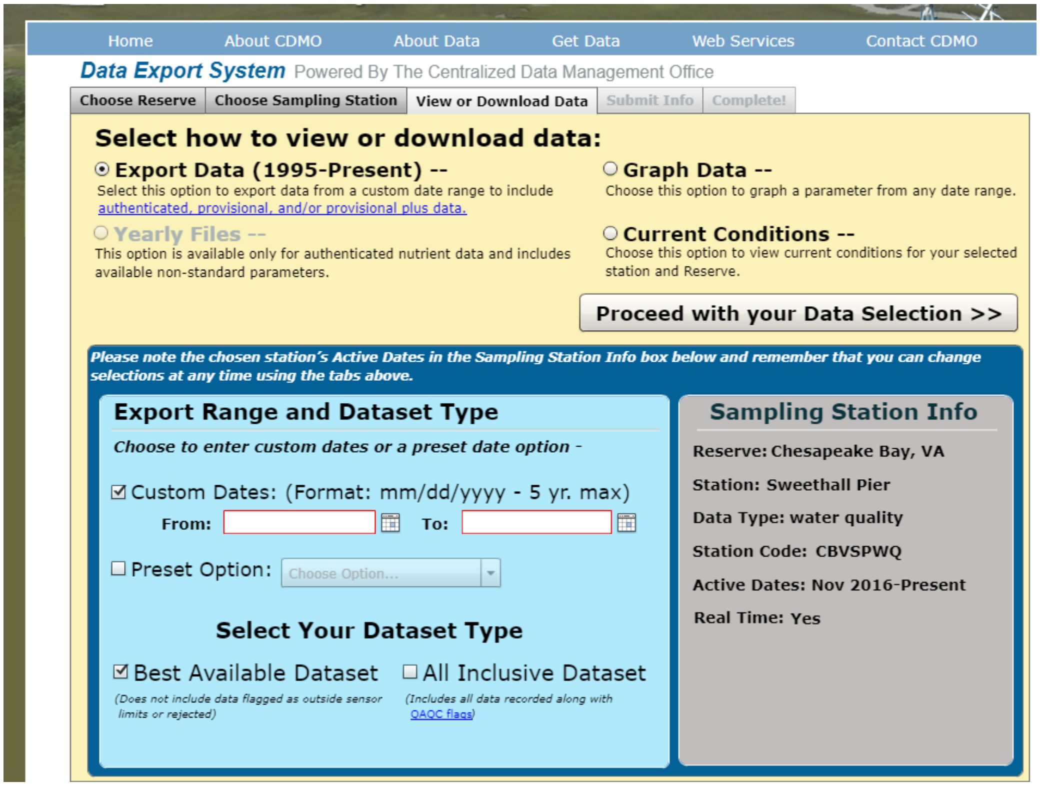
Figure 7: Agile Development Feedback Loops and Design Iteration in BCube

To some extent, agility requires an environment in which movement and change is possible. Established data access policies and idiosyncrasies of existing modes of access limited scenario implementation; lack of access to the source code of the broker instituted what was essentially a ticketing system for modifying the broker; and competition among the various similar activities within the EarthCube set of projects limited access to the altruistic participation potentially present among database managers and others who might be able to ease implementation of the broker within particular data sources.

While the science scenarios were intended to produce a particular set of design priorities, limitation on the independence of team members as discussed above had a tendency to focus work around singular scenarios, often chosen on the basis of being able to produce a demo or representation of the activity of the broker for particular audiences, workshops and conferences. These were not necessarily chosen on the basis of strategy – where reaching out to data managers and presenting to those who might be able to ease the access of the broker may have been strategically valuable it was de-prioritized because of the EarthCube mandate to participate and collaborate with other building blocks. As such the bulk of the work ended up being the management and focus of resources towards low-level and short-term implementations of the broker towards the goal of producing demonstrations of its capabilities. However, this was a bit misleading as the broker is middleware – the demos of its operation tended more to demonstrate the front-end interfacing software or the database rather than the middleware itself. This limited potential collaborations as the operation of the broker was (properly and intentionally) invisible – it seemed to be doing nothing in the demonstrations, which, though ideal in certain ways for properly developed infrastructure, made it much more difficult to encourage participation. As implementation went on there was a discover of 'new middles' between the broker and various service providers and data sources. Properly functioning accessors were a consistent problem alongside the emergence of needed modifications in the brokering software itself – these came up throughout the life of the project, and even at the end of its development not all were resolved. In the final recorded meeting of the Configuration Management Board (CMB) new issues with accessor development were still being exposed and patched. "Further tests reveals*[sic]* other issues." (CMB Minutes, October 2015) is a statement characteristic of the corpus of CMB activities in its simplicity. We had a number of motivated participants, expected to work independently, and with the goal of producing something out of the broker project. As development on the software was bottlenecked through the ticketing system due to a subset of the broker development group not making source code for the broker available or open, as was expected at the outset of the project, group members began to focus on other areas. One piece of this was the replicatability aspect of the broker as mentioned above, others focused more fully on the science scenarios. At the end of the day the majority of the design work of the broker rotated around the science scenarios (similar to use cases) produced, and more specifically implementation efforts revolved around an expansive middle object – the data accessor. Data accessors, tautologically, are pieces of software or plugins for existing software allowing access to a particular set of data. Accessors in BCube accounted for format, metadata, and authentication specifically for each linked data set. These accessors underwent development relatively independently of the development of the brokering software itself, though there were occasions where the accessor required certain changes in the broker itself, which were delivered back to the BCube team as patches. In some cases, the accessor needed to be negotiatied with the database manager in order to work - the PANGEA dataset (https://www.pangaea.de/) , for example, required the intervention of project leadership in order to be implemented due to its authorization procedures. While it was implemented in a version of the broker that is part of EuroGEOSS, discussions about implementing it similarly in BCube stalled at the end of the project – the final set of minutes of the CMB include the open question of whether it was ever finally completed. (CMB Meeting Minutes, October 2015)

In similar vein, implementation of accessors occasionally clashed with credential/authentication procedures when working with some established data sets.

NERRS (National Estuarine Research Reserves) data was a particular example of this problem. While built accessors were capable of accessing NERRS data, there was an issue using that data in order to implement the planned science scenario, which focused on evaluating long-term changes in coastal ecosystems. While data could be drawn from NERRS using the accessor, it tended to fail to get the long-term data needed by the affiliated scientists who developed that scenario. At issue here was the NERRS use of its own web service and interface for delivering data, and only allowed data retrieval from a maximum of 5 years at one time. (Figure 8)



Data Export System Powered By The Centralized Data Management Office

Select how to view or download data:

● **Export Data (1995-Present) --**
Select this option to export data from a custom date range to include authenticated, provisional, and/or provisional plus data.

○ **Graph Data --**
Choose this option to graph a parameter from any date range.

○ **Yearly Files --**
This option is available only for authenticated nutrient data and includes available non-standard parameters.

○ **Current Conditions --**
Choose this option to view current conditions for your selected station and Reserve.

Please note the chosen station's Active Dates in the Sampling Station Info box below and remember that you can change selections at any time using the tabs above.

Export Range and Dataset Type

Choose to enter custom dates or a preset date option -

☑ Custom Dates: (Format: mm/dd/yyyy - 5 yr. max)
From: To:

☐ Preset Option:

Select Your Dataset Type

☑ Best Available Dataset
(Does not include data flagged as outside sensor limits or rejected)

☐ All Inclusive Dataset
(Includes all data recorded along with QAQC flags)

Sampling Station Info

Reserve: Chesapeake Bay, VA
Station: Sweethall Pier
Data Type: water quality
Station Code: CBVSPWQ
Active Dates: Nov 2016-Present
Real Time: Yes

Figure 8: NERRS Data Export System showing date range limit

While technical issues also emerged with the particular web services NERRS employed – which was a combination of WSDL, which is an XML document describing a web service, and SOAP (Simple Object Access Protocol) which allows a user to perform operations remotely by sending commands in an XML document over a transfer protocol like HTTP or SMTP. In practice, the WSDL document describes the operations that can be performed, and SOAP is the means by which those operations are transferred to the server. This would comprise s still further 'middle' of access for the broker – that of data source-defined access procedures. WSDL/SOAP is designed in part to only allow expected uses of the data it provides access to, and does so through XML documents following their particular formats. In terms of the broker, being able to 'read' the WSDL document well enough to process a SOAP request was a significant effort in its own right, requiring adjustments both to the designed accessor and broker itself. In addition to technical issues in parsing and building the XML documents automatically in order to process requests through the broker, the policy issue of only allowing retrieval of data in 5-year increments represented significant barriers to the implementation of the scenario. Even discovery of metadata and formatting did not initially work with broker accessors, and data access never quite managed to work properly. At the end of the BCube project cycle, NERRS access was still buggy and failed to return the expected or correct data – even some unexpected bugs like the broker passing blank xml documents began to crop up as other issues were resolved. Leadership had to negotiate, with little headway, with the managers of this database in an attempt to change access procedure in order to support the long-time goals of the particular scenario, and to plan for and work through alternatives means of getting access to the same data. "[X] negotiating, invoking "Open Data" policy. Fallback is to get data from USGS, [Y] says it may be accessible through VT broker" (Configuration Management Board Minutes, February 2016) And here we seepath dependency effects – the policies and procedures 'built in' to the NERRS data service were closely linked to certain methods of access and specific formats by which they expected access to occur. Negotiating the stack might have required significant change or individually curated data access in order to be successful and, while the BCube team was working on workarounds, negotiating the infrastructural stack already in place represented a significant amount of work and ongoing negotiation not initially accounted for in project plans.

Science Scenario Development

While implementation saw what are largely expected hurdles in development much of the effort of BCube revolved around the development and description of specific science scenarios intended to be a guide for development effort. Each data accessor and new patch was linked to a particular scientific scenario, which were descriptions of a scientific question and the data needed to answer that question. Scenarios were initially grouped into three ‘phases’ of development based on degree of cross-domain collaboration expected. Scenarios were expected to be extremely collaborative in their development and conception, with many parties expected to contribute throughout the scenario design process. (Figure 9) And while the flow chart pictured above was not exactly followed, scenario development did pass through many hands and many revisions prior to being implemented. Phase 1 scenarios were expected to be single-domain, focusing on the research area of the participating scientist who proposed it, phase 2 were scenarios requiring collaboration between two domains, and phase 3 were multi-domain collaborations.

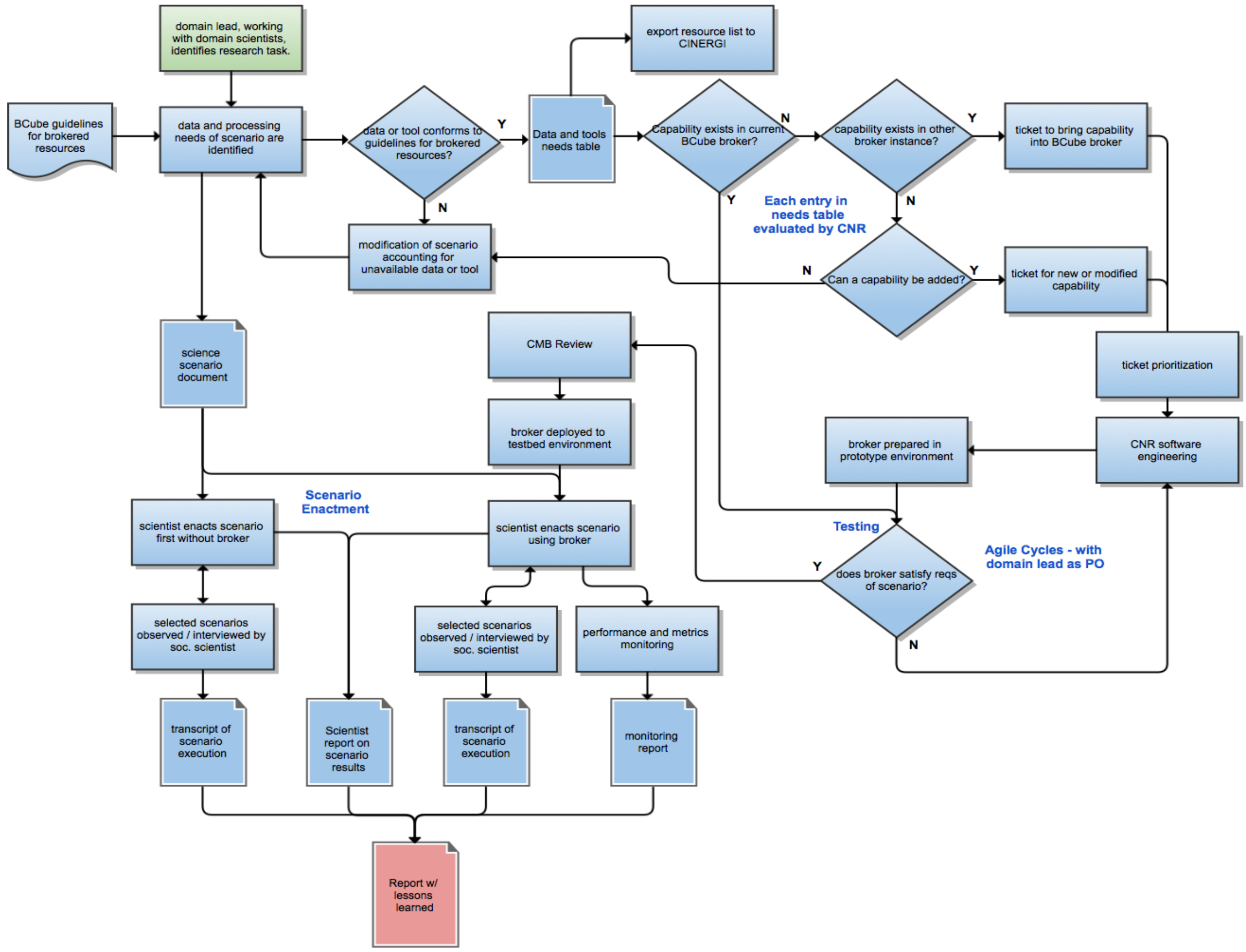


Figure 9: Scenario Development Flow Chart

The interview and representation process resulted in nine science scenarios, all of which were 'phase 1' implementations of the broker that, while limited to single domains, often included the incorporation of disparate data sources. As the work of developing accessors, negotiating with data centers, and modifying the broker itself through patches occupied significant attention, not all of the scenarios captured here were able to be implemented, and some were implemented only partially. Scenarios included were to be drawn from

