A FRAMEWORK FOR ANALYZING THE SUSTAINABILITY OF PEER PRODUCED SCIENCE COMMONS

BY

NICHOLAS M.WEBER

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Graduate School of Library and Information Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Dr. Carole L. Palmer, Chair
Dr. Michael Twidale
Dr. Allen Renear
Dr. Peter Fox

TABLE OF CONTENTS

CHAPTER 1

1.1 The Commons

There is widespread agreement that scientific knowledge is a public good that can be shared broadly without decreasing its value. As a result, the ability to easily draw upon an open stock of scientific knowledge produces externalities that positively benefit society.

Like scientific knowledge, the products of scientific research and development (data, software, journal publications, instruments, etc.) create many positive spillover effects for society. But, unlike knowledge, immediate access and broad sharing of these resources affect their value. Competition for use, social dilemmas around rights of access, and costs related to barriers of entry all impact the way that the products of scientific research are provisioned. Typically, the products of scientific research are subject to one of two traditional forms of economic organization:

1. The market, where intellectual property rights and patents create exclusivity for private ownership, or

2. The state, which regulate and intervenes in markets by subsiding innovation, and consequently creating open-access resource systems.

In science and technology policy the two approaches are often referred to as a tension between the commercialization of science on the one hand, and the so called "norms"of science on the other (Merton, 1963; Eisenberg 1989; Rai 1999; Reichman Uhlir 2003; Mirowski, 2012; Madison, 2014).

Both the the market and the state model are costly in a contemporary research and development environment that requires broad collaboration: Markets tend to create wealth by providing limited protection (e.g. patents) to innovators, but these barriers to access often impede the types of transformative breakthroughs that are necessary for combating global scale issues, such as the sharing of data around an Ebola outbreak in Western Africa (Lofgren et al., 2014); Governments tend to create equitable access and opportunity for innovation, but rarely do scientific funding agnecies have the capacity to sustain investments in research projects longer than a five year grant cycle; leaving innovation unrealized (Mirowksi, 2011).

Throughout the economic literature of the 20th century, the market and the state appear as two conceptual poles for the production and provisioning of scientific research, but they are not the only options available (Ostrom, 2009). A third approach is the commons - a term informally used to refer to any area of economic activity where there is "freedom-to-operate under symmetric constraints, available to an open, or undefined, class of users" (Benkler, 2014). In practice, the commons resemble institutional arrangements predicated on self-organization and community governance at local levels; and broad cooperation, shared norms, and collective rule-making procedures at larger-scales (Ostrom, 1990). The commons offer an institutional alternative to markets and governments, but are not fully separate from government subsidy, nor without answer to contemporary capital-based marketplace models. The commons offer an institutional arrangement which can mix elements of free market ideology alongside hierarchical governance structures. In short, the commons are an institutional hybrid, but not a compromise (Frischmann, 2005).

The commons have proven to be an institutional arrangement especially effective at sustaining socioecological systems that couple people, technologies, and environmental issues of a complex nature (Dietz et al., 2010). Examples of these success can be found at a number of different scales, ranging from effective international treaties negotiated to curb harmful atmospheric emission, such as the Montreal Protocol (Epstein et al., 2014), to the preservation of irrigation systems throughout local municipalities, such as those found in Nepal (Lam, 1998). But, the commons are not a panacea for all of the myriad complex social dilemmas that face the production and provisioning of shared resource systems. Like governments and marketplaces, the commons has its own set of costs and benefits. The work of political economists like Elinor Ostrom shows us that governance models that work in one socioecological setting, may prove ineffective in another (Ostrom, 2010).

The overaching goal of this dissertation is to understand which governance models work best, in which sociotechnical settings. In particular, I focus on research and develop settings in contemporary science.

1.1.1 The Science Commons

Various forms of commons are in use or currently being explored within organizations invested in scientific knowledge production. These include small community resources such as the Weather Research and Forecasting model (WRF) model in meteorology, all the way up to national level funding strategies found in the European Commission's 8th Framework Program for Research and Technological Development. The National Institutes of Health (NIH) Big Data 2 Knowledge (BD2K) Program has initially described its mission as commons-based:

> "The Commons is above all else a conceptual framework for a digital environment to allow efficient storage, manipulation, and sharing of research objects...The Commons belongs to and affects the whole research community... the concept relates to the entire global biomedical research enterprise, the NIH does not own it, so it is not the NIH Commons; similarly it is not just for data and hence is not the Data Commons. Rather it is the concept of sharing digital research objects from any domain, where sharing implies finding, using, reusing and attributing. The Commons could be considered analogous to the Internet each user has his/her own definition of exactly what the Internet is, but all are able to use it every day for their own purposes. No one seems to own it yet it works because each participant abides by a simple set of agreed-upon rules." (NIH, 2014)

While commons governance *can* be just a simple set of rules they can also be complex, multi-scale and multi-level institutional arrangements. Just as in socioecological systems, sociotechnical systems that enable scientific knowledge production are likely require different kinds of governance systems in different contexts. Sustaining the different sets of resources systems needed to address grand-challenge science issues like climate change, the spread of infectious diseases, or biodiversity loss will require diverse institutions cooperating effectively in different arrangements. In short, an understanding of the types institutions that most effectively enable cooperative scientific work will require a systematic, and sustained research agenda as much as an individual study.

This research project will contribute to a better understanding of the institutional arrangements that enable long-term scientific cooperation thorugh the development of a systematic approach to sociotechnical systems sustainability. This will be achieved in two ways:

1. The use of an emerging Knowledge Commons Framework (KCF) to conduct a case study of the International Comprehensive Ocean and Atmosphere Dataset (ICOADS), a project in marine climatology that has successfully sustained a cooperative model of knowledge production for over thirty years. I compare the results of this work with three previously completed case studies of genomics, astronomy, and biomedicine.

2. By adapting and modifying a protocol from the Social-Ecological Systems Meta-Analysis Database (SESMAD) for systematically coding variables related to different components of the Knowledge Commons Framework. My adaptation of this protocol focuses specifically on governance. I show how standard coding of relevant system state variables[1] allows for meaningful inter-case comparison with data collected about ICOADS, and has the potential to be used in diverse sociotechnical settings.

Structure of Document

In the rest of this chapter, I will describe existing and emerging approaches to commons governance, outline the difficulties of governing for sustainability, define a set of key concepts used throughout the project, and finally describe the setting and context for a case study of ICOADS.

In Chapter 2, I review relevant literature and further describe the design choices for this case study.

Chapter 3 describes data collection methods, and the approach to organizing and analyzing the results of my case study.

[1]in other words, the characteristics of system at a particular period of time (Walker et al, 2005)

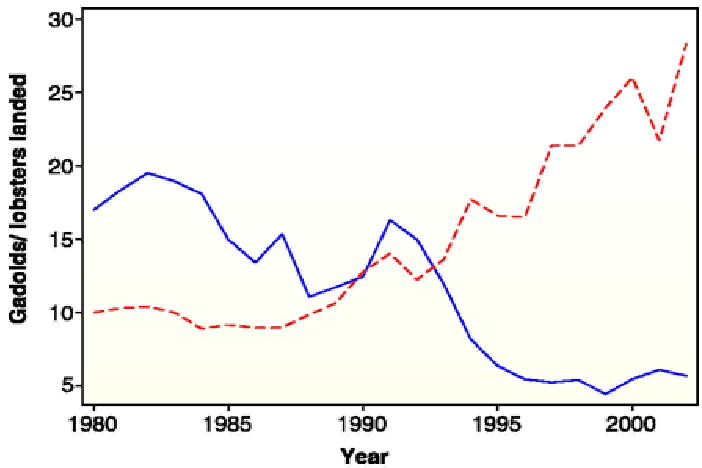
Chapter 4 presents findings from the case study research, which is organized by the Knowledge Commons Framework (Frischmann, Madison, and Strandburg, 2014). Within this framework, I also draw on a systematic approach to coding system state variables related to governance by adapting portions of the Social-Ecological Systems Meta-Analysis Database (SES-MAD) schema (Cox et al., 2014).

Chapter 5 summarizes the findings of the case study, and provides direct answers to a set of formally stated research questions. In doing so, I compare results from the ICOADS case study with previous case studies completed using the Knowledge Commons Framework.

Chapter 6 states the implications of these results for policy, practice, and theory of the commons; and for the governance of sociotechnical systems more generally.

1.1.2 The Struggle to Govern the Commons

The commons is a general term that can refer to a resource management approach, as well as the the shared resource system itself (Hess and Ostrom, 2005). This dual-distinction is evident in the etymology of the word "commons"; the Latin root being "*communis*, which signifies something held in common by a group, but also a user community bound by responsibilities as well as rights "(Disco and Kranarkis 2013, p. 14). Until very recently the study of how commons function and evolve has focused almost exclusively on natural systems; examples include grazing pastures in Switzerland, aquifers in Mexico, and forests in Western Africa. Previous scholarship in this domain, including the "tragedy of the commons" (Hardin, 1968), had created an artificial distinction between private ownership and government regulation. The refutation of this commonly accepted wisdom was pioneered by Elinor Ostrom, who focused on how institutions for collective action develop rules to govern shared resource systems, and what role these institutional arrangements have in creating cooperative, sustainable commons (1990). Ostrom and colleagues used a variety of methodological approaches, including laboratory work, field studies, and surveys, to show that state or market models rarely correspond with the way that successful shared resource systems are governed in the real word. Instead, they often found that users, producers,



Two landings in millions of kilograms over time. Dotted lines show lobster populations thriving, and Gadoids collapse, both in the same location, but under different governance regimes.

and provisioners of a natural resource system created self-governing systems which outperformed state and marketplace models.

A salient example of the unintuitive results of this work comes from the fisheries literature where landings regulations (total amounts of a catch measured in millions of kilograms) are managed under different governance regimes. In the same geographic locations facing the same biophysical conditions, certain fisheries collapse, like the *gadoids* managed under State regulation, while other fishiered thrive, such as *lobsters* managed under a self-organized commons regime.

To put it plainly, governance matters. A large body of research on the commons shows us that the effectiveness of rules, sanctions, and policy instruments depends greatly on the context in which they are deployed; gadoids and lobsters may have different reproductive cycles which require different harvesting rules; technologies for trawling gadoids grounds may have advanced more rapidly than the caging techniques used in the lobster industry; or, it may be that a chemical change in the feeding grounds impacted the two species differently. Without a systematic and controlled way to study these different variables, the evaluation of policy effectiveness is thin, and likely incomplete. In developing a comprehensive framework for studying these types of related socioecological systems, commons scholars like Ostrom have been able to generate a deep understanding of which types of social dilemmas require which kinds of governance systems (Acheson, 2012). An empirical, systematic approach greatly increases the chances that a particular policy or a particular design intervention will be effective for sustaining cooperative arrangements.

1.2 A Systematic Approach to Governing Knowledge Commons

The NIH BD2K program may see the commons as a solution to all of the social dilemmas faced by producing and provisioning a research infrastructure, but it's unclear how effective this model will be, for whom will it be effective, and for how long. What's needed is a systematic approach to science commons in sociotechnical systems, just as a systematic approach to natural commons has been taken for socioecological systems. Just as what works in one socioecological context may not work in another, a one-to-one mapping will not be perfect between the natural resource systems of Ostrom's study, and the purposefully designed and engineered systems that are the subject of this dissertation. Sociotechnical systems are subject to a variety of different social dilemmas than the natural commons; including the coupling of social and technical phenomena which are not well suited for the types of frameworks that have been developed to study "biophysical characteristics" of the natural commons.