Polar, Weather, Oceans, and Hydrology scientists with a goal of integrating across these domains. (Figure 10)

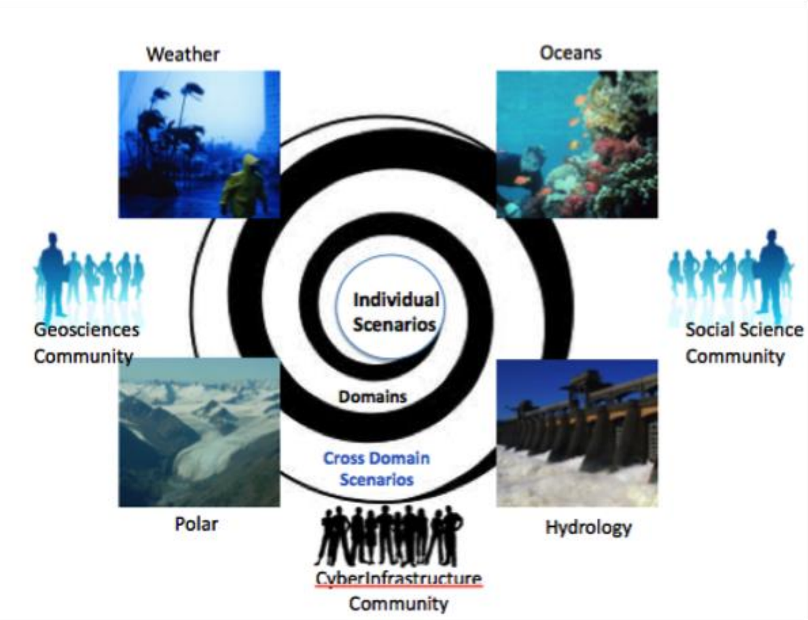


Figure 10: Spiraling up from individual scenarios to cross-domain scenarios.

The science scenarios were developed through a structured set of interviews with nine participating domain experts. These experts were usually (with two exceptions) not funded members of the team. They participated in some organizing activities and were present at the kick-off and all-hands meetings, but their primary contribution to the project was participation in scenario development. There was some hope early in the project that the domain experts would also be able to participate in testing the scenarios once in place, but few scenarios (and none of the interdisciplinary scenarios) were successfully implemented by the end of the project's scheduled life. Despite the initial plan of performing each interview separately twice, only three scenario experts were available for doubled interviews – the remainder were team-interviewed with both a domain expert interviewer and computing expert interviewer. Each interview was relatively structured, and collected a general narrative of the science goal of the scenario, a summary of how the scenario would be implemented, the preconditions necessary for the scenario to be implemented, an example workflow and general notes. There was some range in the level of detail and complexity in the interview results (responses were coordinated between the two interviewers into a single response document, limiting the possibility of comparison). To get a better idea of the sorts of requirements captured and the kinds of scenarios being developed, we will take a closer look at two representative science scenarios, once of which was created through the planned two-interview process, and one that was the result of a single interview.

TITLE: GCM Chemistry North American Corroboration GOAL: “GCCM (Global climate chemistry models) are used to examine the ways in which climate effects air chemistry and the ways that surface emissions affect the overall earth-atmospheric systems.The GCCMs are predominantly used by atmospheric scientists, though there is a desire to integrate data from sensor networks not traditionally used in the atmospheric modeling fields whose formats are commonly processed. An example of such a non-traditional network that is of current interest is the EPA air quality monitoring network.”	TITLE: Distributed network predictive skill expert "The science goal is here is to: Deploy and utilize a distributed observation and ensemble modeling to study a wide a range of science questions dealing with the eastern seaboard during fall 2014. One unifying theme for the effort, but is not limited to, is understanding the dynamics of the Cold Pool. The Cold Pool is a distinctive along-shelf band of remnant winter bottom water over the mid and outer continental shelf peculiar to the area between the Northeast Peak of Georges Bank and Cape Hatteras that, beginning in April and May each year is defined by its low temperature after the water column stratifies due to increased solar insolation, local fresh water runoff, and reduced wind mixing. Because the Cold Pool is bottom-trapped, it is virtually undetectable by ocean surface remote sensing measurements. Thus existing pictures of Cold Pool distribution are aliased and incomplete and so one of the most poorly understood major MAB features. The dynamics in this water masses has a disproportionately large role in structuring the biogeochemistry and ecology of the Mid-

	Atlantic continental shelf. For example the cold pool defines the thermal refuge provided by the Cold Pool for cold-water fisheries species. This water masses also is directly relevant to the tropical storm and hurricane intensity. "
Summary: The scientist should be able to run a shell based utility, similar in nature to commonly used gridding utilities (CDO), on a model netcdf output file. This utility will send a http post to the broker with the the grid description (coordinates, projection, time, etc) and EPA AQMN field desired, to have the broker return a NetCDF of the field in the same grid format."	"The scenario scientist would use the broker accessing technologies to access: - access distributed data distributed through the MARACOOS asset map. - access model simulation data collected from 4 different ocean numerical models -access fish data from NOAA (this grey data is tertiary)"
Basic Flow: "1) Scientist runs GCM with surface chemistry output to be sent to NetCDF file. 2) Scientist runs shell based utility on output NetCDF with AQMN field desired from EPA. Utility extracts timesteps and grid information which is sent to broker with http post. 3) broker grabs AQMN spatial field from the point location field database for the timesteps desired, grids them per sent grid information and returns them to the utility. 4) Scientist compares model output chemistry to EPA measured air quality."	"1) Data from observational platforms and models to be collected as real-time as possible 2) Data will need made available for visualization, so the it could be used in adapting the observational strategies to allow for adaptive sampling during the experimental effort" "Data flows from models and observational platform and is discoverable to the MARACOOS data portal and the broker. The data is generally available via Thredds server."
NOTES: "For this study to succeed, it is expected that the download and preprocessing of the EPA AQMN datasets to conform to the requirements of the initial conditions and boundary conditions of the initialization and output gridded datasets."	

Table 1 Comparison of Two-person and individual interview results in science scenario development.

The GCCM Scenario, in short, was designed to use unusual datasets to confirm climate models, while the other sought to better understand the dynamics of Cold Pool, a particular ocean phenomenon that is both disproportionately influential on Mid-Atlantic ecosystems but difficult to observe as it doesn’t affect surface water temperatures. The first set of interview results (GCCM scenario) was prepared by a two-person interview group. The second (The Cold Pool scenario) by a domain scientist interviewing solo. While these are somewhat too specific to their own particular scenarios and the individual being interviewed for their differences to really be generalizable, there are a couple of interesting things to note here. First, the GCCM scenario is much more detailed in terms of data science requirements. Both input and output data formats and conditions are presented, the general queries refer to particular database fields, and overall the interview results contain a sort of specialized knowledge held by someone working directly in data access and curation. The Cold Pool scenario refers to particular data sources but overall seeks a much more general set of data, with significantly less specificity in terms of the data formats, and particular details about the database being accessed. While a couple of particular data sources are mentioned, little effort is made in describing how the data would be accessed by the broker or what technical steps are to be taken. Instead, the Cold Pool scenario is focused more on the intended scientific outcomes of the analysis of that data. It needs to be ‘made available for visualization’ to be used in adapting observation strategy, and collected as real-time as possible. The data is generally available via a particular server - ‘Thredds’ - but there is not much discussion of the situations and circumstances in which it is unavailable or inaccessible (something that eventually bore some significant consequence in terms of the implementation of the broker). The GCCM scenario references the specific transformations of the data necessary (point data to gridded in order to compare the gridded database to the point database), and the specific set of circumstances necessary for that to happen. While these few finished scenarios for BCube represent a sample that is somewhat too small to make significant generalizations about the community or data science as a whole, they do confirm a division in approaches to data use that I observed throughout the life of the project. Namely, there was a noticeable divide between those data scientists that I would characterize as *data experts* and *data savants*. Data experts (those with comprehensive knowledge of a data space) were those scientists with strong general knowledge of the data space – they were the ones who knew not only how to conduct their own scientific work but also had significant knowledge of alternative data sources, formats and methods in that space. Data savants (those who are learned in data), on the other hand, had strong specialized knowledge of their 'home' data, techniques, and methods without necessarily needing to reference or be drawn in to data in other spaces. I selected these two interviews for a closer reading in no small part because they represent this distinction in action, with the GCCM scenario showing data expertise, and the Cold Pool scenario showing data savantism. In the GCCM scenario the data itself is very much the object of the scientific work to be done: it is at its core about developing and integrating specific data into a usable model without much reference to particular science outcomes. The Cold Pool scenario, while still essentially a modeling and integration task similar to the GCCM scenario, shows a much greater focus on scientific outcomes: understanding the Cold Pool ocean phenomenon and incorporating its effects into general oceanographic models. The expert data work focuses closely on the data itself as an object and outcome – the scientific goal here is in part the incorporation of surface sensor measurement into an existing model – as opposed to the particular understanding of the world that might arise from their work.

The corpus of scenarios developed for BCube as a group are all naturally concerned with the work of integrating heterogeneous data sets using the broker to demonstrate that the brokered work can lower the barrier to entry or initial work in space; can validate or improve existing models; or in some other way work faster than extant methods (though there is a bit of an outlier in an education-specific scenario). However, only one of the scenarios had an explicit comparative goal – that of the Oceans scenario, which would have compared access over the broker to direct access from the project website. This particular scenario was the one drawing on NERRS data as discussed above, and was not sufficiently implemented for that comparison to take place.

More interesting than the design work falling into the easiest channels of change was the way the development of the science scenarios encapsulated a process of infrastructural fitting. While there was no 'knowledge infrastructure' of EarthCube at the time of scenario development, the infrastructural stack upon which it was intended to operate is visibly present in the scientific scenarios devised. As middleware BCube approaches the infrastructure from two directions, as it were – both as something that needs to be negotiated with throughout the process of design and implementation but also as an infrastructural component in its own right, negotiated with the local conditions of use and appealing towards particular scientific activities and user groups. Interestingly, the interview-driven use case development process employed within BCube also was taken up by the EarthCube TAC (Technology and Architecture Committee) in conjunction with the Science Committee in their own use case building project – perhaps not surprising considering a leadership member of BCube was also an active advocate on each of those committees.

EarthCube level use cases were similarly interview-driven, but sought a much greater level of detail than those employed within BCube. Where the BCube scenarios had a relatively limited set of fairly open descriptive fields (Goal, Summary, Preconditions, Basic Flow) that allowed for significant variation between different scenarios in terms of level of detail, the EarthCube use cases had significantly more required details. Their use case template included the categories of Actors, representing the people and institutions involved in the use case, critical existing cyberinfrastructure, measures of success, basic flow, alternative flow, activity diagrams, major outcomes, references and notes. In addition, a technical section specifically requested data sources, formats, volume and velocity, as well as variety, variability and veracity. Data standards, visualization and analytics, required software, and metadata were also described. Rather than being direct guides towards the building of particular components, the EarthCube use cases were more a process of community assessment, and were not well-linked to work funded under EarthCube at that time. “There appears to be no shortage of science drivers, but the notion that 'use cases' will distill these into manageable problems that technologists can address is a fallacy.” (Leadership Member of EarthCube, posted to Science Committee listserv 1/27/16) While each scientist contributing a use case was interviewed, they were interviewed only singly, unlike the two interviews used in BCube. However, the interview was expected to cross domains somewhat – there was the expectation that the interviews would be technologists working through the template with a volunteer scientist.