An emerging concept in the commons literature is the notion of "culturally constructed" resources managed as "knowledge commons."(Madison, Frischmann, and Strandburg, 2010). Where natural commons are the institutions, resource sets, and interactions governed in a socioecological system, the knowledge commons are the cooperative arrangements that share, curate, produce, provision, and sustain informational resources within a sociotechnical system. Governance in the knowledge commons is, similarly, critical to sustainable knowledge production. As Frischmann, Madisson and Standburg, explain, "The nested, multi-tiered character of productive and sustainable knowledge and information systems and the diversity of attributes that contribute to successful governance regimes are key to understanding knowledge commons as mechanisms for knowledge production, collection, curation, and distribution in the context of modern information and IP law regimes." (2014)

Recognizing that a binary distinction between markets and states is an oversimplification of the organizational models of governance in contemporary science, the commons begin to appear ubiquitous in contemporary research and development settings: knowledge commons are the shared libraries of open-source software that run high-performance computing centers; knowledge commons are the research data archives which enable whole earth climate simulations on a petabyte scale; and, knowledge commons are the instrumentation at the center of astronomical observatories, such as the aptly named "Very Large Telescope" in Chile.

But, although knowledge commons appear to be a pervasive phenomena in science and technology settings there is little empirical understanding of how these types of informal institutional arrangements are successful over time. There is a lack of understanding about which governance arrangements enable sustainable knowledge production in which contexts, And, there is yet to emerge a systematic way of collecting data to compare across these different contexts. This dissertation therefore pursues two broad research questions:

RQ 1. What are the effective institutional arrangements (governance) for sustainable scientific knowledge commons?

RQ 2. How do these arrangements differ between domains of knowledge production?

These questions are answered by conducting a case study of a knowledge commons in the domain of climate science. By adapting two frameworks used to study socioecological systems to analyze data collected in the case study, I attempt to show the benefit of this approach as applied to the sustainability of knowledge commons in science. I achieve this by comparing these case study results to three previously completed case studies, and synthesizing the findings across all four cases.

Before introducing the case study subject, the International Comprehensive Ocean and Atmosphere Dataset (ICOADS), I describe a set of basic concepts that will be used throughout this document.

1.3 A Controlled Vocabulary

This section includes a controlled vocabulary for a set of terms which are important to the overall thesis being pursued. What follows should not be read as strict definitions, but rather a guide for set of concepts which may be used in different ways, by different disciplines.

Commons

The commons is a generic term that can refer to a resource management approach, as well as the shared resource system being managed. In the broadest sense, the commons are marked by "privileges and immunities for an undefined public, rather than rights and powers for a defined person or persons...The main function of commons is to institutionalize freedom to operate, free of the particular risk that any other can deny us use of that resource set, subject to symmetric known constraints and the risk of congestion applicable to that resource set, under those rules, within the expected population of users." (Benkler, 2014). The two classes of commons that have been described thus far are, on the one hand pastures, forests, and irrigation districts of the natural world, and on the other, high-performance computing, software libraries, and data archives of the digital realm. This dissertation deals exclusively with the latter type of commons.

Sustainability

Sustainability can be a relative term. In the context of this project, a sustainable knowledge commons is a sociotechnical system that over the long run, enhances both the quality and the resource base on which science depends, provides for the continued support of resources (such as data, or software, or instrumentation), is economically viable, and enhances the quality of science being conducted. In this sense, a sustainable knowledge commons doesn't just persist over time, but evolves given different external and internal pressures. This idea is a central theme of Chapter 5.

Sociotechnical Systems

This project conceives of a sociotechnical system as the mutual constitution of people and technologies in social, political and economic settings that require collective action in order to effectively function over time (Sawyer and Jahairi, 2013). The contextual and embedded nature of sociotechnical systems makes governance, institutional arrangements, and symmetric information exchange paramount to their success.

Infrastructure

Infrastructure, like sustainability, is an inherently relational concept. In a science and technology setting, infrastructures can be defined in relation to organized practices of communities, disciplines, or fields of study (Star and Ruhleder, 199). The term of art for scientific infrastructures has recently become cyberinfrastructure (Atkins, 2000), which I take to mean "...the set of organizational practices, technical infrastructure and social norms that collectively provide for the smooth operation of scientific work at a distnace. All three sets are objects of design and engineering; a cyberinfrastructure will fail if any one is ignored"(Edwards et al., 2007, p. 6)

Given these explanations, a natural question is what are the blurred lines between infrastructures, cyberinfratructures, sociotechnical systems, and commons?

Commons focus on the institutional arrangements, with an emphasis on the rules and governance of people, resources, and the bundle of property rights that are negotiated for their long-term sustainability. Infrastructure, and cyberinfrastructures, are a congruent, and complimentary view of these interconnected elements. Infrastructures can (and often are) managed as a common property (Frischmann, 2005), for which collective action is required to keep the "smooth operation of scientific work at a distance" occurring. Although two literatures - commons and infrastructure studies - make reference to one another, analysis of their relationships are rarely combined (a notable exception, Frischmann, 2005). For instance, commons scholars increasingly acknowledge the importance of developing an account of the infrastructural resources that shape, and are shaped by long-term interactions within shared resource systems (Dietz et al., 2010). Part of the goal of this research project is to better align the findings from a long line of cyberinfrastructure studies, and the emerging knowledge commons frameworks described in Chapter 2.

Governance

A helpful definition of governance is that it includes, "a complex of public and/or private coordinating, steering and regulatory processes established and conducted for social (or collective) purposes where powers are distributed among multiple agents, according to formal and informal rules" (Burns and Sthr 2011: 234). As stated above, a sociotechnical perspective recognizes that collective action is needed to sustain the social relations, orderings, and enforcement of cultural norms, as well as the technical components that allow a commons to effectively function over time. A governance model, a term also used throughout the dissertation, implies the sets of“institutional arrangements (such as rules, policies, and governance activities) that are used by one or more actor groups to interact with and govern” shared resources (Cox et al., 2015).

Governance models typically differ in their centralization (or decentralization) of decision making power - such as self-governing or monocentric governance. This dissertation explores polycentric governance models that nest authority at multiple levels, types, sectors, or jurisdictions. A preliminary defintion of polycentric is it is a “structural feature of social systems of many decision centers having limited and autonomous prerogatives and operating under an overarching set of rules.” (Aligica and Tarko, 2011). The overlap of these different levels create collective action dilemmas, which require multiple rule types to function efficiently. Polycentric models, as I discuss in the following chapter, are increasingly effective for helping sociotechnical systems cope with social dilemmas related to sustainability.

Resilience

The notion of resilience in socioecological systems is partially evolutionary, and partially ecological. Holling formulated this idea in the early 1970s, defining resilience as a “measure of the persistence of systems and of their ability to absorb change and disturbance and still maintain the same relationships between populations or state variables.” (Holling, 1974) In Chapter 5 I focus on transitions of ICOADS between different governance regimes; the ability of a commons to routinely make these governance transitions is described through *resiliency processes*. The NSF’s Critical Resilient Interdependent Infrastructure Systems and Processes (CRISP) program offers a helpful definition of a resilient process for infrastructures; these are “the features of an infrastructures inherent capacity to resist disturbances, initial loss of service quality, and trajectory of service restoration. Conceived as a process, infrastructure resiliency can be achieved by a myriad strategies in addition to simple repair and replacement.” (2013). The “myriad strategies” of ICOADS resiliency are a subject explored throughout this project.

1.4 Research Context and Rationale

I turn next to an introduction of the subject of my case study, and the major subject of data collected for this dissertation. A thorough discussion of the background environment for ICOADS appears in Chapter 4.

Context

The International Comprehensive Ocean and Atmosphere Dataset (ICOADS) is a cooperative project that curates, develops and distributes quality controlled data, metadata, historical documentation, and software to the climate science community. The project, originally named COADS, was initiated in 1981 by researchers at the Earth System Research Laboratory (ESRL), the National Climatic Data Center (NCDC), and the National Center for Atmospheric Research (NCAR). Over time the project grew to include international collaborators and the name was changed to the "International COADS" in order to reflect the contributions of organizations like the World Meteorological Organisation (WMO), the Intergovernmental Oceanographic Commission (IOC), and the Technical Commission for Oceanography and Marine Meteorology (JCOMM).

Contemporary data curated by ICOADS come from a variety of sources, including the Global Telecommunications System (GTS), in-situ measurements taken by sea-faring vessels, earth observing satellites, and both drifting and moored buoys. Historical data come from an effort in the early 1980s to aggregate existing marine data records from maritime archives around the world, as well as a continual stream of historical records that have been rediscovered and newly digitized; this now includes the use of crowd-sourcing efforts to transcribe weather recordings taken by military, shipping, and whaling voyages from the 17th, 18th, and 19th century. These early records are significant culturally and historically, as "Sailors were among the first to systematically record the weather because the states of ocean and atmosphere controlled their progress and survival (Woodruff et al., 1986 citing Quayle, 1977)".

The labor-intensive process of taking heterogeneous records from different observing platforms, and uniformly processing and integrating the data into a larger set of historical observations is the major value of ICOADS ongoing work. This includes the preservation of provenance metadata that is recorded for each individual record, allowing for researchers to trace backwards in time to verify whether the source of a climate anomaly is genuine or the result of a data processing error.

Free and open access to ICOADS has helped it become recognized by the climate community as the "most complete and heterogeneous collection of surface marine data in existence" (Woodruff et al., 2011). ICOADS data have been used extensively in international climate assessments such as the IPCC AR4 and AR5 reports, as well as reanalysis projects that combine historical data with contemporary weather observations to create authoritative datasets for the climate modeling community (Kalnay et al., 1996).

The success and widespread use of ICOADS has not resulted in greater stability for the funding of the project. This is due in part to the difficulty in calculating the research impact of many different ICAODS products (Weber et al., 2014), as well as an overall decrease in federal funding for the maintenance of research infrastructures (Berman and Cerf, 2013). The politicization of climate related research has also impacted ICOADS funding in recent years as congressional pressure to defund "climate research" continues to mount. In the winter of 2012, this became a major issue for the sustainability of the project, as NOAA announced that:

> "For budgetary reasons, stemming from pending large cuts at the NOAA Climate Program Office (CPO), ESRL Directors have determined that it is no longer feasible for its Physical Science Division (PSD) to continue supporting any further ICOADS work effective immediately" (Lawrimore, 2012)

In response, project partners at the National Center for Atmospheric Research (NCAR), the UK Meteorological Office, and the Deutscher Wetterdienst (German Meteorological Service) signed a memorandum of understanding to continue contributing to the maintenance of the project, and its various resources. The research of this dissertation was conducted beginning at the point that signatures for a memorandum of understanding were signed.

Rationale

For many of the reasons described above, ICOADS offers a unique case study of sustainability in a knowledge commons. These include, but are not limited to:

- Project partners are diverse in terms of their organizational affiliations, and expertise. As such, ICOADS is nested within a number of overlapping governance structures.

- Climate science is a unique domain of knowledge production in that it requires a broad scheme of cooperation in order to generate verifiable research results. And further, these results are some of the most scrutinized, and politically charged forms of knowledge production.