Parallel Development Tracks and Coopetition

The 24 funded end-user workshops were intended to guide the development of the various building blocks within EarthCube. These workshops addressed a broad geoscientific community in a variety of areas, and there was at least an implicit requirement when proposing a building block or conceptual design that it address some set of science drivers emerging from those workshops. There was even a workshop specific to the PIs of building blocks where the results of the prior workshops were synthesized and presented as guides for proposals and potential collaboration. However, some aspects of the parallel development and relatively undefined architectural space tended to create further barriers to collaboration on the basis of things like standards negotiation, formatting, and institutional support - "The irony here is that the structural obstacles within EarthCube resemble the very sociotechnical issues that EarthCube is meant to address!" (Posted 11/6/15 to the Science Committee listserv)

Among those building blocks attempting to resolve the end-user identified problem of data re-use, transformation and reformatting is the BCube middleware broker. From its beginning as a funded building block of the infrastructure, the BCube middleware was fundamentally complete from a design perspective; the work being done has largely been configuration and implementation of the middleware data broker framework. As data sources need to provide accessors for the broker – accessors being points of data access that allow the automated middleware to consistently and without the need for multiple authentications query the database on demand – data managers must champion the middleware technology to a certain extent in order for its implementation to succeed. The implementation of BCube was significantly limited by reverse salient of information access. Access restrictions designed for individual users don't scale to automated access; metadata and structural inconsistencies that are relatively easy to resolve at the level of individual data sets become intractable at a scale; and individuals have trouble conceiving middleware as an independent object without reference to both its GUI and its data sources.

Part of the way in which BCube was expected to work was the creation of a registry of data sources with information on their metadata, formatting and access policies in order to more easily aid data discovery – a point repeatedly appearing in the end-user workshops. Issues of data discovery, access, and integration were discussed to some extent in all 24 end-user workshops (including the final PIs workshop), and of the building blocks funded nine of them included some aspect of data discovery and integration. Three building blocks together each proposed to develop some registry of data sources and services for the EarthCube community: GeoWS, CINERGI, and BCube. In the following section I describe each of these briefly in terms of how they run parallel to each other, and then I will briefly account how each of these projects were able to collaborate (or not) through their initial funding cycle.

GeoWS

The GeoWS project, titled Deploying Web Services across Multiple Geoscience Domains, or Geo Web Services, sought to address the problem of 'dark data' by simplifying access to and creating a registry of data centers that can employ some kind of web service in providing access to their data. The GeoWS model of data access involved the use of URLs to indicate parameters of the data they are retrieving. GeoWS focused around a single use case in geodynamics, but enrolled a variety of partner institutions in order to demonstrate a more broad applicability within the Geoscientific space.

CINERGI

CINERGI, or the Community Inventory of EarthCube Resources, sought to solve the problem of data discovery through the curation of a list of available information resources, the collection of metadata and its re-presentation in a standardized interface. XML documents containing metadata, as well as other projects and specific outreach efforts went in to building this inventory of resources. In a later round of funding, CINERGI was extended to create the EarhCube Data Discovery Hub, which is a means of presenting and curating the metadata records collected by CINERGI throughout its initial funding.

BCube

BCube's broker implementation included the development of a data crawler – an automated means of scanning available XML documents describing datasets, harvesting their metadata, and producing a registry of said sites alongside an autoconfigurator that would automate access through the broker itself. The crawler was potentially able to access thousands of such documents, and could potentially result in a significant registry in its own right.

Each of the above three projects produced a registry of resources, a set of metadata documents, and a means by which the end-user workshop defined problem of data discovery might potentially be solved. Each project also was expected to engage in some mode of collaboration, and potentially would be among those tools incorporated into a final EarthCube architecture. However, from the perspective of a non-EarthCube data center, each of these would present a very similar value proposition with repetitive work and no real indication of which would have significant staying power as an available resource. In order to partner with GeoWS, access to data would have to follow their very specific URL-based query structure and the general web services design method. CINERGI required the production of metadata documents in order to be included in their registry, but did not expect any changes in access method or data presentation. BCube's crawler required some form of XML representation of data assets and metadata, but was fairly free in format that document could take, but in many cases in practice required some modification of access policy or software in order to work properly. Near the end of its project cycle BCube was able to implement an accessor for GeoWS data, but there was little ability for BCube and CINERGI to collaborate, despite several meetings attempting to do so. The XML documents produced by the BCube crawler did not contain all of the information required by CINERGI, and CINERGI's documents did not contain quite the correct information for the crawler to auto-configure. The result of these meetings was a general agreement to pursue their development in parallel and hopefully find ways to collaborate in the next round of funding.

These parallel development tracks, concurrent with the early focus on developing governance and community around EarthCube, are particularly relevant to the design of BCube as a middleware broker. Repositories and domain scientists that may have considered collaborating in the development of accessors and incorporating the middleware broker into their workflow were presented with a preponderance of similar, incompatible, and labor-intensive building blocks all claiming to achieve, in a very broad conceptual sense, the same end goals with no representation of the architecture of the final infrastructure (architecture conceptual design projects were not funded until the second round of EarthCube funding) and no assurance of the longevity of any particular building block. Why change access policies to accommodate middleware when that middleware may be gone in a couple years? There was little reason to commit to a particular solution when there was little evidence of commitment to that solution even within the larger infrastructure project. Scientists outside of BCube were not asked to commit to any particular solution, and the NSF was not planning on committing to a solution until the end of the first round of funding was complete. And yet, the level of work necessary to even demonstrate a given technical solution that may have been asked of various repositories and data managers was not trivial. Some application might require the generation of a specifically formatted descriptive XML document to make collections amenable to discovery, where another would require a different, but similar, XML document describing those datasets for use in semantic web applications. BCube itself would need a couple of repositories (like the NERRS repository above) to make significant changes to their access policies in order to even allow the development of a software accessor. The parallel development tracks worked, then, to undermine the willingness of non-EC repositories, domain scientists, and other entities to adjust their workflow, policies, or systems to accommodate any particular building block. Even BCube, one of whose major selling points included the development of accessors (modifying existing ones for repositories built to particular standards, new ones for others) such that repositories and scientists would see little change in their workflows, required some level of commitment and change to access policy in order to implement their planned scientific scenarios.

> "The Science Plan articulated several high level problems that EC might address. However, even then their diversity made them less than appealing to NSF and technologists alike -- primarily because their solution requires significant capital investment, as well as bringing together groups of scientists to articulate TO technologists what needs to be done. Moreover, there are clear planning and implementation stages to such projects (as those of us that have been involved in other big-science projects well know). By contrast EC has by-passed the planning stage (using the End-User workshops and RCNs as a substitute)." (Leadership Member of EarthCube, posted to Science Committee listserv 1/27/16)

No data-management, interpretive, or discovery tool can be implemented in an existing workflow neutrally – there is always an attendant set of buy-ins, negotiations and adjustments in the implementation of that tool. Even theoretically system-neutral middleware is impeded by the change in workflows necessary to accommodate is function: for instance, authentication and log-in procedures are in large part built to limit multiple, automated or anonymized requests (which would be needed for many forms of automated data aggregation), and it is quite difficult to convince system administration of the need to allow potentially risky access for an experimental project. This requires something of a 'policy mandate' – somebody needs to do some work in order to scale. Issues of contribution, reward and participation become paramount to the account of a given infrastructural effort, and organizational issues tend to overshadow all others in describing a narrative of success, partial success, and failure. "There is a tension between the technical ideal of a stable, interoperable infrastructure for data sharing and reuse, and the reality of knowledge as evolving, socially and locally constructed, and often disputed." (Ure, et al. 2009)

> "So, where do we go from there? I think re-centering things on the science would be essential. And having the technologists sit down and listen to that, instead of "systems of systems", would be nothing short of revolutionary. The RCNs (well, some of them) are a successful example of what happens when you put scientists together to geek out some tech solutions to their problems. So that might be where we start: RCNs report on their work, and the technologists sit down and listen and think how they can help. There will be an exam at the end, and if they fail to understand what we need, they fail to get another round of funding. Plain & simple." (Posted 1/26/2016 to the Science Committee listserv)

Reciprocal Subtension: Locating and Negotiating New Infrastructure

"Revolutionary science will come about when scientists can work collectively and diverse points of view are brought to bear on a common problem," and information technology plays a major role in resolving some of the major tensions endemic to working with a distant, remote team. (Olson, Zimmerman, Bos, 2008, pg. 74) However, it also introduces new tensions – the introduction of new technology requires adjustments to workflow, training, and even evaluation. While success, especially among collaborations, may come in a variety of forms, there is significant space for designed technology to play a role in allowing larger and more diverse groups to work together. Technology, in CI and other scientific collaborations, works in part to enable the existence a trading zone that limits the effect of distance, lack of common understanding, and other tensions present in remote work. According to Galison, a trading zone is an area where "radically different activities can be locally, but not globally, coordinated." (Galison, 1996, pg. 119) In a trading zone activity is coordinated at the small scale, strategically and locally. Rather than relying on shared ontology, standardization or other sorts of community agreement, work in a trading zone is strategically coordinated and temporary. Speaking of the relationships between many areas of inquiry that became linked through their use of simulation and modeling in the emergence of computing, "discrete scientific fields were linked by strategies of practice that had previously been separated by object of inquiry." (Galison, 1996, pg. 157) Galison notes that as the study of physics moved from strict reliance on the observation of nature to the application and interpretation of simulations and models there emerged a seeming inter-discipline in the ability to create, manage, and interpret computed simulations. The trading zone does not initially refer to exchanges between relatively stable groups of people (unlike the boundary object, which stands between communities) though those exchanges may occur, it is created of the interactions of individuals. The simulation physicists (who Galison refers to as pariahs in their own community) pioneered a mode of work that became central to the broader narrative of their discipline.

In seeking to build infrastructure, change cascades across the infrastructures currently in use. Infrastructure is a technological system in reciprocal subtension. When I refer to infrastructure in reciprocal subtension, I am not necessarily referring to all infrastructure together – or rather I'm not interested in all of the various subtensions I can possibly trace. Reciprocal subtension is not just the fact of infrastructure being build atop other infrastructures, but a particular relationship of the subtending infrastructures to each other. Take, for example, the telephone system as the subtending infrastructure to the modern internet. While telephony still exists, the material needs of internet infrastructure have fundamentally changed its subtending infrastructure. While the internet was originally built 'atop' telephone infrastructure we now see (almost skeumorphically) telephone infrastructure being built back atop the infrastructure of the internet in the form of VoIP and various applications and programs that support 'live' communication. The basic cable, wires, transformers and poles that characterize telephone infrastructure still exist, as does something that is evidently infrastructure supporting telephone communication, but in subtending a new infrastructure the subtended infrastructure itself undergoes significant change. They coexist, but the design characteristics of each mutually constitute the characteristics of both the subtending and subtended infrastructure to the point that it is difficult to determine where each underlies the other. It is these sorts of infrastructure that can be said to be in reciprocal subtension with each other, where the line between built upon and built blurs. This is not limited to physical infrastructure, systems intended to provide access to legacy data or data collected for other purposes simultaneously provide their own infrastructure of metadata, access and representation while working with and upon the characteristics of the data and systems they are accessing. As built and built upon blur, so too does the concept of working with compared to working on – expanding and providing access to disparately described data is an 'inverting' of the existing infrastructure (Bowker, 1994) to provide a different sort of access, a different set of connections, and often to accommodate other, similarly expansive infrastructures of standards, technology and organizations.

Knowledge infrastructures can also be said to be in reciprocal subtension with each other, where change bears consequence not just for the infrastructure 'in question' but also for each subtending infrastructure. Changing a data standard to encourage interoperability as part of a CI project may also require a substantive set of changes in the standards employed by data centers, in the organizational practices of scientists making use of that data and working with those data centers, and so on. The full implementation of the BCube broker as a centralizing architecture for EarthCube likely would have required significant investment on the part of data centers in accommodating their particular methods of data discovery and access. Existing organizational policy and the path dependence of particular technologies stood 'above' BCube in this case in part because of community investment and the presence of extant work. NERRS's data access was designed to serve its community, and certain expectations of how that data would be used and what would be needed (no more than five years at a time, localized to particular collection sites) represented a significant barrier to the kind of change necessary to implementing the broker. As the NERRS system already subtending significant activity and had limited resources for change, it represented an infrastructure that would be significantly resistant to change outside of policy regimes. GeoWS, on the other hand, in addressing 'long-tail' data, dark data, and other less-infrastructured data sources was more easily able to collaborate with BCube – the development of an accessor to this particular method was achieved with little reported difficulty near the end of the BCube project cycle.

Infrastructures in reciprocal subtension with each other cascade change, or introduce moments of systemic failure when the subtending infrastructure does not or counter-intuitively responds to change. This is something I observed in BCube. While the broker is 'infrastructural' itself, it also is subtended by the standards and processing infrastructures of data centers (potentially even other brokers). However, in BCube the implementation of the broker represents a change in access authorization from some data providers –the broker makes data requests in a way different from an individual scientist doing their own data work. Data centers, then, would need to change their authentication practice in order to allow this different retrieval mode. In many workflows, the implementation of BCube represents a new way of doing the same thing – and that newness can be invisible in many different potential use cases. Like VoIP across telephone wires, it is both ‘underneath’ and ‘above’ the data work it enables. It is identical to the original telephone from a certain perspective but fundamentally different in the standards and technologies enrolled, the requirements for necessary maintenance and repair, and responsive to technology change from different places.

The sheer scale of resources necessary for infrastructure development place results in increasingly-large decision making groups and stakeholder populations. (Galison and Hevly, 1992) Public scrutiny is a natural consequence of such substantial projects, particularly when they are government funded. It can be difficult to parse the difference between a project exploring a variety of standards and policies and the notion that that project represents a funded mandate towards particular forms of interoperability. With aggregation from a variety of sources, metadata is unlikely to capture the full extent of information needed to make the data work. Particularly in the area of troubleshooting, it would be difficult to pinpoint where problems emerge from the data itself (flawed sensor, software bug in the broker) without substantial documentation and understanding of the broker’s function on the part of the users. Some will feel that they do not have enough from the metadata to make an informed decision, and others will have an insufficient sense of control. The next step in development is not in adding features to the broker. Extant features are disbelieved or ill-understood, and future development and demonstration at each stage that is evident to me requires the investment of some community of users – either data managers, EarthCube committees, or the wider scientific communities. Assuming EarthCube results in a success, buy in from any of these increases the likelihood of good placement of the broker – as middleware it does not work on its own.

Conclusion

As discussed in the last chapter, the vision of EarthCube early on was driven by an adjustment on the part of the NSF – they had monitoring regimes similar to the point and general sensors/models present in the TMDLs and water management examples and were adjusting in a similar way. Working towards a relevant goal (novel science) and attempting to build an infrastructure to support it in a variety of ways. The NSF design seeks to better understand and monitor the discursive state of its communities in order to capitalize on relevant, novel science and its techniques. The initial Test Governance committee was designed to self-obsolete and become a support office – it was primarily interested in studying the community of initial researchers engaged in EarthCube and to provide a forum for developing a governance charter for the project as a whole. This is governance building governance and saw a similar introduction of monitoring regimes – the governance committee and eventual voted-in steering committee use open comments, surveys, and digital trace data to understand their communities.

While the NSF sought to understand the community for whom it was building through relatively long-term workshops and the generation of high-level, aspirational goals the community sought to understand itself and its own needs through particular, ground-level descriptions of task, technology and goal. While end-user workshops did produce some use cases it is significant that there is philosophical distinction between interpreting the techniques of governance and self-understanding in building the infrastructure. This emerges in part from differing techniques for self-understanding, and reveals the difference in motivations and sought outcomes from the policy perspective compared to the science perspective. Organization-building and serving particular, defined, and well-understood user needs (in line with a computing technology development tack) fell by the wayside when a demonstrable, centralized architecture needed to be present. The large picture produced in the end user workshops was very general, and didn't translate to many of the domain-neutral technology applications. Projects like BCube were faced simultaneously with issues of parallel development, difficulty enrolling external collaborators and data centers, and the overall difficulty of negotiating the reciprocal subtension of the ecology of existing infrastructures.

CHAPTER 7: Conclusion and a Note on Sustainability

So, to answer my question from the introduction, how does policy-level knowledge influence, account for, and respond to the production of scientific knowledge in terms of regulation, funding, and infrastructure development?

The production and character of science is inevitably responsive to, enabled by, or otherwise engaged with policy at a variety of scales. Scientists, policymakers, and society more broadly work together to co-produce a concept of relevance that guides both financial and time investment in particular areas of study. Regulatory bodies regularly employ scientific principles and enabling technologies in order to implement monitoring and management regimes, and fund science to improve and adapt those principles towards particular management goals. Organizational policy provides both a limit and support for particular scientific activities through structures of reward and recognition, applying mandates to particular standards, formats, or data sources, and serve as a negotiation point for the process of infrastructural strange. Basic science is funded by government agencies in relatively consistent ways and responds to input from scientific communities in both voluntary solicitations of opinion like white papers and potential projects in the form of proposals as well as directed surveying and design-oriented activities. Such scalar and leveraging devices elucidated boundaries, stabilities, stakeholders, provided points of contact and access for policy organizations, worked to identify expertise and experts, and produced a discourse on major issues in that field that identified theoretical and practical points of emphasis. In addition, the knowledge produced for the purposes of policy implementation worked to guide the development and assessment process by defining goals, managing expectations, and producing new sites for ongoing work.

Knowledge production about and the means of pursuing science occurs among both scientific and political actors. Regulatory regimes acknowledge certain models, technologies and techniques as 'good' science through funding, legislation, and the selection of advisors and experts, while scientists actively work to represent to those regimes the state of the art of their science and the its conclusions. Large-scale funding programs begin with an assessment and constitution of a scientific community, and the development of that community is driven simultaneously by concerns endogenous to the science itself as well as those emerging from issues of management, organization and regulation. The early history of the regulation of oyster populations in the Chesapeake Bay shows not only how novel science is accounted for and acknowledged in legislation, regulation and management, but also how that science represents its findings towards certain regulatory goals. EarthCube showed how the process of workshopping and solicitation allowed the self-constitution of a field of study around leveraging devices like umbrella terms and discontext bits. The feedback mechanisms within EarthCube as well as more broadly in the TMDL and Chesapeake Bay regulations bound and defined a group of stakeholders and 'major players,' provided a basis for self-assessment within a particular scientific community, created a space for imaginative activity and established channels of communication through which processes of negotiation and mutual adjustment can take place.

Policy and science, writ large, work together to produce concepts of both scientific and social relevance through a process of mutual adjustment, negotiation, and iteration often extending beyond particular projects, regulatory actions and systems. Tools, like BCube, developed for a particular infrastructural project do not cease to exist just by nature of not being durable parts of that project. The BCube brokering approach, in some form or another, was part of GEON and EuroGEOSS as well as EarthCube, and even if it is not part of those projects in the future the lessons learned from implementation attempts will not simply disappear. In similar vein it is reasonable to expect that experts, technologists, scientists, and others with special interest in cyberinfrastructure will not universally cease to pursue that agenda once a particular project is no longer the site of active work. The rise and fall of particular funded projects is less a sign of the lack of viability of that approach or science and more accurately might represent a process of iterative design and institutional learning that extends beyond individuals, particular calls for funding, and particular institutions. While in the regulatory world these feedback loops and iterative processes are very clearly outlined, this is somewhat less visible in the funding of science – particularly so given the general focus of the NSF on novelty and transformative science.

Various knowledge-producing tools, abstractions, and modes of work enable and support regimes of funding, constitute scientific communities, and provide a means to respond to changing conditions. The design vision of EarthCube changed dramatically during the initial design charrettes in terms of both the overall goals of the project and its scope. What was once intended to produce a working prototype four years out became a decade-long investigation and representation of scientific goals, needs, and capacity intended to give rise to a community-led infrastructure and architecture. In similar vein, ongoing scientific work on regulatory objects like the TMDL enables regulation to account for new knowledge about how pollutants enter water, account for pollutants not previously defined, and creates an actionable abstraction of the state and character of a watershed that does not require close knowledge of its particulars.

EarthCube represents in this work not only a piece of nascent infrastructure in development, but also a site of knowledge production about the geosciences as a whole and the goals and concerns of individuals within funding agencies. BCube responded to the stated concerns of the NSF in its structure and internal goals, and produced knowledge about its collaborating sciences as part of its desing and implementation process. Both EarthCube and BCube provide a window into the necessary negotiation of governmental, organizational, and funding policy from the perspective of instigating infrastructural change, and similarly provide a means of understanding how change is both enabled and limited by the landscape of organizational policy and the pre-existing infrastructural stack. This understanding becomes vital in understand how to build infrastructure in the ecology of infrastructures already at play while also providing a means of interpreting how scientists and technologists conceive of and pursue novelty and innovation in their fields, as well as means by which systems, tools, and communities might make the turn to infrastructure in their own right. GBIF provides a significant example of how, despite a ‘lack of fit’ with established scientific practice, an infrastructural resource might become sustainable through continued investment, ongoing work, and the support of policymakers at the macro level.

Throughout the course of this writing I have worked to characterize infrastructural investment as a significant site of visibility for the interpenetration of political and scientific concerns. Supported by direct funding, with consistent engagement from funded scientists, the presence and interest of government science funders in the development of EarthCube works to constitute a community of researchers, build an understanding of that community, and encourage their active participation in the goals of the project. During my time studying EarthCube the project was very much alive in terms of still receiving resources, support and attention from the NSF as well as other institutions like data centers and NGOs – and will be until at least 2021. However, the story of infrastructure is one where, as funding dries up, so too does activity, participation and engagement – this often leads to a perception where infrastructure projects tend to be perceived to 'fail' on a consistent basis. (Bowker and Ribes, 2008)

Aramis, Latour's study of France's personal transportation network, is a close investigation of just such a 'failed' project. (Latour, 1996) Latour's "real love story" (pg. 149) was experimental writing, an experiment in *scientifiction*, taking the form of a mystery novel with the central question of "Who killed Aramis?" However, throughout the course of the novel he finds that nobody truly Aramis, it was simply that nobody loved it quite enough to see it come to designed fruition. As various points Aramis became more and less real as the ontological landscape changed around it, and as it faded as a real project, a car in which a person might ride, it instead emerged a sit of learning and coordination, where the skills, associations and knowledge produced in the process of design persisted past the designed object and were realized to some extent through it even in the absence of its reality. While nobody in particular killed Aramis, there were moments where its life might well have been saved – had there been support of necessary research, had maintenance and implementation work persisted, there might be an Aramis to point to, ride upon and talk about today. Instead there is a floating concept of what might have been, whose effects linger without their original object of reference. While nobody killed Aramis, ongoing maintenance was in many ways what would have kept it alive:

> "Nothing happens between two elements of Aramis that the engineers aren't obliged to relay through their own bodies. The motor breaks down, the onboard steering shakes and shatters, the automatic features are still heteromats overpopulated with people in blue and white smocks. Chase away the people and I return to an inert state. Bring the people back and I am aroused again, but my life belongs to the engineers who are pushing me, pulling me, repairing me, deciding about me, cursing me, steering me." (Latour, 1996, pg. 123)

However, rather than understanding this fading away as death, in the mode of Latour on Aramis, I would prefer to characterize it as metamorphosis. The rise and fall of projects is part of the cycle of iteration, feedback and adjustment I have been discussing throughout this document, and represents not the death or lack of love for that technology but instead its persistence in different forms. While GEON ended and is no longer being worked upon, components from GEON, which included the software core of the BCube broker, were repurposed for other projects. Scientists in the geosciences did not fail to continue their work without GEON, but it did inform the structure and goals of EarthCube. Without GEON it is unlikely EarthCube would have adopted its middle-out design model, and without EarthCube it is unlikely the next instantiation of some geoscientific or other infrastructure project would take the shape that it does. Memory and learning in the realm of governmental and organizational policy takes place in iteration, in the cycle that begins with the end of a particular set of funded projects and the initiation of work on a new set.

Throughout this writing I have explored the moments where policymaker, regulators, or managers work to constitute scientific communities, where scientists represent themselves and their work, and processes of negotiation, iteration and imagination that go into large collaborative ventures. However, these moments only exist in moments of sustained attention on both sides of the discussion. The production of relevance here becomes paramount to understanding how projects rise and fall, and provides some key insights into how to make infrastructure sustainable. “Sustainability of a cyberinfrastructure over time is a process of ongoing maintenance of infrastructural relationships among people, organizations, and technologies. This maintenance work is taking place against a constant backdrop of change” (Bietz, Ferro, and Lee, 2012) While techniques for the procurement of long-term funding for maintenance in the face of ongoing change is not within the scope of my project or analysis, the notion of relevance as directing both scientific and political attention and effort is a key direction for future discussion and the understanding of how such infrastructure projects might persist.

While the process of feedback, iteration, and adjustment might imply that it is best to let large-scale infrastructure projects fade as interest wanes and new techniques or models for its design emerge, the end result is a lack of ongoing infrastructure. One of the ways in which scientific relevance is produced is through the ongoing availability of funding, but it is likely that there are other pathways to maintaining that relevance and attention, and sustaining the work on that infrastructure. Implied in the research I have presented here is the question of when we can consider an infrastructure complete, or sufficient to its task. Future research work following from this project could focus on understanding how to produce ongoing relevance, how to work through policy channels in support of sustainable research, and how to involve the long-term in infrastructural efforts despite short funding cycles and shifts in administration. Or, in another phrase, how to produce a system that people love enough to keep the project alive.

REFERENCES

Agarwal, D. A., S. R. Sachs, and W. E. Johnston. 1998. "The reality of collaboratories." *Computer Physics Communications* 110, no. 1-3: 134-141.