- The resource sets provisioned by ICOADS have persisted over a thirty year period that has seen a number of fluctuations in funding, the politicization of the subject matter, and rapid technological change. ICOADS therefore presents an opportunity to better understand how diverse resource sets, consisting of software, data, human expertise, and computational infrastructures are sustained over time.

The specific research questions that this case study will answer are related to the ways in which, over time, ICOADS project governance has evolved in order to sustain a unique set of shared resources. Stated formally, the research questions answered by this case study will be:

- **CS RQ 1** : What are the different governance models that ICOADS has effectively used to manage shared resources over it's thirty year existence (1983-present)?

- **CS RQ 2** What causes a governance system to shift from one regime to another?

- **CS RQ 3** What types of disturbances are ICOADS resilient or vulnerable to?

1.5 Summary

In this chapter I have defined a set of research problems that have been under-examined in the current literature, namely how knowledge commons in scientific research and development settings are sustained over time. I have identified a number of systematic approaches used to study socioecological systems, and suggested ways in which these can be a usefully adapted to the study of sociotechnical systems sustainability. I offered a controlled vocabulary for concepts that will be important to the entirety of this document, including sustainability, commons, governance, sociotechnical systems, infrastructure, and resilience. Finally, I briefly described the subject of my case study, ICOADS, and justified the choice of this case.

I also offered two sets of research questions; the first set of questions are aimed at understanding sustainability in knowledge commons more generally. This set of questions will be answered by comparing the results of a case study of ICOADS with previously completed case studies in Biology, Astronomy, and Biomedical research networks. The second set of questions - which are specific to the ICOADS case study - are answered through the use of a range of data collected and analyzed through a framework that is described in detail in Chapter 3. In the next chapter, I review literature relevant to this study, and further contextualize this research project.

CHAPTER 2

2.1 Introduction

This chapter reviews relevant literature on governance and sustainability of shared resource systems, including the various conceptual models which inform commons theory. I focus in particular on aligning the rules-based approach of socio-ecological systems with those of sociotechnical systems.

I begin with an overview of economic theories of organizing around shared resource systems. I then describe the emergence of commons theory, and its relationship to sustaining socio-ecological systems. Finally, I relate previous research on the natural commons to emerging notions of the knowledge commons, which brings with it new models of peer production, and use of networked information communication technologies.

2.2 History of the Commons

Economic theory in the first half of the twentieth-century generally recognized two models of organization: the market and the state. Classical economic thinking holds that the government (state) should be responsible for provisioning the types of public services that may go under-produced if left open to pure marketplace competition. In contrast, the market should be an unregulated domain where pricing signals direct self-interested actors exchanging goods and services to satisfy their preferences. In short, the market is the domain of production and exchange between bakers, butchers and brewers; and the government is the domain of provision, such as those required for lighthouses, railways and infrastructures to effective operate.

2.2.1 The Market: Coase, Williamson, and Transaction Economics

In 1937 Ronald Coase began to upend common assumptions about economic efficiency of organizing in the marketplace. This began with a publication titled "The Nature of the Firm"; a short, and largely anecdotal article in which Coase asks a simple question: Why do firms - hierarchicaly organized models of production- exist if markets are efficient for satisfying individual preferences? His explanation follows:

> "...an alternative form of economic organisation which could achieve the same result at less cost than would be incurred by using the market would enable the value of production to be raised...the firm represents such an alternative to organising production through market transactions... *Within the firm individual bargains between the various cooperating factors of production are eliminated and for a market transaction is substituted an administrative decision. The rearrangement of production then takes place without the need for bargains between the owners of the factors of production.*" (Coase, 1932)

According to Coase, the relationship between firms and markets is as follows - when the cost of achieving an outcome is cheaper by organizing activities in a hierarchy, firms will emerge, but when the marketplace creates a pricing system with less costs, individuals (or small institutions) will follow pricing signals. Therefore, economists interested in the private sector should focus on how choices between such alternatives are made. Further, Coase turned marketplace behavior from a theoretical problem, to an empirical one by showing the market vs. firm distinction is an organizational phenomena. As Coase put it, the task for those interested in the organization of economic systems of production is this:

> ...to bridge what appears to be a gap in [standard] economic theory between the assumption (made for some purposes) that resources are allocated by means of the price mechanism and the

> assumption (made for other purposes) that that allocation is dependent on the entrepreneur-coordinator. We have to explain the basis on which, in practice, this choice between alternatives is effected. (1932, p. 389)

The focus of this line of thinking then became the grounds for transaction cost[1] economics" which Oliver Williamson would go on to use for a predictive theory of economic organization, which he described was meant to "breathe operational content into governance and organization" of production systems, where governance could be defined as "...the means by which to infuse order, thereby to mitigate conflict and realize mutual gain..." (Williamson, 2010).

The research agenda of transaction economics then required reformulating the problem of economic organization, from the individual behavioral explanation of rational actors, to one built on comparative contractual terms. This was accomplished by the following steps:

- Naming the key attributes with respect to which transactions differ,
- Describing the clusters of attributes that define alternative modes of governance (of which markets and hierarchies are two),
- Joining these parts by appealing to the efficient alignment hypothesis, wherein
- Predictions would be derived to which empirical tests would be applied and
- Public policy ramifications could be worked up. (Williamson, 2010)

In parallel with Williamson, Elinor Ostrom was working to "breathe empirical life" into the governance assumptions that classical economics had made about the provisioning of environmental goods, especially community managed resource systems. This research agenda took on assumptions that solutions to the provision of public goods, those typically the domain of the state, were most efficient when organized by a central authority. Through a diverse body of work that included the study of police departments, watersheds, fisheries, and forests, Ostrom and colleagues showed that, like Coase

[1]In this work, the unit of analysis is a transaction, defined as "the three principles of conflict, mutuality, and order. This unit is a transaction" (Common, 1932, p. 4).

and Williamson, classical economic ideas about efficiency rarely matched reality. What the parallels in these two research agendas teach scholars of political economy, in any domain, is that understanding attributes of governance, how they differ in efficient application, and the policies that can be consequently derived, will require a systematic empirical approach that can be applied across cases.

2.2.2 Section Summary

Both of these ideas - the emergence of the firm, and transaction cost economics - were critical for the study of alternative institutional arrangements and governance mechanisms in the marketplace. First, Coase showed that it was not happenstance, nor ideology that drove organizational structures to adopt one governance regime over another; instead this was the result of individuals trying to efficiently solve marketplace dilemmas by assessing costs and benefits related to following a hierarchy, or following pricing signals. Later, Williamson demonstrated that an analysis of transaction costs could be systematized, making the study of contracts and institutions more useful to public policy decisions that regulate marketplaces.

I turn next to the debates about property and the organization of governance regimes for the provisioning of public goods; the controversies out of which commons scholarship was borne. The next two sections trace the intellectual history of the commons. I then describe the influence that new modes of production, namely peer-production, are having on organizing shared resources systems in order to establish the commons based research agenda to which the findings of this dissertation will contribute.

2.2.3 ‘'Tragedy of the Commons’

In 1968 Garret Hardin the bioligist penned a short, but highly influential article in ‘Science’ that described a hypothetical dilemma facing herdsman in the shared resource system of an open pasture where consumtpion is unrestrained. As he writes,

> Each man is locked into a system that compels him to increase his herd without limit - in a world that is limited. Ruin is the

> destination toward which all men rush, each pursuing his own best interest in a society that believes in the freedom of the commons. Freedom in a commons brings ruin to all. (Hardin, 1968 p. 1244)

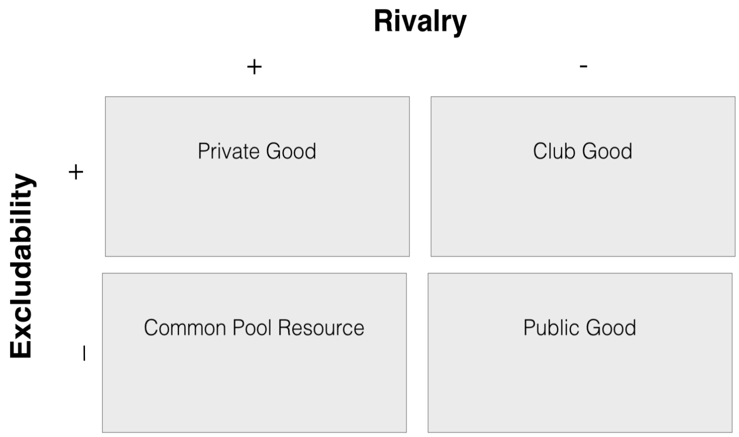
Hardin's 'Tragedy of the Commons' is a standard externality problem that is the result of failed collective action, built on an assumption that resource users act independently, and of their own self-interest. The solution set that Hardin offers consists of two approaches from classical economics: Enclosure, in which the land is privately parceled out for individual ownership, or management by a centralized state government which could prohibit overgrazing through regulation.

A number of scholars have since pointed out, Hardin's scenario is flawed in a number of ways; the resource system he describes is not necessarily a commons, but an open-access regime, and the assumptions made about rational actions independent of collective conscious turns out to be greatly oversimplified when compared with real world scenarios (Taylor, 1990). But, Hardin was correct in recognizing the troubles that befall shared resource systems when communication between users is limited.

To better understand shared resource management, and how it operates in practice it is necessary to first review the classification of goods based on two dimension: their rivalry and their subtractability. Then, by looking closely at the governance regimes that are adopted to manage access and use of goods across this classification, it will become both clearer why the tragedy of the commons scenario is flawed, and what is really at stake in the design and adoption of different governance models for shared resource systems.

2.2.4 Classifying Goods: From Samuelson to Ostrom

The classical economist Paul Samuelson was the first to formally describe differences between private and public goods. He was careful to point out that although his examples represented two extremes most government activity could be analyzed as some blend of two poles: private and finite, or public and infinite (1950). Samuelson arrived at these two distinctions by way of a theory about rivalrous vs. non-rivalrous commodities. A good with rivalry can be owned or consumed by a limited number of individuals - the good is



The classification of goods by degrees of rivalry (substitution) and excludability

finite in this sense, and its value is largely a function of the limited nature of its ownership. A good with no, or even low degrees of rivalry can be shared among a large number of individuals without the value of the good being significantly harmed - its value then is a function of its accessibility and shareability.

Initially, it was rivalry that concerned political economists thinking about the production and provision of goods. Over time, Sameulson's second dimension was more carefully considered, and it it is from this body of literature that common pool resource theory emerged (Agrawal, and Boettke, 2011). In the second division of goods by type, the question theoretically turns on whether or not it is economically efficient to create barriers (exclusion) to the consumption of a resource.

A swimming pool is an excludable resource where it is economically efficient to bar entry so that a select (proper) number of swimmers can enjoy the pool. The ocean, on the other hand, is a resource for which it is economically inefficient to prohibit swimming access. Artificially creating a barrier to prevent use of an ocean by swimmers or sailors would be prohibitively expensive for each nation of the world, and no one person has any greater claim of right to access the ocean than any other. The resource therefore remains relatively non-excludable [2].

[2] Beaches of course have swimming rules, and there are boundaries erected for the safety of swimming in open waters, but the larger point is that creating barrier to entry, for sheer size and right to access, is prohibitive

Each of these categorizations have to do with resources (goods) as property - where some can have exclusive ownership, while others can have the power of ownerships symmetrically distributed (collectively shared) by a community of interested parties. Commons - including open access commons described by Hardin - are marked by the absence of asymmetrical power to determine the rights of the resource. In other words "they are marked by privileges and immunities for an undefined public, rather than rights and powers for a defined person or persons." (Benkler, 2014 p. 3) Hardin's argument in the tragedy of the commons confused the type of resource with its governance regime. Hardin described a completely open access public goods regime used to manage common-pool resources, claiming that the failure was the result of self-interested individuals incapable of acting in a way that did not maximize their own self-interest.

Elinor Ostrom pioneered the study of "common pool resources" (CPRs), or goods with rivalrous and non-exclusive access such as grazing pastures, aquifers, or forests (1990). Over three decades worth of work, Ostrom and the "Bloomington School" of scholars convincingly showed that there is not a one size fits all solution for governance of CPRs (Aligica and Boettke, 2009). Common-pool resources can just as often be successfully managed by communities with bundled legal rights (a commons) as they can with formal government intervention. But, a repeated finding of this work is that for small, well-bounded common-pool resources, a community-based self-governance model outperforms a state model, especially in negotiating terms for long-term sustainable collective action (Ostrom, 2010).

In short, there is no governance panacea; a single property-rights regime can produce, and effectively provision several kinds of goods, and several kinds of goods can be effectively produced or provisioned under different regimes (Hann 1998; Berge 2002; Benda-Beckman et al. 2006). What works in one domain may fail spectacularly in another. The questions that result from this realization are twofold:

1. Which type of governance regimes most effectively provision the types of goods or services which require collective action (e.g. public goods, common-pool resources, infrastructures, etc.)?

2. How do novelties in the design of these institutions impact their ability to provision resource sets? That is, if context matters, what effect does

it have on the design of similar commons in different contexts, and different commons in similar contexts.

Across the different contexts that she studied, Ostrom's work establishes five different kinds of communal property rights:

1. Access the right to enter a defined physical area and enjoy non-subtractive benefits (e.g. hiking, skiing);

2. Extraction the right to take products of a resource system (e.g. to fish, divert water);

3. Management the right to regulate use patterns and make improvements on the resource;

4. Exclusion the right to determine who will have access and withdrawal rights;

5. Alienation the right to sell or transfer management and exclusion rights. Varying combinations of these rights are associated with statuses. Owners have all of these rights; authorized users can only access a property and withdraw resources.

In most commons settings these bundled rights are established through rules, some formally and legally enforceable and others informal and community policed. Ostrom draws a distinction between these two rules - those written down and codified being the "rules-in-form" and those unspoken, tacit, and community enforced being the "rules-in-use". Critical to understanding the proper functioning, and long-term success of commons then is finding ways to observe rules-in-use for effectively dealing with social dilemmas, such as collective action problems found in the 'Tragedy of the Commons'. Michael Taylor offers a succinct definition of collective action, as follows "Collective-action problems occur when there is a divergence between the interests of the individual and those of the society. In these cases, it is not rational for individuals to cooperate, even though cooperation would bring positive results for all." (Taylor 1990) In some sense - this is the broad economic concern for efficiency - asking what system, arrangement of resources, or policy intervention creates the greatest good. But collective action is a micro-economic concern, because it focuses on how these types of dilemmas are solved amongst individuals with competing interest. These solutions, especially in the context of environmental common-pool resources, can be attributed to rule types that establish and in some cases prescribe a number of different expectations for actors in the commons:

- "Boundary rules specify how actors are to be chosen to enter or leave a situation
- Position rules specify a set of positions (council members, president, etc.) and how many actors hold each one
- Information rules specify channels of communication among actors and what information must, may, or must not be shared.
- Authority rules specify which actions are assigned to a position at a node
- Aggregation rules (such as majority or unanimity rules) specify how the decisions of actors at a node are to be mapped to intermediate or final outcomes
- Scope rules specify the outcomes that could be affected
- Payoff rules specify how benefits and costs are to be distributed to actors in positions." (McGinnis, 2011)

In addition to rule sets, Ostrom and colleagues synthesized many studies of CPRs throughout the world by way of a systematic frameworks for both gathering and analyzing empirical data. The results from this synthesis are a set of design principles for shared resource systems that, at a very high level, account for the features which often (but not always) lead to commons success:

- Clearly defined boundaries should be in place.
- Rules in use are well matched to local needs and conditions.
- Individuals affected by these rules can usually participate in modifying the rules.

- The right of community members to devise their own rules is respected by external authorities.

- A system for self-monitoring members' behavior has been established.

- A graduated system of sanctions is available.

- Community members have access to low-cost conflict-resolution mechanisms.

- Nested enterprises that is, appropriation, provision, monitoring and sanctioning, conflict resolution, and other governance activities are organized in a nested structure with multiple layers of activities. (Ostrom 1990, 90102)

According to Ostrom (1999, 2000), community governance is most successful in sustaining shared resources when social capital is symmetrically distributed (i.e. few gatekeepers), communities are small, local, well mapped geographically, and those that are most dependent on the resource have a what is known as a "low discount rate" or a "willingness to sacrifice current payoffs for higher payoffs in the future." (Acheson, 2011).

Comedy of the Commons

While Ostrom and the Bloomington School thrust commons theory into the academic spotlight during the early 1990's, it is often assumed that the social dilemmas of common-pool resources are the social dilemmas faced by all commons. The types of commons that deal with socio-ecological systems (pastures, forests, aquifers, huertas, etc.) often face dilemmas that are based on excludability - such as the free rider problem which emerges in open-access regimes (e.g. Acheson, 1990), or the zero-contribution thesis of collective action which occurs in large cooperatives (Olson, 1965). But, excludability is only one type of social dilemma faced by commons. In commons that include the production and provision of public goods, which are often assumed to be efficiently provisioned by governments, problems of congestion (overuse) and free-riding emerge that are similar to, but different than those faced by CPRs.

Work that addresses these dilemmas began with Carol Rose, a lawyer, in the late 1980's writing about the "inherent public property" of a certain class of goods - such as waterways, the atmosphere, and public beaches. Rose asked how, and to what degree, common law establishes a set of rights for public access to these resources (1986), recognizing that it is both costly to exclude users, and that in the absence of congestion these types of goods create positive spillover effects. In short, public goods are the infrastructure for a number of downstream innovations - and the freedom to operate within those infrastructures should be inherent, not just by virtue of their being large and difficult to exclude access, but in their governance. Rose coined a phrase to refer to the spillover effects of these common rights as a "comedy of the commons"; and she then went on to show, through a number of case studies, how in distributing access and usage rights of public goods to individuals made resource users complicit in the effective provisioning of the shared resource system. As Rose argues, investment in such coordinating costs would be prohibitively expensive using either a market or state based model of organization. It was only through a commons, with symmetrical rights to operate effectively shared by all resources users, that such cooperation could be sustained.

Section Summary

In the following sections, I explore commons outside of the CPR context. I look specifically at how governance regimes for sustainable commons evolve with the introduction of networked information communication technologies, focusing on science and technology research settings. Much of this work is predicated on the idea that for informational goods, collective action problems emerge around both the production and provisioning of goods. Where scholars like Ostrom, Acheson, and Rose are concerned with resource sets that are already in place, other commons which are produced, such as open-source software (Schweik and English, 2012); Genomic data archives (Contreras, 2012); and Telecommunications infrastructures (Frischmann, 2005) face dilemmas of collective action at both a point of initial production, as well as sustaining and provisioning the resources over time. In these domains, the initial terms upon which cooperation is founded play an important, and non-trivial role in the structure of later governance models (O'mahony and Farraro, 2007).

To conclude this section, I've described three things:

1. A widely used categorization of resource (goods) based on two criteria: Rivalry and Excludability
2. A distinction between the classification of property types (goods), and governance; and, how confusions in this distinction has impacted the historical conceptions of the commons.
3. The empirical research agenda launched by Elinor Ostrom to study the governance of commons, including CPRs where social dilemmas related to excludability often emerge, but are sustained through establishing and enforcing rules of two types: Rules-in-Use, and Rules-in-Form.

2.3 Commons Unmodified

As Yochaie Benkler notes, Rose's work on public goods was an important precursor to commons theory developed for networked communication technologies (Benkler, 2001), and public infrastructures (Frischmann, 2005) as well as informational goods such as creative works in the public domain (Lessig, 1999) or open-source software (Lakhani and VonHippel, 2003).

Benkler provides an framework for distinguishing between Rose and Ostrom's commons:

- **Modified** commons are aimed at describing the types of resource systems studied by Ostrom and colleagues - physical, well bounded, and long-enduring. Pastures, aquifers, huertas, are the quintessential case of the modified commons.
- **Unmodified commons** refer to public goods, which share the a of excludability with modified commons, but also have low degrees of rivalry in their consumption. Environmental unmodified commons include beaches, open-roads, the atmosphere, and oceans. In the informational sphere, unmodified commons are the platforms and infrastructures upon which many facets of the modern economy are based; examples include the broadband spectrum, Internet exchange protocols, suites of open-source software, and weather data.

For modified commons studied by the Bloomington School, the following were generally true:

1. Resource boundaries were clear.
2. Resource systems were small and easy to observe
3. Solving social dilemmas was of high importance to appropriators (funders)
4. Institutions were long enduring (centuries), and evolved rules and sanctions over equally long periods of time. (Hess and Ostrom, 2005)

Nearly each of these features are the opposite in an unmodified commons. These resource systems are typically large, and have unclear boundaries. Information-based unmodified commons, or knowledge commons, often form rapidly, and without plans for long-term maintenance in mind. This is especially true of the types of research and development in science, which is typically funded through five year grants of government subsidy (Mirowski, 2011). Finally, solving social dilemmas is of high importance to appropriators, but the size, lack of clear boundaries, and the rapid formation of unmodified commons greatly complicate this process.

The Knowledge Commons, Unmodified

Recall, from Chapter 1, that knowledge commons - including resource sets made up of informational goods - are to sociotechnical systems, as the natural commons - containing environmental resources - are to socioecological systems. Given these differences, two other important distinctions mark the unmodified knowledge commons, and I argue that both of these features are critical to understanding the role of governance for sustaining shared resource systems of this variety:

1. Unmodified knowledge commmons, are almost always purposefully engineered systems. That is to say, they are designed and built instead of naturally occurring. Simon's treatise on a ‘’science of the artificial’ is an important antecedent to this thread of unmodified commons work. Simon draws a distinction between what he calls science as an analysis of nature, and engineering as synthesis of the artificial (1981). His distinction is aimed at supporting a thesis about design, which explains that artificial phenomena “are as they are only because of a system's

being modeled, by goals or purposes, to the environment in which it lives. If natural phenomena have an air of ''necessity' about them in their subservience to natural law, artificial phenomena have an air of "contingency" in their malleability by environment." (1981, p. X)

According to this distinction, artificial things differ from nature in the following ways:

(a) Artificial things are synthesized (though not always or usually with full forethought) by human beings.

(b) Artificial things may imitate appearances in natural things while lacking, in one or many respects, the reality of the latter.

(c) Artificial things can be characterized in terms of functions, goals, adaptation.