Almeida, C., & Báscolo, E. 2006. Use of research results in policy decision-making, formulation, and implementation: a review of the literature. Cadernos de Saúde Pública, 22: S7-S19.

Althusser, Louis. 2006. "Ideology and ideological state apparatuses (notes towards an investigation)." *The anthropology of the state: A reader,* 86-111.

Alvesson, Mats, and Dan Karreman. 2000. "Varieties of discourse: On the study of organizations through discourse analysis." *Human relations* 53, no. 9: 1125-1149.

Anderson, Chris. 2008. "The end of theory: The data deluge makes the scientific method obsolete." *Wired Magazine* 16 no. 7

Arizona Geological Survey. 2001. “EarthCube End-User Workshops: Executive Summaries”

Atkins, Daniel. "Electronic Collaboratories and Digital Libraries." *Neuroimage* 4, no. 3 (1996): S55-S58.

---. 2003. *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure*. (January)

---. 2005. "Cyberinfrastructure and the next wave of collaboration." *EDUCAUSE Australia*.

Atkinson, P., & Hammersley, M. 1994. “Ethnography and participant observation.” In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research,* 248-261. Thousand Oaks, CA: Sage.

Aufderheide, Patricia. 1999. *Communications policy and the public interest: The Telecommunications Act of 1996*. Guilford Press.

Avallone, Linnea M., and Brigitte Baeuerle. 2016 "A 20-year history of NSF-supported atmospheric science field campaigns: statistics and demographics." *Bulletin of the American Meteorological Society* 2016.

Baker, Karen S., David Ribes, Florence Millerand, and Geoffrey C Bowker. 2005. "Interoperability strategies for scientific cyberinfrastructure: Research and practice." *Proceedings of the Association for Information Science and Technology,* 42, no. 1.

Barnes, William, Myles Gartland, and Martin Stack. 2004. "Old habits die hard: path dependency and behavioral lock-in." *Journal of Economic Issues* 38, no. 2: 371-377.

Barry, Frank J. 1970 "The Evolution of the Enforcement Provisions of the Federal Water Pollution Control Act: A Study of the Difficulty in Developing Effective Legislation." *Michigan Law Review* 68, no. 6: 1103-1130.

Baru, Chaitan, Sandeep Chandra, Kai Lin, Ashraf Memon, and Choonhan Youn. 2009. "The GEON service-oriented architecture for Earth Science applications." *International Journal of Digital Earth*2, no. S1: 62-78.

Bates, Marcia J. 1998. "Indexing and access for digital libraries and the Internet: Human, database, and domain factors." *Journal of the Association for Information Science and Technology* 49, no. 13: 1185-1205.

Biernacki, Patrick, and Dan Waldorf. 1981. "Snowball sampling: Problems and techniques of chain referral sampling." *Sociological methods & research* 10, no. 2: 141-163.

Bietz, Matthew J., Drew Paine, and Charlotte P. Lee. 2013. "The work of developing cyberinfrastructure middleware projects." In *Proceedings of the 2013 conference on Computer supported cooperative work:* 1527-1538.

Beck, Kent, Mike Beedle, Arie Van Bennekum, Alistair Cockburn, Ward Cunningham, Martin Fowler, James Grenning et al. 2001. "Manifesto for agile software development."

Beniger, James. 2009. *The control revolution: Technological and economic origins of the information society*. Cambridge: Harvard University Press.

Berman, P. The study of macro- and micro-implementation*.* 1978*. Public Policy*, 26 no. 2 (1978): 157–184.

Bernstein, Philip A. 1996. "Middleware: a model for distributed system services." *Communications of the ACM* 39, no. 2: 86-98.

Bishop, Ann Peterson. 1999. "Document structure and digital libraries: how researchers mobilize information in journal articles." *Information Processing & Management* 35, no. 3: 255-279.

Bishop, Ann Peterson, Laura J. Neumann, Susan Leigh Star, Cecelia Merkel, Emily Ignacio, and Robert J. Sandusky. 2000. "Digital libraries: Situating use in changing information infrastructure." *Journal of the American Society for Information Science* 51, no. 4: 394-413.

Borah, Deva K., and Maitreyee Bera. 2004. "Watershed-scale hydrologic and nonpoint-source pollution models: Review of applications." *Transactions of the ASAE* 47, no. 3: 789.

Borah, D. K., G. Yagow, A. Saleh, P. L. Barnes, W. Rosenthal, E. C. Krug, and L. M. Hauck. 2006. "Sediment and nutrient modeling for TMDL development and implementation." *Transactions of the ASABE* 49, no. 4: 967-986.

Borgman, Christine L. 1999. "What are digital libraries? Competing visions." *Inf. Process. Manage.* 35, no. 3: 227-243.

Borah, Porismita. 2011. "Conceptual issues in framing theory: A systematic examination of a decade's literature." *Journal of communication* 61, no. 2: 246-263.

Borsuk, Mark E., Craig A. Stow, and Kenneth H. Reckhow. 2002. "Predicting the frequency of water quality standard violations: A probabilistic approach for TMDL development." *Environmental science & technology* 36, no. 10: 2109-2115.

Bos, Nathan, Ann Cook, Judith Olson, Jude Yew, Jason Yerkie, Erik Dahl, and Gary Olson. 2007. "From shared databases to communities of practice: A taxonomy of collaboratories." *Journal of Computer-Mediated Communication* 12, no. 2: 652-672.

Boullier, Dominique. 2014. "Habitele: mobile technologies reshaping urban life."*URBE* 6, no. 1: 13-16.

Bowker, Geoffrey C. 1994. *Science on the run: Information management and industrial geophysics at Schlumberger, 1920-1940*. Cambridge: MIT Press.

---. 2000. "Biodiversity datadiversity." *Social studies of science* 30, no. 5: 643-683.

---. 2005. Memory practices in the sciences (pp. 269-298). Cambridge, MA: MIT Press

Bowker, Geoffrey C., and Susan Leigh Star. 2000. *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.

Bowker, Geoffrey C., and Bruno Latour. 1987. "A booming discipline short of discipline:(social) studies of science in France." *Social Studies of Science* 17, no. 4: 715-748.

Bowker, Geoffrey C. and David Ribes. 2008 "Organizing for Multidisciplinary Collaboration: The Case of the Geosciences Network." in *Scientific collaboration on the Internet*, edited by Olson, Gary M., Ann Zimmerman, and Nathan Bos: 311-330. Cambridge, MA: The MIT Press.

Bowker, Geoffrey C., Karen Baker, Florence Millerand, and David Ribes. 2010. "Toward information infrastructure studies: Ways of knowing in a networked environment." In *International handbook of internet research*, 97-117. Springer Netherlands.

Bowman, I. (Chairman). 1948. Report of the Commission on Conservation of Natural Resources to the Governor of Maryland, Annapolis, Maryland.

Brand, Stewart. 1995. *How buildings learn: What happens after they're built*. London, UK: Penguin.

Braeckel, Aaron, Bob Barron, Arnaud Dumont, and Bruce Carmichael. 2011. EarthCube Design Approach. White paper released 9/26/2011

Arizona Geological Survey. 2012. EarthCube Governance Framework: A Proposal to the Community - Documentation, Research, and Recommendations.

Brickman, Ronald. 1984. "Science and the politics of toxic chemical regulation: US and European contrasts." *Science, Technology, & Human Values* 9, no. 1: 107-111.

Brooks, William Keith. 1891. *The Oyster: A Popular Summary of a Scientific Study*. Baltimore: Johns Hopkins Press.

Brown, G. and Yule, G. 1983. *Discourse Analysis*, Cambridge: Cambridge University Press

Budds, Jessica. 2004. "Power, nature and neoliberalism: the political ecology of water in Chile." *Singapore Journal of Tropical Geography* 25, no. 3: 322-342.

Busch, Lawrence. 2011. *Standards: Recipes for reality*. Cambridge, MA: MIT Press.

Bush, Vannevar. 1945. "Science: The endless frontier." *Transactions of the Kansas Academy of Science (1903-)* 48, no. 3: 231-264.

Callon, Michel. 1984. "Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay." *The Sociological Review* 32, no. 1_suppl: 196-233.

---. 1986. "The Sociology of an Actor Network: The Case of the Electric Vehicle." In *Mapping the Dynamics of Science and Technology*, edited by A. Rip, M. Callon and J. Law, 19-36. London: MacMillan.

---. 1990. "Techno-economic networks and irreversibility." *The Sociological Review* 38, no. S1: 132-161.

---. 1994. "Is science a public good? Fifth Mullins lecture, Virginia Polytechnic Institute, 23 march 1993." *Science, Technology, & Human Values* 19, no. 4: 395-424.

Latour, Bruno, and Michel Callon. 1992. "Don't throw the baby out with the Bath school." in *Science as practice and culture,* edited by James Pickering, 343-368. Chicago: University of Chicago Press.

Carnegie Classification of Institutions of Higher Education. 2016. *The Carnegie Classification of Institutions of Higher Education.* Retrieved from http://carnegieclassifications.iu.edu/index.php

Carse, A. 2012. “Nature as infrastructure: Making and managing the Panama Canal watershed.” *Social Studies of Science* 42, no. 4: 539-563

Carayannis, Elias G., and David FJ Campbell. 2009. "'Mode 3'and'Quadruple Helix': toward a 21st century fractal innovation ecosystem." *International journal of technology management* 46, no. 3-4: 201-234.

Cash, David W., Jonathan C. Borck, and Anthony G. Patt. 2006. "Countering the loading-dock approach to linking science and decision making: comparative analysis of El Niño/Southern Oscillation (ENSO) forecasting systems." *Science, technology, & human values* 31, no. 4: 465-494.

Clarke, Adele E. 2007. "Grounded theory: Critiques, debates, and situational analysis." *The SAGE handbook of social science methodology*: 423-442.

Clegg, S.R. 1989. *Frameworks of power*. London: Sage.

Cogburn, Derrick L. 2003. "HCI in the so-called developing world: what's in it for everyone." *interactions* 10, no. 2: 80-87.

Collen, B., Ram, M., Zamin, T., & McRae, L. 2008. The tropical biodiversity data gap: addressing disparity in global monitoring. Tropical Conservation Science, 1 no. 2: 75-88.

Collins, C. 1997. The dialogics of “community”: language and identity in a Scottish housing scheme. in P. Hoggett (ed.). *Contested Communities: Experiences, Struggles, Policies*. Policy Press, Bristol.

Collins, Harry M. 1983. "An empirical relativist programme in the sociology of scientific knowledge." in *Science observed: Perspectives on the social study of science* edited by Knorr-Cetina, K. and Mulkay, M., 85-113. London: Sage, 1983.

---. 2008. "Actors’ and analysts’ categories in the social analysis of science." *Clashes of Knowledge*: 101-110.

Collins, Harry M., and Robert Evans. 2002. "The third wave of science studies: Studies of expertise and experience." *Social studies of science* 32, no. 2: 235-296.

Cook, John, Dana Nuccitelli, Sarah A. Green, Mark Richardson, Bärbel Winkler, Rob Painting, Robert Way, Peter Jacobs, and Andrew Skuce. 2013. "Quantifying the consensus on anthropogenic global warming in the scientific literature." *Environmental research letters* 8, no. 2: 024024.

Copeland, Claudia. 2012. "Clean Water Act and Pollutant Total Maximum Daily Loads (TMDLs)." In *CRS Report for Congress Prepared for Members and Committees of Congress*.

Costanza, Robert, Ralph d'Arge, Rudolf De Groot, Stephen Farber, Monica Grasso, Bruce Hannon, Karin Limburg et al. 1997. "The value of the world's ecosystem services and natural capital." *nature* 387, no. 6630: 253-260.

Craver, James M., and Robert S. Gold .2002. "Research collaboratories: their potential for health behavior researchers." *American Journal of Health Behavior* 26, no. 6: 504-509.

Cronin, L. Eugene. 1986. "Chesapeake fisheries and resource stress in the 19th century." *Journal of the Washington Academy of Sciences*: 188-198.

David, P. A. 1990. "The dynamo and the computer." *The American Economic Review* 80, no. 2: 355-361.

D'angelo, Paul. 2002. "News framing as a multiparadigmatic research program: A response to Entman." *Journal of communication* 52, no. 4: 870-888.

Deleuze, Gilles. 1992. "Postscript on the Societies of Control." *October*: 3-7.

De Borgman, Tine, and Jan Luiten Van Zanden. 2008. "Do ut des (I give so that you give back): Collaboratories as a new method for scholarly communication and cooperation for global history." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 41, no. 2: 67-80.

Demeritt, David. 2000. "The new social contract for science: accountability, relevance, and value in US and UK science and research policy." *Antipode* 32, no. 3: 308-329.

DeNardis, Laura.2009. *Protocol politics: The globalization of Internet governance*. Cambridge, MA: MIT Press.

---. 2010. "The emerging field of Internet governance." *Yale Information Society Project Working Paper Series*.