(d) Artificial things are often discussed, particularly when they are being designed, in terms of imperatives as well as descriptives. (Simon, 1981 p. 13-17)

The purposeful design, and implementation of sociotechnical systems brings with it - as Simon notes of the artificial - a need to attend to the normative implications of those built environments. To analyze the effectiveness of governance in the unmodified commons, one needs to also be concerned with a set of values, or a background environment in which "artificial things" are designed and built.

Additionally, Simon notes that artificial things can be characterized by their functions, goals, and adaptation to contexts. Artificial things, like informational goods, are abiotic, but they nevertheless evolve in a context of use. One of the central concepts of Chapter 5's analysis is that like biotic things, selective pressures play a major role in how artificial things are adapted to meet those needs. I argue that an account of this brand of commons then needs to attend to the intentions of design, and to the adaptation of artificial things, in use. Similar to Ostrom's rules, Simon's point about artificial objects having functions and adaptations within different contexts creates a distinction between "things-in-form" and "things-in-use" - and as Simon instructs, one understands the "evolution and future of a system through its history." (1981, p. 56)

2. Related to the normative implications of design and engineering, there is a blurring of traditional user, producer, or consumer roles within the unmodified commons. This is especially true within shared resource systems related to informational goods, where resource flows are no longer bi-directional (producer to consumer), but are instead n-dimensional; producers of an informational good in one context often become users of the same resource in a different context. Producers are often working to provision goods, and users are often working to produce goods. Within contemporary sociotechnical systems many of the same actors are playing many different roles.. For instance, IBM both produces software that is proprietary and is simultaneously one of the largest contributors to open-source projects in the world. It provisions certain open-source libraries by providing labor and funding for their maintenance, and simultaneously it uses those libraries in its own distribution of software packages (O'mahony Lakhani, 2011). This breakdown in traditional roles requires analysis of the functioning of an unmodified commons over time, and across different levels and scales of analysis.

   Many studies from informatics (e.g. Wiggins and Crowston, 2011) and social network analysis (e.g. Borondo et al., 2013) have shown a defining characteristic of contemporary scientific work is that the traditional boundaries between user and producer are blurred. Stakeholders both use pooled resources, and contribute to the provisioning of those resources over time (Benkler, 2006). These blurred distinctions result in a number of social dilemmas, such as the continuation of a "Matthew Effect" in hyper-authorship networks (Glnzel and Schuber, 2005), identity disambiguation in the process of attributing knowledge claims (Fegley and Torvik, 2013), large collaborations that do not have a clear chain of ownership for intellectual property (Reichman and Uhlir, 2003), and the distributed nature of most eScience infrastructure development, which complicates a clear distinction between production and provision (Fry, Schroder and den Besten, 2011). Often the most successful collaborative projects include scenarios where an actor is enabled to modify, change, or improve a given technology for his or her own research while simultaneously making those modifications, changes and improvements available for others to use.

Purposefully built systems containing “artificial things” that evolve through use; and the collapse between traditional roles of economic actors in producing, using, provisioning informational goods are what marks sociotechnical systems as a unique branch of the unmodified commons research agenda. Next, I turn to reviewing literature which further describes social dilemmas that result from these unique factors, and the ways in which institutions successfully overcome these dilemmas. I focus specifically on previous studies of sustainability in science and technology settings, and conclude by describing the emergence of peer-production and its relevance to the research project at hand.

2.4 Sustainability

The socioecological systems literature contains a number of different conceptualizations of sustainability, three of which I'll review here: resource sufficiency, functional integrity, and for lack of a precise term the “normative account”. I describe these approaches in order to characterize contemporary research on the sustainability of sociotechnical systems found in science and technology studies.

Resource sufficiency is a technical approach to environmental stewardship that is generally concerned with two measurement problems related to sustainable practices. The first measurement problem is to accurately capture the rate of consumption of a resource. The second measurement problem is concerned with estimating the stock of available resources. To characterize sustainability of a socioecological system through resource sufficiency is to match consumption practices with rates of stock regeneration.

Functional integrity is a systems-based approach to sustainability practices. In this paradigm, a “practice that creates a threat to the system's capacity for reproducing itself over time is said to be unsustainable” (Thompson 1997). A functional integrity approach requires modeling reproduction cycles, and the practices that effect these cycles over time. Functional integrity could also be conceived of as risk management approach to systems sustainability.

Comparing and contrasting these two approaches: resource sufficiency is utilitarian in that it conceives of sustainability as an outcome of balancing resource stocks and flows (Moore, 2011). Functional integrity leans towards a communal, albeit cautious, view of risk in relation to the stability of broad systems, including social, economic, political, as well as environmental.

A third approach offered by both environmental economists and philosophers is to extend the systems thinking of functional integrity to a normative, values-based model. A **normative approach** emphasizes the impact that production and provisioning - otherwise grouped together as development practices - have on the environment, society, and the economy. An example of this approach is as follows, "We define sustainable agricultural development in this paper as an agricultural system which over the long run, enhances environmental quality and the resource base on which agriculture depends, provides for basic human food and fiber needs, is economically viable, and enhances the quality of life of farmers and society as a whole." (Davis and Lanham, 1995, p. 21-22). The normative approach is unique in that it conceives of a target that sustainable development practices or policies *should* be aimed.

Sustaining Sociotechnical Systems

In the following section, I map these three notions of socioecological sustainability to contemporary work on sociotechnical systems sustainability. Literature reviewed in this section comes from the field of science and technology studies (STS), and Computer Supported Cooperative Work (CSCW).

Ribes and Finholt - 2009

Ribes and Finholt studied the design and deployment of "e-infrastructure" by doing a longitudinal cross-case analysis of the daily work practices of "sustaining research facilities" in universities and federally funded research centers (FFRDCs) (2009). They describe the problematic nature of research funding cycles that are misaligned with sustainability work, saying in particular that long-term infrastructure is primarily an institutional consideration, beyond the scope of any single project or discipline. (2009 p. 377) Their findings are broadly construed as implications for the funding of future cyberinfrastructure work, as they describe the various tensions in the deployment of a technology, and its maintenance over a relatively short period of time (2 years).

Overall, this study uses a *resource sufficiency* approach to systems sustainability. The authors describe how resource flows, constrained by financial support from the USA government, impact stocks of knowledge and infrastructural maintenance required for a sustainable research enterprise.

Bietz, Baumer, and Lee

A series of studies by Bietz, Baumer, and Lee explains the "work" of sustaining genomics cyberinfrastructures, which they describe through the concept "synergizing." Sustainable sociotechnical systems in this domain are made possible through the "...work that developers of infrastructure do to build and maintain productive relationships among people, organizations, and technologies. "(2012) Their concern for invisible or overlooked work in genomics database maintenance emphasizes the emergent phenomena of a complex system. As such, they conclude that cyberinfrastructure sustainability is "less about maintaining any particular technology than it is about being prepared to accommodate technological, scientific and organizational change"; accommodating this change is best achieved when maintaining systems level functionality (2012, p. 8). They briefly touch on the flexibility in the rules developed by the genomics community to govern shared resources, but do not tie these to sustainability per se.

Bietz, Baumer, and Lee conceive of sustainability as *functional integrity* of a system, acknowledging that "emergent" phenomena within a work-group are impacted by removing one or another member. The authors stress the need for accommodating change by identifying disruptive forces within a cycle of knowledge production, and suggest that the work of adaptation be labeled "synergy"

Kee and Browning

Kee and Browning focused on the funding infrastructure of developing new, shared high-performance computing centers for scientific applicaitons (2010). They propose to treat these issues as dialectical tensions , with five sets of seemingly opposing forces on three levels of institutions, individuals, and ideologies (2010, p. 285) Kee and Brownings conclusions are that the longevity and sustainability of an infrastructure is compromised when it is based on short-term commitments and part-time attention of a distributed group of technologists, its progress is compromised. (2010 p. 286) Their proposal is conduct further work on the long-term, historical implications of such small funding cycles.

Like Ribes and Finholt, this study takes a *resource sufficiency* approach to sustainability. The author's focus on balancing inputs to a sociotechnical system, such as dollars in grant funding and private sector support, with tangible, measurable outputs. The "dialectical tensions" are positioned as naturally occurring between desires to increase one set of stocks or flows over another.

Vertesi and Dourish

At NASA's Jet Propulsion Laboratory (JPL) Vertesi and Dourish studied two research teams working on the same spacecraft mission. This work produced a cross-case comparison of data sharing and resource pooling. They describe a number of sociotechnical features that complicate the open exchange of data, including an analogy to different types of data having currency within a larger research "data economy" (2011). Their findings emphasize a need to develop systems with a concern for data infrastructures that reflect and secure data sharing practices specific to each scientific collaboration. (2011, p. 9)

Vertesi and Dourish are somewhat unique in this set of studies in that they focus on a specific target for whom a system should be sustainably developed. They are concerned with how policy aimed at increasing data sharing in federal laboratories will impact the sustainability of effective collaborative arrangements at JPL. But, in practice this work resembles a *functional integrity* approach to sustainability by focusing on cycles in which data are produced, and how the practices and policies of managing that data then impact long-term collaborations.

Fry, Schroder and den Besten

In a series of in-depth interviews with principle investigators of eScience projects in the UK, Fry, Schroder and den Bensten investigated "research governance at the institutional level and local research practices at the project level." (2008, p.6) Their work asked research questions about both the sustainability of eScience infrastructure models, and further, to what extent they promoted openness in sharing research products and collaborating. Their findings indicate that, “The fundamental challenge in resolving openness in practice and policy, and thereby moving towards a sustainable infrastructure for e-Science, is the coordination and integration of goals across e-Science efforts, rather than one of resolving IPR (Intellectual Property Rights) issues, which has been the central focus of openness debates thus far”. (2008, p.6) The authors note that they identified a number of conflicts between contractual governance at a micro-level and project settings (institutional contexts) at a macro-level which introduced uncertainty for the sustainability and trust between different collaborators.

Fry, Schroder and den Besten are engaged in a *normative* account of sustainability, as they conclude with a discussion of different approaches to evaluation between participants engaged in science research, and their funding agents; “Such a schism may be because contractual and institutional arrangements in science have traditionally focused on final products (or outcomes) of projects such as publications, compounds or genes, rather than engaging at the level of knowledge creation. This is a particular issue for e-Research as there is more of an impetus to disseminate and share by-products such as software code and data not used in final results. ”(2008, p. 23). They are concerned not just with target communities of sustainability, but with an approach that sees sustainability in terms of justice for future generations of scientists that will need to draw upon a stock of knowledge produced by this work.

Synthesis of Studies

Socioecological studies, even when conceptualizing of sustainability differently, are able to accomplish integration of results for two reasons:

1. They can rely on well-accepted quantitative measures, such as landings yields (fisheries) or wood density (forests), and

2. Over time, the sub-domains that study these different systems have developed ways to conduct meta-analysis by way of systematic frameworks that guide the collection of new data, or re-analysis of existing data.

Using either of these methods of integrating results will be challenging for sociotechnical systems analysis.

In general, there is lack of reliable quantitative measures for sociotechnical systems' performance. None of the studies reviewed above have a quantitative component, and it remains unclear how sustainability might be correlated with metrics, even exploratory metrics (e.g. Bollen et al, 2010), for evaluating performance of these types of systems. Additionaly, the lack of rigor in evaluating these systems from a quantitative standpoint, as I argue in the next chapter, may in fact lead to their instability.

Developing systematic empirical frameworks is a more tenable, immediate step for sociotechnical systems sustainability. I demonstrate in Chapter 4 that socioecological frameworks can be adapted to a sociotechnical settings, and that this can lead to better integration of findings about sustainability.

In final section, I describe research about the governance of systems for peer-production. I argue that this is an important, but underconsidered concept in literature that deals with the governance of contemporary research and development settings.

2.4.1 Peer-production

The decentralized and non-proprietary phenomena of organizing work on the Internet found early success in Free / Libre Open Source Software projects, such as Linux and Apache (Lahkani, 2001). In these settings people who were paid to contribute often worked alongside volunteers, in an almost non-hierarchical structure. Motivation for participating in these types of production systems include; the freedom to reap rewards of one's own work, the prospect of future employment, status within a community, reciprocity for help with similar software projects, social justice activism, and simply wanting interesting problems to solve (Scheick, 2007).

Benkler describes the organization of this type of work as a model of peer-production :

> "a form of open creation and sharing performed by groups online that: set and execute goals in a decentralized manner; harness a diverse range of participant motivations, particularly non-monetary motivations; and separate governance and management

> relations from exclusive forms of property and relational contracts (i.e., projects are governed as open commons or common property regimes and organizational governance utilizes combinations of participatory, meritocratic and charismatic, rather than proprietary or contractual, models)" (2013).

The new mode of organizing that is enabled can be further refined as commons-based peer production, which is

> "...radically decentralized, collaborative, and non-proprietary; based on sharing resources and outputs among widely distributed, loosely connected individuals who cooperate with each other without relying on either market signals or managerial commands." (2006, p. 60)

While this early definition characterized participation in peer-production as loosely connected Haythornthwaite later offered two models of peer-production on the Internet - Lightweight and Heavyweight:

- The lightweight model of peer-production is recognizable in the open-source software projects mentioned above (e.g Apache, Mozilla, or even Linux). In a lightweight model of peer production many loosely connected individuals contribute effort to accomplishing small well-defined tasks, that when combined add up tp complex, large technologies. The strength of a lightweight model of peer-production is that participants can self-select tasks, and maximize their own skills in contributing while simultaneously learning new skills in collaborating.

- The heavyweight model of peer-production is exemplified by virtual organizations (VOs) made up of strongly connected and highly committed members whose tasks are loosely coordinated, and their contributions are accepted based on quality control mechanisms like peer review (2009). Heavyweight peer production often includes the development of a high-quality dataset, or a piece of software with a very specific use.

Research on the process of collective action in peer production systems (i.e. Wikipedia, reddit, slashdot, etc.) demonstrate that a mix of both experts and non-experts to coordinate, and cooperatively produce complex works that are (often) of superior quality than their market-based competitors (e.g. Wikipedia vs.Encarta). These systems are often described, in general terms, as having an egalitarian nature of "communal information goods" (Fulk, Flanagin, Kalman,Monge, Ryan, 1996; as quoted in Shaw and Hill, 2014). Some of the peer-production systems listed above have created democratic organizations that have had transformative and lasting success in industries that are traditionally dominated by private firms (i.e. Apache Server software, Linux and Android operating system, etc.) But, just as often FLOSS projects are the work of a single individual (Ingo, 2006) and by most measures highly uncooperative (Hill, Burger, Jesse, Bacon, 2008;); Wikipedia is a single success in what were many similar wiki-based encyclopedia projects that failed to attract participants (Kittur and Krut, 2008; Ortega, 2009); and wikis - lauded for being the most democratic of peer production systems tend to resemble oligarchies when studied over time (Hill and Shaw, 2013). Most recently, commons and peer production scholars have started to turn their attention from what has made the Wikipedias of peer production successful towards what has caused similarly structured projects to fail (Shaw et al; 2014).

All of this is to say that while peer-production is a novel organization form, it does not necessarily lead to successful or even democratic divisions of labor. Peer production satisfies costs and benefits of a networked communication structure just as any other method of production, but it can outperform traditional models of organizing labor given the right context. These contexts remain under-theorized, and the effects of governance on peer-production remain an understudied phenomenon in the current literature of science and technology studies. Where research on open-source software projects suggests that the introduction of peering is reflected in their governance structures (O'mahony and Ferraro, 2008), there have been few similar studies conducted on the impact of peering in research and development settings.

2.5 Summary

In the following Chapters, I will demonstrate how new modes of production (including peer-production) have been governed within the ICOADS community. I will attempt to show how, in combination, changes in provisioning and producing shared resources introduces social dilemmas for a sociotechnical system, like ICOADS, to solve through commons governance. And, I will demonstrate how adapting systematic approaches from the study of commons governance in socioecological systems can help reduce the complexity of understanding these issues as they relate to sociotechnical systems.

In this chapter I have:

- Reviewed the classical economic approach to production and provisioning of goods; including the market and state models.
- Described how, initially, Coase and Williamson studied the role of governance in organizing at an institutional level.
- Discussed the missteps of the "Tragedy of the Commons" logic, and how a classification of goods based on two dimensions - rivalry and excludability - helps make sense of the types of externality dilemmas that Hardin had described.
- I reviewed a number of different findings from Ostrom's work on the successful design and arrangement of institutions for collective action, including a set of rules and design principles that have proved to have high explanatory power for long-enduring socioecological systems.
- I then reviewed conceptions of sustainability from the sociecological systems literature, and attempted to show how these are manifested in contemporary science and technology studies. I used this exercise to argue that a systematic approach to studying these phenomena was necessary to build knowledge about which arrangements are successful for sustaining sociotechnical systems.
- I introduced the concept of peer-production, and delimited its scope so as to understand that it is one of several models of explanation for the division of labor in contemporary research and development settings.

CHAPTER 3

3.1 Introduction

This chapter describes in detail the range of empirical research methods used to carry out this project. This includes the design of a case study, as well as the various methods used for data collection and analysis

I begin by describing the case study design, including the overall structure of the case and its background. I then describe the methods of data collection used in this work, including a set of ethnographic studies carried out over a three year period, three informetric studies that provide a quantitative account of ICOADS use and acknowledgment, and a set of semi-structured interviews conducted in year three of the study. I conclude by describing the use of two existing frameworks to organize and analyze these different data sources, and the validity constructs that have guided this analysis.

3.2 Case Study Design

The marked features of a case study are that it:

- Investigates a contemporary phenomenon within its real-life context
- Is appropriate when the boundaries between phenomenon and context are not clearly evident
- Copes with the technically distinctive situation in which there will be many more variables of interest than data points
- Relies on multiple sources of evidence, with data needing to converge in a triangulating fashion

- Benefits from the prior development of theoretical propositions to guide data collection and analysis (Quoted from **?**).

In social science research that takes things like organizations, institutions or systems as a unit of analysis, the case study approach can be used to test, extend, or further develop existing theories, or empirical frameworks (**?**). This case study will adapt and extent existing socioecological frameworks for studying sustainability issues in sociotechnical systems settings.

The research design of a case study requires the following elements:

1. Clear definition of the case type
2. Research Questions
3. Propositions
4. Unit(s) of Analysis
5. Logic linking the data to the propositions
6. Criteria for interpreting the findings (Yin, 2014).

Below, I describe the design choices made for the ICOADS case study.

Research Design for Case Study of ICOADS

3.2.1 Type of Case Study

Distinctions in case study types are generally made along categorical lines such as exploratory, descriptive, explanatory, or critical (**?**). The following case study of ICOADS is most comparable to the descritptive approach, as it attempts to describe how and why a certain institutional arrangement has been successful at sustaining a shared resource set over a period of time.

3.2.2 Research Questions

The specific research questions that this case study will answer are:

- **CS RQ 1** : What are the different governance models that ICOADS has used to manage shared resources over its thirty year existence (1983-present)?

  To reiterate from Chapter 1, I rely on the definition of a governance system as "a set of institutional arrangements (such as rules, policies, and governance activities) that are used by one or more actor groups to interact with and govern interactions." (Cox et al., 2014).

- **CS RQ 2** What causes a governance system to shift from one regime to another?

  A regime shift is, following Smith, Stirling, and Berkhout (2005), understood to be a function of two processes:

  1. Shifting selection pressures (either external or internal pressure to change) bearing on the regime.
  2. The coordination of resources available inside and outside the regime to adapt to these pressures.

- **CS RQ 3** What types of disturbances are ICOADS resilient or vulnerable to?

In order to identify selection pressures, governance shifts, and disturbance types, I use a systematic approach to code system state variables related to ICOADS different governance models. This approach is described in the "Criteria for Interpreting Results of Case Study" section below.

The broader research questions being pursued in this dissertation are:

- What are the effective institutional arrangements (governance) for sustainable community based science infrastructures?

- How and why do they differ between domains of knowledge production?

These two research questions are addressed by synthesizing the results of this case study with those from previous cases using the same analytical framework. I describe this synthesis work in detail at the end of this chapter.

3.2.3 Unit(s) of Analysis

In short, a unit of analysis is the subject of a case study, and units of observation are individual measures of the subject (Long, 2004.

The unit of analysis in this case study is ICOADS as a socio-technical system. This is a single case study, but inter-case comparison will be conducted through exploration of different governance regimes adopted over the history of ICOADS. Smith, Stirling, and Berkhout provide a nice description of the need to expand a unit of analysis in this type of work,

> "Conventional economic analysis of technical change tends to focus on pressures that operate visibly at the level of the firm (such as pricing, competition, contracts, taxes and charges, regulations, standards, liability, profitability, skills and knowledge). Analysis at the level of the socio-technical regime, on the other hand, includes such factors, but goes beyond them to consider less economically visible pressures emanating from institutional structures and conventions, including changes in broad political economic landscapes, or wider socio-cultural attitudes and trend"(Smith, Stirling Berkhout, 2005).

In this case study, there are multiple units of observation, including, users, producers, provisioners, and stakeholders of the ICOADS project, as well as the digital resources themselves (e.g. ICOADS software and data)

3.3 Data Collection Methods

A number of different methods were used to collect data for this case study. With the exception of the semi-structured interviews completed in year three each of these studies have been published in peer-reviewed outlets. Where this is the case, I provide bibliographic references for that work. Below I describe these methods in detail. Each of the sections that follow are structured to include the following:

- A description of the intention of research method used (what particular phenomenon it was useful for understanding),

- A brief review of the background literature that informed my operationalization of this approach, and
- Description of the data (range, type, etc.) that were produced by an individual study.

3.3.1 Preparation for Fieldwork

Collecting data of any kind requires a certain amount of familiarity with the topic, and subject that will be studied. This is an especially important and diverse topic for contemporary science studies where interdisciplinary research is conducted (Jirotka, Lee, and Olson, 2013). In preparation for doing fieldwork amongst experts in Marine Climatology, and for further enchaining my understanding of this field, I undertook following activities:

- In the Fall of 2012, and 2014 I enrolled in ATMS 591 Atmospheric Sciences Seminar at the University of Illinois, Urbana-Champaign. This is an open discussion and lecture series featuring topics of importance to the field of meteorology, atmospheric science, and earth systems science more generally.
- I also completed four on-line (MOOC) courses with passing marks in the following related subjects:
  - "Marine Megafauna: An Introduction to Marine Science and Conservation" (Duke University)
  - "Climate Literacy: Navigating Climate Change Conversations" (University of British Columbia)
  - "Global Warming: The Science and Modeling of Climate Change" (The University of Chicago)
  - "Climate Change" (The University of Melbourne)
- During both residencies at NCAR (described below) I attended weekly lectures at NCAR, including the Advanced Studies Program lecture series. This gave me the opportunity to both interact with visiting scientists, and describe the research agenda that I was involved in, further adding to my own competency in the science behind my participant's work.