Deutsch, David. 1998. *The fabric of reality*. London, UK: Penguin.

De Vreese, Claes H. 2012. "New avenues for framing research." *American Behavioral Scientist* 56, no. 3: 365-375.

Dilling, Lisa, and Maria Carmen Lemos. 2011. "Creating usable science: Opportunities and constraints for climate knowledge use and their implications for science policy." *Global environmental change* 21, no. 2: 680-689.

Doran, Peter T., and Maggie Kendall Zimmerman. 2009. "Examining the scientific consensus on climate change." *Eos, Transactions American Geophysical Union* 90, no. 3: 22-23.

Dourish, Paul, and Melissa Mazmanian. 2013. "Media as material: Information representations as material foundations for organizational practice." *How Matter Matters: Objects, Artifacts, and Materiality in Organization Studies* 3: 92.

Durkheim, Emile. 1884. "The division of labor in society." *Journal des Economistes*: 211.

Dye, Jane F., Irene M. Schatz, Brian A. Rosenberg, and Susanne T. Coleman. 2000. "Constant comparison method: A kaleidoscope of data." *The qualitative report* 4, no. 1: 1-10.

EarthCube Tools Inventory. 2017. Retrieved 10/20/2017 at https://www.earthcube.org/tools-inventory

EarthCube: Past, Present and Future. 2014. Yolanda Gil, Marjorie Chan, Basil Gomez and Bruce Caron (Eds.) Retrieved 10/20/2017 at https://www.earthcube.org/document/2014/earthcube-past-present-future

Edwards, Paul N. 2010. *A vast machine: Computer models, climate data, and the politics of global warming*. Mit Press, 2010.

Edwards, Paul N., and Gabrielle Hecht. 2010. "History and the technopolitics of identity: the case of apartheid South Africa." *Journal of Southern African Studies* 36, no. 3 (2010): 619-639.

Edwards, Paul N., Steven J. Jackson, Melissa K. Chalmers, Geoffrey C. Bowker, Christine L. Borgman, David Ribes, Matt Burton, and Scout Calvert. 2013. "Knowledge infrastructures: Intellectual frameworks and research challenges." (2013). Available at http://knowledgeinfrastructures.org/

Egyedi, T. 2014. “Standards and sustainable infrastructures: matching compatibility strategies with system flexibility objectives.” *Papers on Open Innovation! A collection of papers on Open Innovation from leading researchers in the field.*, edited by S. Coughlan, 55–80. OpenForum Europe LTD, 2014.

Egyedi, Tineke M., Donna C. Mehos, and Wim G. Vree. 2012. "Introducing inverse infrastructures." *Egyedi, TM, Mehos, DC (eds., 2012)* (2012): 1-16.

Emerson, Robert M., Rachel I. Fretz, and Linda L. Shaw. 2011. *Writing ethnographic fieldnotes*. University of Chicago Press, 2011.

Engestrom, Yrjo. 1990. "When is a tool? Multiple meanings of artifacts in human activity." *Learning, Working and Imagining* (1990): 171-195.

---. 2005. "Knotworking to create collaborative intentionality capital in fluid organizational fields." In *Collaborative capital: Creating intangible value*, pp. 307-336. Emerald Group Publishing Limited, 2005.

Entman, Robert M. 1993. "Framing: Toward clarification of a fractured paradigm." *Journal of communication* 43, no. 4 (1993): 51-58.

Etzkowitz, Henry, and Loet Leydesdorff. 2000. "The dynamics of innovation: from National Systems and “Mode 2” to a Triple Helix of university–industry–government relations." *Research policy* 29, no. 2 (2000): 109-123.

Finholt, Thomas A. 2003. "Collaboratories as a new form of scientific organization." *Economics of innovation and new technology* 12, no. 1 (2003): 5-25.

Finholt, Thomas A., and Gary M. Olson. 1997. "From laboratories to collaboratories: A new organizational form for scientific collaboration." *Psychological Science* 8, no. 1 (1997): 28-36.

Fischer, F. 1993. ‘Policy discourse and the politics of Washington think tanks’, in F. Fischer and J. Forester (eds.) (1993), *The Argumentative Turn in Policy Analysis and Planning*, UCL Press, London.

---. 2003. *Reframing public policy: Discursive politics and deliberative practices*. Oxford University Press, 2003.

Flanagan, M., Howe, D., & Nissenbaum, H. 2008. Embodying values in technology: Theory and practice. *Information technology and moral philosophy*: 322-353.

Foucault, M. 1981. The order of discourse. In R. Young (Ed.), *Untying the text: A post-structural anthology*: 48-78. Boston, MA: Routledge & Kegan Paul.

---. 1991. “Governmentality”. In *The Foucault effect: Studies in governmentality*, edited by Foucault, Michel, Graham Burchell, Colin Gordon, and Peter Miller. Chicago, IL: University of Chicago Press.

---. 2002. *The order of things: An archaeology of the human sciences*. New York, NY: Routledge Press.

Friedberg, Erhard, and Michel Crozier. 1980. *Actors and systems: The politics of collective action*. Chicago, IL: University of Chicago Press.

Friedman, B., Howe, D. C., & Felten, E. 2002. Informed consent in the Mozilla browser: implementing Value-Sensitive Design. In System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on...

Friedman, B., & Nissenbaum, H. 1996. Bias in computer systems. ACM Transactions on Information Systems (TOIS), 14 no. 3: 330-347.

Fujimura, Joan H. 1988. "The molecular biological bandwagon in cancer research: Where social worlds meet." *Social Problems* 35, no. 3: 261-283.

Galison, P. 1992. *Big Science: The Growth of Large-Scale Research*. Stanford, CA: Stanford University Press.

---. 1996. Computer simulations and the trading zone. In *The disunity of science: Boundaries, contexts, and power*, edited by Galison, P., & Stump, D. J., 118-157. Stanford, CA: Stanford University Press.

Gamson, William A., and Kathryn E. Lasch. 1981. "The political culture of social welfare policy." *CRSO Working Paper No. 242*

Georgiadou, Yola, Francis Harvey, and Gianluca Miscione. 2009. "A bigger picture: information systems and spatial data infrastructure research perspectives." In *Global Spatial Data Infrastructure 11 World Conference: SDI Convergence: Building SDI Bridges to Address Global Challenges, Rotterdam, The Netherlands, 15-19 June, 2009*.

GeoWS – Geoscience Web Services. Accessed at https://www.earthcube.org/group/geows-geoscience-web-services

GEO Vision working group. GEO Vision. 2009 Accessed at https://www.nsf.gov/geo/acgeo/geovision/nsf_ac-geo_vision_10_2009.pdf

Gerlitz, Carolin. 2012. "Acting on data. Temporality and self-evaluation in social media." *tbc*: 1-18.

Gerlitz, Carolin, and Celia Lury. 2014. "Social media and self-evaluating assemblages: on numbers, orderings and values." *Distinktion: Scandinavian Journal of Social Theory* 15, no. 2: 174-188.

Gibbons, Michael. 2003. "Mode 2 society and the emergence of context-sensitive science." *Science and public policy* 27, no. 3: 159-163.

Gillespie, Tarleton. 2011. "Can an algorithm be wrong? Twitter Trends, the specter of censorship, and our faith in the algorithms around us." *Culture Digitally* 19. Retrieved at https://socialmediacollective.org/2011/10/19/can-an-algorithm-be-wrong/

Glaser, Barney G. 1965. "The constant comparative method of qualitative analysis." *Social problems* 12, no. 4: 436-445.

Gonçalves, Marcos André, Edward A. Fox, Layne T. Watson, and Neill A. Kipp. 2004. "Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries." *ACM transactions on information systems (TOIS)* 22, no. 2: 270-312.

Gould, Stephen Jay. 1982. "Darwinism and the expansion of evolutionary theory." *Science* 216, no. 4544: 380-387.

Graham, Stephen, and Simon Marvin. 2001. *Splintering urbanism: networked infrastructures, technological mobilities and the urban condition*. New York: Routledge.

Grant, David, Tom W. Keenoy, and Cliff Oswick, eds. 1998. *Discourse and organization*. London, UK: Sage.

Grave, Caswell. 1912. *A Manual of Oyster Culture in Maryland* Baltimore: George W. King.

Gregory, J. 1998, June. Envisioning and Historicizing: Incomplete Utopian Projects. Unpublished Paper, *1SCRAT konference Aarhus, Denmark*.

Habicht, J. P., Victora, C. G., & Vaughan, J. P. 1999. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. International journal of epidemiology, 28 no. 1: 10-18.

Hajer, Maarten. 2003. "Policy without polity? Policy analysis and the institutional void." *Policy sciences* 36, no. 2: 175-195.

---. 2009. *Authoritative governance: Policy making in the age of mediatization*. Oxford: Oxford University Press.

Hannemyr, Gisle. 2003. "The Internet as hyperbole: A critical examination of adoption rates." *The Information Society* 19, no. 2: 111-121.

Hedstrom, Margaret. 1997. "Digital preservation: a time bomb for digital libraries." *Computers and the Humanities* 31, no. 3: 189-202.

Hepsø, Vidar, Eric Monteiro, and Knut H. Rolland. 2009. "Ecologies of e-Infrastructures." *Journal of the Association for Information Systems* 10, no. 5: 2.

Hetherington, Kregg, and Jeremy M. Campbell. 2014. "Nature, infrastructure, and the state: Rethinking development in Latin America." *The Journal of Latin American and Caribbean Anthropology* 19, no. 2: 191-194.

Hodge, Gail. 2000. *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Digital Library Federation, Council on Library and Information Resources. Washington, DC.

Hong, Weiyin, James YL Thong, and Kar-Yan Tam Wai-Man Wong. 2002. "Determinants of user acceptance of digital libraries: an empirical examination of individual differences and system characteristics." *Journal of Management Information Systems* 18, no. 3: 97-124.

Hook, Derek. 2001. "Discourse, Knowledge, Materiality, History Foucault and Discourse Analysis." Theory & Psychology 11, no. 4.

Houck, Oliver A. 2002. *The clean water act TMDL program: Law, policy, and implementation*. Washington, DC: Environmental Law Institute.

Hughes, Thomas Parke. 1993. *Networks of power: electrification in Western society, 1880-1930*. Baltimore, MD: JHU Press.

Van Hulst, Merlijn, and Dvora Yanow. 2016. "From policy “frames” to “framing” theorizing a more dynamic, political approach." *The American Review of Public Administration* 46, no. 1: 92-112.

Introna, L. D., & Nissenbaum, H. 2000. Shaping the Web: Why the politics of search engines matters. The information society, 16, no. 3: 169-185.

Jackson, Steven J., and Ayse Buyuktur. 2014. "Who killed WATERS? Mess, method, and forensic explanation in the making and unmaking of large-scale science networks." *Science, Technology, & Human Values* 39, no. 2: 285-308.

Jackson, Steven J., Paul N. Edwards, Geoffrey C. Bowker, and Cory P. Knobel. 2007. "Understanding infrastructure: History, heuristics and cyberinfrastructure policy." *First Monday* 12, no. 6.

Jackson, Steven J., Tarleton Gillespie, and Sandy Payette. 2014. "The policy knot: Re-integrating policy, practice and design in CSCW studies of social computing." In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*: 588-602.

Jansson, Janet K., and Neslihan Taş. 2014. "The microbial ecology of permafrost." *Nature Reviews Microbiology* 12, no. 6: 414-425.

Jasanoff, Sheila, ed. 2004. *States of knowledge: the co-production of science and the social order*. London, UK: Routledge.

---. 2009. *The fifth branch: Science advisers as policymakers*. Cambridge, MA: Harvard University Press.

Jensen, Casper Bruun. 2008. "Power, technology and social studies of health care: An infrastructural inversion." *Health Care Analysis* 16, no. 4: 355-374.

Joerges, Bernward. 1999. "Do politics have artefacts?." *Social studies of science*29, no. 3: 411-431.

John, Richard R. 2009. *Spreading the news: The American postal system from Franklin to Morse*. Cambridge, MA: Harvard University Press.

Jorgensen, Danny L. 1989. *Participant observation*. New York: John Wiley & Sons, Inc.

Kaghan, William N., and Geoffrey C. Bowker. 2001. "Out of machine age?: complexity, sociotechnical systems and actor network theory." *Journal of Engineering and Technology Management* 18, no. 3: 253-269.

Kaltenbrunner, Wolfgang. 2015. "Infrastructural inversion as a generative resource in digital scholarship." *Science as Culture* 24, no. 1: 1-23.

Kang, M. S., S. W. Park, J. J. Lee, and K. H. Yoo. 2006. "Applying SWAT for TMDL programs to a small watershed containing rice paddy fields." *Agricultural Water Management* 79, no. 1: 72-92.

Kauffman, Stuart A. 1993. *The origins of order: Self-organization and selection in evolution*. Oxford: Oxford University Press.

Keenoy, Tom, and Cliff Oswick. 2004. "Organizing textscapes." *Organization Studies* 25, no. 1: 135-142.