3.4 Ethnographic Study of ICOADS

Two forms of ethnographic inquiry informed this project: In June of 2012 I began by using an interactional approach borrowed from the field of Science and Technology Studies (STS), and over time I also adopted a complimentary 'historical ethnographic' approach as it is conducted in the field of Organizational Behavior. I describe each approach in detail below.

Participant Observation: Interactional Expertise

Methodological Intention

Participant observation allows for ethnographic data to be collected in a natural context, adding a dimension of inter-personal authenticity to what are the otherwise sterile or staged interactions of a formal interview process. My intention in using this method of data collection was to capture the everyday lived experiences of developers, scientists, and engineers engaged in the ICOADS community. As these individuals are members of diverse academic, social, and cultural backgrounds they have a very different view of the phenomena of governance. In observing, participating in, and often arguing about climate science I was able to better understand and interpret behaviors within the community that I was engaged in studying.

The traditional focus of this type of ethnography in on "unobtrusive observation" - that is, entering, observing, and leaving a site with minimal impact on the everyday lives of the participants (Fine, 1993). Some of the earliest studies of laboratory cultures in STS contested the value of this strict form of participant observation; as Woolgar and Latour argued, the need to be inconspicuous and innocuous was detrimental to their participants own understanding and reflection on basic, routinized steps in a research process (Woolgar and Latour, 1979 p. 40). My intention in using an ethnographic approach to data collection is

Background Literature

Over a three year period (June 2012 April 2015) I engaged the ICOADS community - including developers, funders, end-users, collaborators and previous project participants - at multiple sites of study, including a significant amount of time (15 months total) as a research fellow at the National Center for Atmospheric Research (NCAR) in Boulder, Colorado. My collection of data in this setting follows what Harry Collins calls participant comprehension. This mode of inquiry is informed by traditional ethnomethodology as found in cultural anthropology (e.g. Garfinkel, 196X) and science studies (e.g. Lynch, 1997). Harry Collins' work on participant comprehension began in 1984, when studying the material culture of gravitational wave physics he described "an interpretation of participant observation under which the field-worker tries to acquire as high a degree of native competence as possible and interaction is maximized without worrying about disturbing the field site." (1998, p. 297).

Collins' motivation for developing the extended form of description that participant comprehension generates is in pursuit of an interactional expertise, or a third kind of knowledge that sits between formal or propositional knowledge and informal or tacit knowledge (2004, p. 125-7). Collins' argument is that during his time studying gravitational wave physics he was able to achieve an expertise which was something different than the kind that comes with formally studying and being a part of a discipline as a practitioner. Through linguistics socialization, traditional ethnographic observation and extended field work, Collins could read, participate in discussions, and defend his arguments about physics, without having the ability to actually do the physics himself.

My time working in this field, as well as studying and using ICOADS has allowed me to develop something like an interactional expert I understand the limitations and uses of the data in ways that allow me to converse with its community of users. Six months into this project, I was elected to the American Meteorological Society's Board of Data Stewardship.After a year and a half of interaction, I was even been invited to give a talk at a biennial marine climatology meeting (Weber and Worley, 2014). Neither of these accolades mean that I am able to do marine climatology research, but they demonstrate that I can developed a level of mutual respect within this community, and was able to eventually access, manipulate, and contribute to data and curation in my performing an ethnographic mode of inquiry.

Operationalized Study

Like Collins, my interaction is maximized by informally interviewing, talking with, and observing people who produce and use ICOADS; The operationalized study included the following periods of data collection:

- From June - August of 2012 I was an Advanced Studies Program (ASP) fellow at the National Center for Atmospehric Research (NCAR) in Boulder, Colorado. While there, I observed and interacted with a team of four individuals that curate ICOADS at NCAR. I also informally interviewed scientists who used ICOADS, asking basic questions about their experience, and their interaction with the project over time. During this time, I also established a formal relationship with one of ICOADS original developers who was willing to become my "key informant" (Payne, 2004). This individual sponsored my project within the ICOADS community, and became a mentor of sorts to me in the field of marine climatology, and data science more generally.

- After leaving the site, I continued to work with this team of four curators, exchanging emails, and calling into teleconferences about their work on ICOADS.

- In August, 2013 I returned to NCAR to collaborate with this team of four individuals, and a larger network of visiting scientists and ICOADS partners in the Boulder, CO area.

- In August 2014 I left NCAR, but continued to interact with broader network, including regular participation in teleconference calls with my main informant and a steering committee established to govern ICOADS.

- During this time I also attended the following meetings where I either presented my on-going informetric work, or observed my participants presenting their own work: 2013 American Geophysical Union (AGU) Winter Meeting; 2011-2014 American Meteorological Society (AMS) Annual Meetings; 2012-2015 Earth Science Informatics Partners (ESIP) Winter and Summer meetings; the 2014 Joint Technical Commission for Oceanography and Marine Meteorology (JCOMM) Workshop on Advances in Marine Climatology (CLIMAR-IV).

Data Collected

Forms of data collected in the ethnographic work includes field and interview notes, memos (which summarized my observations, and were later refined and shared with my participants for validation (Emerson, 2011); I subscribed to a current email list and participated in conversations there; I gathered general documentation about the production of ICOADS itself-meaning user manuals, production notes for programmers working on transformation of different versions of ICOADS, as well as formally published proceedings of meetings, meeting minutes, and journal articles describing the project dating from as early as 1981.

I made a conscious decision not to record interviews conducted during this stage of research. I came to this decision after initially attempting to record conversations and observing discomfort of my participants in having their conversations taped. I made every effort to announce, and clearly communicate that my participation was part of a dissertation focusing on ICOADS governance. I also sent an email to project partners in ICOADS outlining the major goals of my dissertation, which they approved and sent back their encouragement.

3.4.1 Document Analysis: Historical Ethnography

Methodological Intention

In order to meaningfully answer research questions about the historical evolution of ICOADS' governance regimes, I also used an approach to gathering historical data from the field of Organizational Behavior. Historical Ethnography is a technique pioneered by Diane Vaughn in her work with NASA and a USA Presidential Panel in the aftermath of the *Challenger disaster* (Vaughn, 2006; 2009). The approach requires balancing archival work with present-day validation of participants' memory of those past events. It can be described loosely a a process of "digging into the past, deliberately reconstructing history in order to identify and then track the processes connecting past and present." (1999) What makes my work a historical ethnography, as opposed to pure history, is that it includes a sensitivity to present day events and a deep commitment to tying those events to the history out of which they were borne. This is the major source of data for the "Background Environment" section of Chapter 4.

Background Literature

Vaughn's work with NASA was based on what she called "documentary practices" within NASA as an organization (Vaughn, 2006). Trying to better understand on-going Congressional testimony, she sought an explanation of how deviance - through preference for risk, and safety standards - diverged between engineering and management cultures. The breakdown of communication between the two groups was traced through reports filed by engineers in previous shuttle missions, as well as *Challenger* flight simulations. By connecting these reports, their language, and the documentary trail of decision making processes that informed a NASA management decisions, she could ultimately provide a more thorough explanation of the disaster based on organizational breakdowns in safety reportage.

Another germane historical ethnography can be found in John Middleton's study of early Modern Era Swahili merchants. Middleton drew upon Vaughn's method of historical ethnography in order to show how certain artifacts (coins, dock locations, merchant societies, etc.) have persisted in structural shape throughout Eastern African. This continuity created a porous boundary between present and past in divisions of class, race, and gender. Middleton's research, although based on historical documents, was grounded by 'collective memory' of cultures he observed and lived within, and by 'extrapolation from modern ethnography' in current mercantile society's of East Africa (Middleton, 2013).

Operationalized Study & Data Collected

I conducted archival work in two physical locations: the National Center for Atmospheric Research archives, as well as NOAA's Earth System Research Laboratory (ESRL) which has a small but useful set of papers related to the COADS project; both are located in Boulder, Colorado. I also asked each participant that I interacted with to provide me with email archives, past meeting minutes, and any documentation related to previous ICOADS development. NOAA has a number of digitized historical texts available online at (http://www.history.noaa.gov/index.html), as well as links out to historical material that covers marine climatology and its history. JCOMM and NOAA digitally archive the proceedings of two major conferences related to marine climatology (CLIMAR and MARCDAT) dating back to the first meeting in 1999. Through my key informant, I was also able to obtain an archived discussion forum that recorded early emails exchanged by ICOADS developers.

In total, I gathered over 10 GB of digital material - of which I read through, created memos, and used for discussion with my informants. This was also an iterative process of tracing citations, and trying to establish the provenance of certain ideas or even ideologies. This process eventually led me to Matthew Fontaine Muary, a USA Naval officer who played a key role in both naval history, and weather data standardization. Maury's personal papers, as well as valuable secondary sources were digitized and are available online through the Virginia Military Institutes's University Archives[1].

3.5 Informetric Studies

One of the social dilemmas that we will see ICOADS face is related to evaluation in that - like an infrastructure - the diversity of metrics defining success of the project is one of its chief hurdles in securing reliable funding. The informetric studies conducted in this dissertation are a heuristic attempt at gathering "demand-side" criteria for the provisioning of ICOADS. As opposed to supply side criteria, such as the amount of funding ICOADS has received, the number of records in its archive, the number of papers it has produced, etc, demand side informetric work tries to determine, who uses, what, when, and why?

I stress that these are imperfect studies in the traditional Information Science sense of having robust statistical measures of significance, however they are another data source that provides a means for triangulation on what are a broader set of research questions about the sustainability of a shared resource. I return to the discussion of metrics in Chapter 5.

3.5.1 Citation Content Analysis

Methodological Intention

The premise of this study is to use close reading techniques to better understand the context in which ICOADS has been referred to in the formal literature of marine climatology (1985-2014). As a result of failed experiments in calculating traditional forms of bibliometric impact for the ICOADS project, this study was designed to give a more thorough account of not only how frequently something is cited, but for which reasons, and with what sentiment. This approach may be especially salient for the earth systems

[1] http://www.vmi.edu/archives.aspx?id=19209fulltext

science community, because of the tradition of writing data "papers" which announce or describe a new data-related product (Mayernik et al., 2014). These publications are seen as a way for data producers to announce an update (newly digitized material, newly processed or interpolated data points, new grid resolutions etc.), or to comprehensively describe the process of aggregating and developing a new archive of software and data. The context in which citations are made to papers about ICOADS can offer a way to deepen an understanding of how the archive has been used over time, and how users have judged its quality, reliability, utility and value.

A "data paper" may be cited in a number of ways - to document the use of a certain data source (example 1 below), or it may be cited for the propositional content that the data paper contains - such as assertions it makes about the coverage or completeness of the data being released (example 2, below). All of this is to recognize that journal publications are complex documents that play a number of different and sometimes conflicting roles.