Kinman, Ken. 2002, August 21. Library on Fire (triage). [Electronic mailing list message] retrieved from http://mailman.nhm.ku.edu/pipermail/taxacom/2002-August/041477.html

Zimmerman, Christine. 1998. "WK Brooks and the oyster question: science, politics, and resource management in Maryland, 1880–1930." *Journal of the History of Biology* 31, no. 3: 383-424.

Kennedy, Victor S., and Kent Mountford. 2001. "Human influences on aquatic resources in the Chesapeake Bay watershed." *Discovering the Chesapeake: the history of an ecosystem*: 191-219.

Kennedy, Victor S., and Linda L. Breisch. 1983. "Sixteen decades of political management of the oyster fishery in Maryland's Chesapeake Bay." *Journal of Environmental Management* 16, no. 2: 153-171.

Khalsa and Nativi. 2012. Presentation Materials for BCube Kickoff Meeting, November 2012.

Kintigh, Keith W. 1984. "Measuring archaeological diversity by comparison with simulated assemblages." *American Antiquity* 49, no. 1: 44-54.

Klose, Alexander, and Charles Marcrum. 2015. *The Container Principle: How a Box Changes the Way We Think*. Cambridge, MA: MIT Press.

Knobel, Cory. 2010. Ontic Occlusion and Exposure in Sociotechnical Systems. Thesis prepared for the University of Michigan.

Knobel, C., & Bowker, G. C. 2011. Values in design. Communications of the ACM, 54, no. 7: 26-28.

Knorr-Cetina, Karin. 1999. *Epistemic communities*. Cambridge, MA: Harvard Educational Press.

Knorr-Cetina, K. and Mulkay, M. 1983. *Science observed: perspectives on the social study of science.* London, UK: Sage.

Korn, Matthias, and Amy Voida. 1996. "Creating friction: infrastructuring civic engagement in everyday life." In *Proceedings of The Fifth Decennial Aarhus Conference on Critical Alternatives*, pp. 145-156. Aarhus University Press, 2015.

Kouzes, Richard T., James D. Myers, and William Wulf. 1996. "Collaboratories: Doing science on the Internet." *Computer* 29, no. 8: 40-46.

Kramer, Adam DI, Jamie E. Guillory, and Jeffrey T. Hancock. 2014. "Experimental evidence of massive-scale emotional contagion through social networks."*Proceedings of the National Academy of Sciences* 111, no. 24: 8788-8790.

Kranzberg, Melvin. 1986. "Technology and History: Kranzberg's Laws."*Technology and culture*: 544-560.

Kuhn, Thomas S. 2012. *The structure of scientific revolutions*. Chicago, Il: University of Chicago press.

Lampland, Martha, and Susan Leigh Star. 2009. *Standards and their stories: how quantifying, classifying, and formalizing practices shape everyday life*. Ithaca, NY: Cornell University Press.

Lakatos, Imre. 1970. Falsification and the methodology of scientific research programmes. In *Criticism and the Growth of* Knowledge, Imre Lakatos and Alan Musgrave, Eds. Cambridge: Cambridge University Press.

Lakoff, George. 1993. "The contemporary theory of metaphor." *Metaphor and thought* 2: 202-251.

Latour, B. 1983. "Give me a laboratory and I will raise the world." In *Science Observed: Perspectives on the Social Study of Science*, edited by Knorr-Cetina, K. and Mulkay, M., London, UK: Sage Publications.

---. 1987. *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard university press.

---. 1992. "Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts." *Shaping technology/building society: studies in sociotechnical change*, in W. E. Bijker and J. Law, 225–258: Cambridge, MA.: MIT Press.

---. 1993. *The pasteurization of France*. Cambridge, MA: Harvard University Press.

---. 1996. *Aramis, or, The love of technology*. Cambridge, MA: Harvard University Press.

---. 2002. "Gabriel Tarde and the End of the Social." in *The social in question: New bearings in history and the social sciences* edited by Joyce, Patrick. New York: Routledge: 117-132

---. 2004. *Politics of nature*. Cambridge, MA: Harvard University Press.

---. 2007. *Reassembling the Social: An Introduction to Actor-Network-Theory*. New York: Oxford University Press.

Latour, Bruno, and Steve Woolgar. 1979. *Laboratory life: The construction of scientific facts*. London, UK: Sage.

Lave, Jean, and Etienne Wenger. 1991. *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Lawrence, Steve, C. Lee Giles, and Kurt Bollacker. 1999. "Digital libraries and autonomous citation indexing." *Computer* 32, no. 6: 67-71.

Lazer, David, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, Nicholas Christakis et al. 2009. "Life in the network: the coming age of computational social science." *Science (New York, NY)* 323, no. 5915: 721.

Lee, Charlotte P., Paul Dourish, and Gloria Mark. 2006. "The human infrastructure of cyberinfrastructure." In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*: 483-492.

Michael Lesk. 1997. *Practical digital libraries: books, bytes, and bucks*. Burlington, MA: Morgan Kaufmann.

Lewandowsky, Stephan, Gilles E. Gignac, and Samuel Vaughan. 2013. "The pivotal role of perceived scientific consensus in acceptance of science." *Nature Climate Change* 3, no. 4: 399-404.

Leydesdorff, Loet. 2013. "Triple Helix of university-industry-government relations." In *Encyclopedia of Creativity, Invention, Innovation and Entrepreneurship*, pp. 1844-1851. New York: Springer New York.

Lima, M. 2011. *Visual Complexity: Displaying Complex Networks and Data Sets.* New York, NY: Princeton Architectural Press.

Lindell, Michael K., and Christina J. Brandt. 2000. "Climate quality and climate consensus as mediators of the relationship between organizational antecedents and outcomes." *Journal of applied psychology* 85, no. 3: 331.

Linker, Lewis C., Richard A. Batiuk, Gary W. Shenk, and Carl F. Cerco. 2013. "Development of the Chesapeake Bay watershed total maximum daily load allocation." *JAWRA Journal of the American Water Resources Association* 49, no. 5: 986-1006.

Lofland, John, and Lyn H. Lofland. 2006. *Analyzing social settings*. Belmont, CA: Wadsworth Publishing Company.

Lutters, Wayne G., Mark S. Ackerman, James Boster, and David W. McDonald. 2000. "Mapping knowledge networks in organizations: creating a knowledge mapping instrument." *AMCIS 2000 Proceedings*

Lynch, M. 1985. *Art and Artefact in Laboratory Science*. London: Routledge.

Lynch, Clifford, and Hector Garcia-Molina. 1996. "Interoperability, Scaling, and the Digital Libraries Research Agenda." *Microcomputers for Information Management* 13, no. 2: 85-132.

Mackenzie, Adrian. 2005. "Untangling the unwired: Wi-Fi and the cultural inversion of infrastructure." *Space and Culture* 8, no. 3: 269-285.

Maes, Joachim, Benis Egoh, Louise Willemen, Camino Liquete, Petteri Vihervaara, Jan Philipp Schägner, Bruna Grizzetti et al. 2012. "Mapping ecosystem services for policy support and decision making in the European Union." *Ecosystem Services* 1, no. 1: 31-39.

Malone, Linda A. 2002. "The Myths and Truths That Ended the 2000 TMDL Program." *Pace Envtl. L. Rev.* 20: 63.

Marcus, George E. 2009. "Multi-sited ethnography: Notes and queries." In Mark-Anthony Falzone, ed. *Multi-sited ethnography: Theory, praxis, and locality in contemporary research*. New York: Routledge. 181-196.

Martin, Robert C. 2002. *Agile software development: principles, patterns, and practices*. Upper Saddle River, NJ: Prentice Hall.

Marx, Karl, and Frederick Engels. 1970. *The German Ideology*. ed. CJ Arthur. New York: International.

Mayernik, Matthew S., Jillian C. Wallis, and Christine L. Borgman. 2013. "Unearthing the infrastructure: humans and sensors in field-based scientific research." *Computer Supported Cooperative Work (CSCW)* 22, no. 1: 65-101.

McCarthy, John David. 1987. “Pro-Life and Pro-Choice Mobilization; Infrastructure Deficits and New Technologies.” In Zald, Mayer N., and John David McCarthy, eds. *Social movements in an organizational society: Collected essays*. Piscataway, NJ: Transaction Publishers.

Merton, Robert K. 1968. "The Matthew effect in science." *Science* 159, no. 3810: 56-63.

Mittermeier RA, Gil PR, Hoffman M, Pilgrim J, Brooks T, et al. 2005. *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions*. Mexico City: Agrupacion Sierra Madre.

Monge, Peter R., and Noshir S. Contractor. 2001. "Emergence of communication networks." *The new handbook of organizational communication: Advances in theory, research, and methods*: 440-502.

Morita, Atsuro. 2017. "Multispecies Infrastructure: Infrastructural Inversion and Involutionary Entanglements in the Chao Phraya Delta, Thailand." *Ethnos* 82, no. 4: 738-757.

Mulgan, Geoffrey J. 1991. *Communication and control: Networks and the new economies of communication*. New York: Guilford Publications, Inc.

Mumford, Lewis, and George Copeland. 1961. *The city in history: Its origins, its transformations, and its prospects*. New York: Harcourt, Brace & World.

Nardi, Bonnie A. 1996. "Studying context: A comparison of activity theory, situated action models, and distributed cognition." *Context and consciousness: Activity theory and human-computer interaction*. Cambridge, MA: MIT Press.

National Research Council. 2001. Assessing the TMDL Approach to Water Quality Management. available at http://books.nap.edu/html/tmdl/

Nativi, S., Khalsa, S., Domenico, B., Craglia, M., Pearlman, J., Mazzetti, P., Rew, R. 2011. The Brokering Approach for Earth Science Cyberinfrastructure. EarthCube White Paper submitted October 10, 2011.

Nativi, S., M. Santoro, L. Vaccari, M. Craglia, M. 2011. Grasso. Specification of EUROGEOSS Initial Operating Capacity, EuroGEOSS Deliverable D.2.2.2.b, March 2011.

Nissenbaum, H. 1998. Values in the design of computer systems. *Computers and Society*, 28, no. 1: 38-39.

Norton, Bryan G. 2005. *Sustainability: A philosophy of adaptive ecosystem management*. University of Chicago Press.

Olson, Gary M., and Judith S. Olson. 2000. "Distance matters." *Human-computer interaction* 15, no. 2: 139-178.

Oreskes, Naomi. 2004. "The scientific consensus on climate change." *Science* 306, no. 5702: 1686-1686.

Østerlund, Carsten, Steve Sawyer, David Ribes, Kalpana Shankar, and Stuart Geiger. 2014. "What to Do with all those Traces People Leave behind: Computing, Culture, and (Bits of) Context?." *iConference 2014 Proceedings*.

Pelizza, Annalisa. 2016. "Developing the Vectorial Glance: Infrastructural Inversion for the New Agenda on Government Information Systems." *Science, Technology, & Human Values* 41, no. 2: 298-321.

Pinch, T.J. and W. E. Bijker. 1987. The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. In The Social Construction of Technological Systems, edited by W. E. Bijker, T.J. Pinch, and T. P. Hughes, 17-50. Cambridge: Cambridge University Press.

Poole, Austin Lane. 1993. *From Domesday Book to Magna Carta, 1087-1216*. Vol. 3. Oxford, UK: Oxford University Press.

Porter, Theodore M. 1996. *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Potter, Jonathan. 1997. "Discourse analysis as a way of analysing naturally occurring talk." *Qualitative research: Theory, method and practice* 2: 200-222.

Potter, Jonathan, and Margaret Wetherell. 1987. *Discourse and social psychology: Beyond attitudes and behaviour*. London, UK: Sage.

Propp, Vladimir. 2010. *Morphology of the Folktale*. Vol. 9. Austin, TX: University of Texas Press.

Pyle, Richard. 2010, January 27. [Taxacom] Was Reproducibility of phylogenetic analysis. [Electronic mailing list message] retrieved from http://mailman.nhm.ku.edu/pipermail/taxacom/2010-January/067852.html

Reckhow, K. H. 2003. "On the need for uncertainty assessment in TMDL modeling and implementation." *Journal of Water Resources Planning and Management* 129, no. 4: 245-246.

Reese, Stephen D. 2007. "The framing project: A bridging model for media research revisited." *Journal of communication* 57, no. 1: 148-154.

Ribes, David. 2014. "Ethnography of scaling, or, how to a fit a national research infrastructure in the room." In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*:158-170.

Ribes, David, and Karen Baker. 2007. "Modes of social science engagement in community infrastructure design." In *Communities and Technologies 2007*. London, UK: Springer London. 107-130

Ribes, David, and Thomas A. Finholt. 2009. "The long now of technology infrastructure: articulating tensions in development." *Journal of the Association for Information Systems* 10, no. 5: 375.

Rip, Arie, and Jan-Peter Voß. 2013. "Umbrella terms as a conduit in the governance of emerging science and technology." *Science, technology & innovation studies* 9, no. 2: 39-59.

Rowland, Kate Mason, G. Mason, Alexander Henderson, Daniel of St. Thomas Jenifer, T. Stone, and Samul Chase. 1888. "The Mount Vernon Convention." *Pennsylvania Magazine of History and Biography*: 410-425.

Rose, Nikolas. 1990. *Governing the soul: the shaping of the private self*. London: Free Association Books.

Rubin, Herbert J., and Irene S. Rubin. 2011. *Qualitative interviewing: The art of hearing data*. London, UK: Sage.

Rubin, Richard E. 1998. *Foundations of Library and Information Science*. New York: Neal-Schuman Publishers.

Ruppert, Evelyn, Penny Harvey, Cellia Lury, Adrian Mackenzie, Ruth McNally, Stephanie Alice Baker, Yannis Kallianos, and Camilla Lewis. 2015. "Socialising Big Data: From concept to practice." *CRESC Working Paper Series* 138.

Sahay, Sundeep, Eric Monteiro, and Margunn Aanestad. 2009. "Configurable politics and asymmetric integration: Health e-infrastructures in India." *Journal of the Association for Information Systems* 10, no. 5: 4.

Said, E. 1983. *The world, the text and the critic*. Cambridge, MA: Harvard University Press.

Sandvig, Christian, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. "Auditing algorithms: Research methods for detecting discrimination on internet platforms." Paper presented to *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*, a preconference at the 64th Annual Meeting of the International Communication Association. May 22, 2014; Seattle, WA, USA.

Sarewitz, Daniel, and Roger A. Pielke. 2007. "The neglected heart of science policy: reconciling supply of and demand for science." *environmental science & policy* 10, no. 1: 5-16.

Sartre, Jean-Paul. 2001. *Colonialism and neocolonialism*. New York: Routledge.

Sawyer, Steve. 2001. "A market-based perspective on information systems development." *Communications of the ACM* 44, no. 11: 97-102.

Schatz, Bruce R. 1997. "Information retrieval in digital libraries: Bringing search to the net." *Science* 275, no. 5298: 327-334.

Schatz, Bruce, and Hsinchun Chen. 1996. "Building large-scale digital libraries." Computer 29, no. 5: 22-26.

Schleyer, Titus KL, Stephanie D. Teasley, and Rishi Bhatnagar. 2005. "Comparative case study of two biomedical research collaboratories." *Journal of medical Internet research* 7, no. 5.

Schüll, Natasha Dow. 2012. *Addiction by design: Machine gambling in Las Vegas*. Princeton, NJ: Princeton University Press.

Seaver, Nick. 2012.Algorithmic recommendations and synaptic functions. *Limn* 1, no. 2.

Sengers, P., Boehner, K., David, S., & Kaye, J. J. 2005. Reflective design. In *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility* ACM: 49-58.

Serres, Michel. 1982. *The parasite* LR Schehr, Trans. Baltimore: The Johns Hopkins University Press.

---. 1995. *The natural contract*. Ann Arbor, MI: University of Michigan Press.

Shannon, Claude E., and Warren Weaver. 2015. *The mathematical theory of communication*. Champaign, IL: University of Illinois press.

Shenk, Gary W., and Lewis C. Linker. 2013. "Development and application of the 2010 Chesapeake Bay watershed total maximum daily load model." *JAWRA Journal of the American Water Resources Association* 49, no. 5: 1042-1056.

Shirmohammadi, A., I. Chaubey, R. D. Harmel, D. D. Bosch, R. Muñoz-Carpena, C. Dharmasri, A. Sexton et al. 2006. "Uncertainty in TMDL models." *Transactions of the ASABE* 49, no. 4: 1033-1049.

Simon, Herbert A. 1972. "Theories of bounded rationality." *Decision and organization* 1, no. 1: 161-176.

Snow, David A., and Robert D. Benford. 1988. "Ideology, frame resonance, and participant mobilization." *International social movement research* 1, no. 1: 197-217.

Soberón, J., & Peterson, T. 2004. Biodiversity informatics: managing and applying primary biodiversity data. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359, no. 1444: 689-698.

Sohn-Rethel, Alfred. 1987. *The economy and class structure of German Fascism*. London: Free Association.

Sokal, Alan D. 1996. "A physicist experiments with cultural studies." *Lingua franca*6, no. 4: 62-64.

de Solla Price, Derek John. 1986. *Little science, big science... and beyond*. New York: Columbia University Press.

Sonnenwald, Diane H., and Bin Li. 2003. "Scientific collaboratories in higher education: exploring learning style preferences and perceptions of technology." *British Journal of Educational Technology* 34, no. 4: 419-431.

Star, Susan Leigh, and Karen Ruhleder. 1996. "Steps toward an ecology of infrastructure: Design and access for large information spaces." *Information systems research* 7, no. 1: 111-134.

Star, Susan Leigh, and Geoffrey C. Bowker. 2006. "How to infrastructure." *Handbook of new media: Social shaping and social consequences of ICTs*: 230-245.

Starosielski, Nicole. 2015. *The Undersea Network*. Durham, NC: Duke University Press.

Stenner, Paul. 2007. "Non-foundational criticality? On the need for a process ontology of the psychosocial." *Outlines: Critical practice studies* 9, no. 2: 44-55.

Strathern, Marilyn. 2004. *Commons and borderlands: Working papers on interdisciplinarity, accountability and the flow of knowledge*. Herefordshire, UK: Sean Kingston Publishing.

Strauss, Anselm, and Juliet Corbin. 1994. "Grounded theory methodology." *Handbook of qualitative research* 17: 273-85.

Strum, Shirley S., and Bruno Latour. 1987. "Redefining the social link: from baboons to humans." *Social Science Information* 26, no. 4: 783-802.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. 2009. eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142, no. 10: 2282-2292.

Sutherland, William J., Susan Armstrong-Brown, Paul R. Armsworth, Brereton Tom, Jonathan Brickland, Colin D. Campbell, Daniel E. Chamberlain et al. 2006. "The identification of 100 ecological questions of high policy relevance in the UK." *Journal of applied ecology* 43, no. 4: 617-627.

Suarez-Villa, Luis. 1997. "Innovative capacity, infrastructure and regional inversion: is there a long-term dynamic?." *Innovative Behaviour in Space and Time*: 291-305.

Suchman, Lucy A. 1987. *Plans and situated actions: The problem of human-machine communication*. Cambridge: Cambridge University Press.

Takacs, David. 1996. *The idea of biodiversity: philosophies of paradise.* Baltimore, MD: Johns Hopkins University Press.

The Chronos Project. 2007. Retrieved 10/14/2017 at http://www.geongrid.org/07/pp_chronos.php

Timmermans, Stefan, and Steven Epstein. 2010. “A World of Standards but Not a Standard World: Toward a Sociology of Standards and Standardization.” Annual Review of Sociology 36: 69–89.

Trist, E. L., and K. W. Bamforth. 1951. "Some social and psychological consequences of the Longwall method." *Human relations* 4, no. 3: 3-38.

Turnhout, Esther, Matthijs Hisschemöller, and Herman Eijsackers. 2007. "Ecological indicators: between the two fires of science and policy." *Ecological indicators* 7, no. 2: 215-228.

Twidale, Michael B., and Ingbert Floyd. 2008. "Infrastructures from the bottom-up and the top-down: can they meet in the middle?." In *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008*, Indiana University: 238-241.

Ure, Jenny, Rob Procter, Yu-wei Lin, Mark Hartswood, Stuart Anderson, Sharon Lloyd, Joanna Wardlaw, Horacio Gonzalez-Velez, and Kate Ho. 2009. "The development of data infrastructures for ehealth: a socio-technical perspective." *Journal of the Association for Information Systems* 10, no. 5: 415.

Van den Hove, Sybille. 2007. "A rationale for science–policy interfaces." *Futures* 39, no. 7: 807-826.

Van der Vleuten, Erik. 2004. "Infrastructures and societal change. A view from the large technical systems field." *Technology Analysis & Strategic Management* 16, no. 3: 395-414.

Van Dijk, T. 1993. Principles of critical discourse analysis. *Discourse & Society* 4: 249-283

Van House, Nancy A., Ann P. Bishop, and Barbara P. Buttenfield. 2003. "Introduction: Digital libraries as sociotechnical systems." *Digital library use: Social practice in design and evaluation*: 1-21.

Van Looy, Jan, and Jan Baetens, eds. 2003. *Close reading new media: Analyzing electronic literature*. Vol. 16. Leuven, Belgium: Leuven University Press.

Veyne, Paul. 2013. *Foucault: His thought, his character*. Hoboken, NJ: John Wiley & Sons.

Voinov, Alexey, and Erica J. Brown Gaddis. 2008. "Lessons for successful participatory watershed modeling: a perspective from modeling practitioners." *Ecological modelling* 216, no. 2: 197-207.

Wackers, Ger. 2009. "Offshore Vulnerability: The Limits of Design and the Ubiquity of the Recursive Process." *RISKY WORK ENVIRONMENTS, supra note* 21: 81-98.

Wainger, Lisa A. 2012. "Opportunities for reducing total maximum daily load (TMDL) compliance costs: Lessons from the Chesapeake Bay." *Environmental science & technology* 46, no. 17: 9256-9265.

Wainger, Lisa A., George Van Houtven, Ross Loomis, Jay Messer, Robert Beach, and Marion Deerhake. 2013. "Tradeoffs among ecosystem services, performance certainty, and cost-efficiency in implementation of the Chesapeake Bay total maximum daily load." *Agricultural and Resource Economics Review* 42, no. 1: 196-224.

Wiener, Norbert. 1961. *Cybernetics or Control and Communication in the Animal and the Machine*. Vol. 25. Cambridge, MA: MIT press.

---. 1988. *The human use of human beings: Cybernetics and society*. No. 320. New York: Perseus Books Group.

Witten, Ian, David Bainbridge, Gordon Paynter, and Stefan Boddie. 2002. "Importing documents and metadata into digital libraries: Requirements analysis and an extensible architecture." *Research and Advanced Technology for Digital Libraries*: 219-229.

Wellman, Barry. 1988. Structural analysis: From method and metaphor to theory and substance. In B. Wellman & S. D. Berkowitz (Eds.), *Social structures: A network approach.* New York: Cambridge University Press: 19-61.

---. 1996. "For a social network analysis of computer networks: a sociological perspective on collaborative work and virtual community." In *Proceedings of the 1996 ACM SIGCPR/SIGMIS conference on Computer personnel research*: 1-11.

Wennersten, John. 2007. *The oyster wars of Chesapeake Bay*. Centerville, MD: Tidewater Publishers.

Williams, Michael R., Gopal Bhatt, Solange Filoso, and Guido Yactayo. 2017. "Stream Restoration Performance and Its Contribution to the Chesapeake Bay TMDL: Challenges Posed by Climate Change in Urban Areas." *Estuaries and Coasts*: 1-20.

Wimsatt, William C. 2001. "Generative entrenchment and the developmental systems approach to evolutionary processes." *Cycles of contingency: Developmental systems and evolution*: 219-237.

Winner, Langdon. 1980. "Do artifacts have politics?." *Daedalus*: 121-136.

Winslow, Francis. 1882. Report on the Oyster Beds of the James River, Virginia, and of Tangier and Pocomoke Sounds, Maryland and Virginia: U.S. Coast and Geodetic Survey Report for 188. Washington, D.C.

Woolgar, Steve. 1986. "On the alleged distinction between discourse and praxis."*Social Studies of Science*: 309-317.

---. 1993. "What's at Stake in the Sociology of Technology? A Reply to Pinch and to Winner." *Science, Technology, and Human Values*: 523-529.

Wulf, W. A. 1989. "Towards a national collaboratory." In *Report of Invitational Workshop at Rockefeller University*.

Wynne, Brian, and Alan Irwin, eds. 1996. *Misunderstanding science?: the public reconstruction of science and technology*. Cambridge, MA: Cambridge University Press, 1996.

Xentity Corporation. 2016. “Technical Response: EarthCube Architecture & Implementation Plan Development” Proposal submitted August 22, 2016.

Yates, C. 1913. Summary of survey of oyster bars of Maryland 1906-1912. Washington, D.C.: U.S. Coast and Geodetic Survey.

Yates, JoAnne. 1993. *Control through communication: The rise of system in American management*. Vol. 6. Baltimore, MD: JHU Press.

Yesson, Chris, Peter W. Brewer, Tim Sutton, Neil Caithness, Jaspreet S. Pahwa, Mikhaila Burgess, W. Alec Gray et al. 2007. "How global is the global biodiversity information facility?." *PLoS One* 2, no. 11: e1124.

Ziek, Paul. 2012. CSR infrastructure for communication and the Nike controversy. *Journal of Management and Sustainability* 3, no. 1: 63.

Zimmerer, Karl S., and Thomas J. Bassett. 2003. Introduction. in Zimmerer, Karl S., and Thomas J. Bassett, eds. *Political ecology: an integrative approach to geography and environment-development studies*. New York: Guilford Press.