- **Citation 1**

  The error properties of both input data in OA procedure are obtained using Comprehensive OceanAtmosphere Dataset (COADS) (Woodruff et al. 1998) ship observations as the base data. (Wheeler et al, 2008)

- **Citation 2**

  Thus, the B150% CFC supersaturation levels observed in the summer of 1988 are probably the result of seasonal warming combined with the substantially lower mean wind speeds in spring and summer (Woodruff et al., 1998) allowing only partial reequilibration of the upper water during the warming. (Lee et al. 2002)

Background Literature

Similar informetric work has focused on developing a theory of citation (Cronin, 1984; and Leydersdorff, 1984), or typing citations in a formulaic way so as to describe their function in a scholarly communication system (White, 2004). Attempts to type citations based on their function within a formal publication dates back as far as the 1970's with Small's work in citation context (1982) and symbol (1978), Pertiz's study of citation roles (1983), and Moravcsik and Murugesan's typology of citation functions (1975). These early studies were conducted in the traditions of library and information science, systems development and analysis, and applied linguistics; as such the classifications of these citation elements were aimed at improving the indexing of a document or information retrieval system. In short, previous works that analyzed the context and content of citations were, as Zhang, Milosevic and Ding put it "...constructed more from the perspective of users needs and perceptions rather than from those of the citing authors, especially in terms of authors citing motivations. "(2013)

Recent work on extracting and coding the context of citations has focused on sentence level content analysis to improve abstracting services (Nakov, Schwartz, and Hearst, 2013), as well as the semantic and syntactic coding of citation contexts to create a comprehensive schema of citation behavior (Zhang, Milosavic and Ding, 2013; hereafter referred to as the CCA approach). Both of these studies use methods of qualitative inquiry to mark-up a text manually, and then use these annotations to improve machine-learning algorithms that automate the classification of citations.

Work at manually coding citation contexts had previously been popular in fields like organizational theory (Anderson, 2006; Golden-Biddle et al., 2006; Lounsbury and Carberry, 2005; Mizruchi and Fein, 1999) but there has yet to be a substantive use of this work to automate or improve manual processes.

There have also been a number of citation typing, and scholarly communications ontologies developed for semantically encoding digital publications. The most successful and fully developed of these are the SPAR (semantic publishing and referencing) suite of ontologies. Included in SPAR are the Citation Typing Ontology (CiTO) (Shotton, 2010) and the Scholarly Contributions and Roles Ontology (SCoRO) (Shotton and Peroni, 2013) which provide an exhaustive vocabulary for typing the content of a citation, and the roles played in producing a publication, respectively.

Operationalized Study

There are four major publications that have announced a new release, or a substantive update to ICOADS.

1. Woodruff, S.D., R.J. Slutz, R.L. Jenne, and P.M. Steurer, 1987: A comprehensive ocean-atmosphere data set. Bulletin of the American

Meteorology Society 68, p. 1239-1250.

2. Woodruff, S.D., H.F. Diaz, J.D. Elms, and S.J. Worley, 1998: COADS Release 2 data and metadata enhancements for improvements of marine surface flux fields. Physical Chemistry of the Earth, 23, p. 517-526.

3. Worley, S.J., S.D. Woodruff, R.W. Reynolds, S.J. Lubker, and N. Lott, 2005: ICOADS Release 2.1 data and products. International Journal of Climatology (CLIMAR-II Special Issue), 25, 823-842. doi:10.1002/joc.1166

4. Woodruff, S.D., S.J. Worley, S.J. Lubker, Z. Ji, J.E. Freeman, D.I. Berry, P. Brohan, E.C. Kent, R.W. Reynolds, S.R. Smith, and C. Wilkinson, ()2011) ICOADS Release 2.5: Extensions and enhancements to the surface marine meteorological archive. International Journal Climatology (CLIMAR-III Special Issue), 31, 951-967. doi: 10.1002/joc.2106

Using the bibliographic databases Scopus and Web-of-Knowledge (WoK), I retrieved all citations made to these publications as of August 1, 2013, and then removed duplicate or overlapping citations. The result was "citing documents corpus" that contained 1,195 documents. I then used a systematic sampling technique to select documents evenly across the 30 years span of ICOADS publications:

	Citations received	**Lapse**	**Documents sampled per year**	**Study sample**
2011	57	2	7	15
2005	190	8	6	48
1998	162	15	3	41
1988	786	25	4	128
				252

Data Collected

Following the practice of Nabokov et al. (2004; 2013) a colleague and I then extracted the sentence in which each ICOADS document was formally cited, and instances where the project was mentioned by name (either formal or by acronym) but was not cited.

We then coded the sentences in which ICOADS was cited or mentioned using the following categories:

- Number of authors

- Relation to cited work: Is the citing publication related to the ICOADS project in some way?
- Location of citations and mentions : Where in the publication is ICOADS cited or mentioned?
- How ICOADS was referenced: Direct citation, endnote or mention
- How ICOADS was referred to, materially : What do the authors call ICOADS? (i.e. a dataset, an archive, etc.)
- Function of the citation: Meant to answer why was ICOADS being cited (Uses CiTO controlled vocabulary)
- Sentiment of the citation (and mentions) : A citation or mention could be positive, negative, or neutral

We chose a small sample (n=10) to initially code together, and iterated on this process for three rounds (n=40 total) until we achieved a consistent inter-rater reliability. We coded all 252 documents in the sample. The result is a set of descriptive statistics for the categories that we coded. The full coding schema and access to bibliographic records can be found in the project repository[2].

Data Usage Index

Methodological Intention

Federally funded research and development centers typically measure the levels of service they provide to end users through descriptive statistics, such as the number of times a data file has been downloaded over a given period of time. To better understand the use of ICOADS in the climate science community I designed a study that drew upon user-log analysis techniques from Information Science (**??**), and in particular a Data Usage Index (hereafter referred to as DUI) from biodiversity informatics.

Background Literature

The DUI was originally designed to measure the impact of institutional contributions to the Global Biodiversity Information Federation Database

[2]http://git.io/vJnPX

(GBIF) by using *indicators* of how data were discovered and accessed (i.e. the number of downloads, page hits, files contributed by an institution, etc.) (Chavan and Ingwersen, 2011). In the DUI, indicators are extracted from user query logs and combined in simple, but unique ways to calculate impact factors[3].

Operationalized Study

This intention of this study was to modify the DUI's indicators from the context of a biodiversity database to the Research Data Archive (RDA) at NCAR, which provisions a number of different climate data and software resources (including ICOADS). To do this we developed a set of use cases based on a researcher's discovery, selection, and mode of access to different data products from the RDA.

From the use cases we identified six key indicators that were captured in the RDA's logs, as well as potentially informative relationships between these indicators (see Table 1 below).

Code	Indicator	Explanation
uu(ds)	Unique Users	Unique users that downloaded data during a time window
n(ds)	Number of Datasets	Number of Datasets assigned DS number by RDA
f(ds)	Files DS	Number of files in Dataset per time window
d(ds)	Download Frequency	Total number of files downloaded per time window
hp(ds)	Homepage Hits	Home Page Hits of Data Set per time window
d(ds) /uu (ds)	Download Density	Average number of files downloaded per unique user
d(ds) / f (ds)	Usage Impact	Total number of downloaded files over total files in dataset
d(ds) / hp(ds)	Usage Balance	Files downloaded by number of homepage hits per time window
hp(ds) / f(ds)	Interest impact	Total homepage hits per number of files in dataset
hp(ds) / uu(ds)	Secondary Interest Impact	Total Homepage hits over unique users
ss(ds) / d(ds)	Subset Ratio	Subset requests over total number files downloaded

The completed use cases also demonstrated that two user types could be identified based on how data were accessed:

- **Programmatic Users**: accessed or downloaded data through a command line tool (e.g. '-curl [ˆcurl] or "wget" [ˆwget]) or through a scripting language.
- **Assisted Users:** access data via the graphical user interface, or by subset requests made through a separate tool developed by the RDA

[3]For more detail, see repository with tables and data http://git.io/vJnPX

staff.

Data Collected

To test the effectiveness of the modified DUI three RDA data products were selected - including the most recent release of ICOADS (version 2.5). Indicators from the completed use-cases were then extracted from the user logs of each dataset over a sixteen month period.

From the index of each dataset we then further calculated two impact factors:

Usage Impact Factor

$$(d(u)/f(u))/\sum_{1}^{n} d/\sum_{1}^{n} f \tag{3.1}$$

where (u) is the given resource unit, (d) is the download frequency of users, (f) is the number of files downloaded per user session, and (n) is the total number of units (files available in the dataset) in the denominator.

Interest Impact Factor

$$(hp(u)/f(u))/\sum_{1}^{n} hp/\sum_{1}^{n} f \tag{3.2}$$

which is identical to UIF except download frequency (d) of users is replaced by the number of homepage hits (hp) a dataset receives.

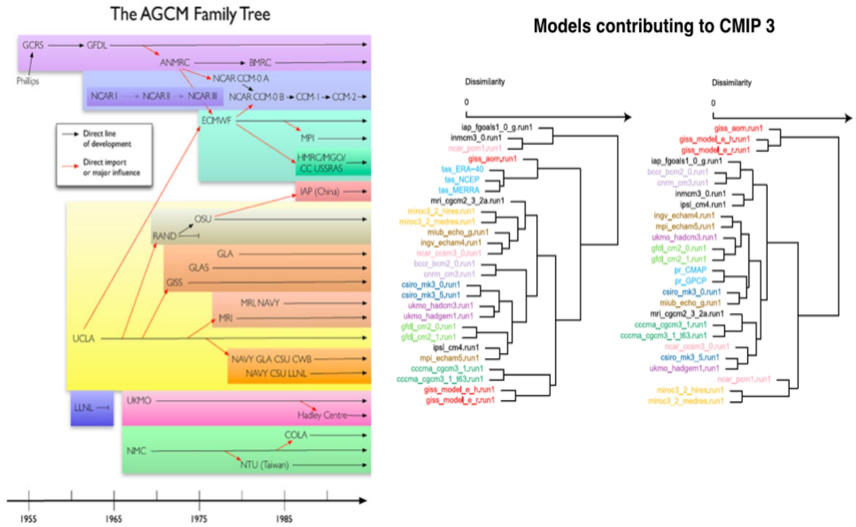
This resulted in a composite index for each of the three datasets[4].

Phylomemetics

Methodological Intention

In my ethnographic work participants often explained that ICOADS was reused in developing new climatology data products. However, operationalizing a tracking of the reuse of ICOADS through citation analysis proved

[4]Found in http://git.io/vJnPX



Edwards' genealogy and Knutti's principle components

limited by the inconsistency in how the project is acknowledged, and which research object is subsequently cited (see Weber, Mayernik, and Worley, 2014 for more details on this dilemma).

To better understand the history of how ICOADS has been reused, we conducted a study using techniques from the field of evolutionary biology (Page, 2009) and cultural analytics (Mace, 2005). This work produced a *genealogy* of ICOADS.

Background Literature

The inspiration for this work came from a set of studies that looked at the history of climate models by tracing their relatedness (See Figure 1.1); The figure on the left is from a study conducted by Edwards, who created a "family tree" of Atmospheric General Circulation Models (AGCM) from archival documents and oral histories (**?**). The figure on right is a dendrogram showing relatedness of climate models that participated in the 5th Climate Model Intercomparrison Project (CMIP), constructed using principal component analysis on documentation available for each model (**?**).

Quantitative phylogeny of digital artifacts

The approach is not without precedence: application of quantitative phylogenetic methods to linguistics and textual criticism is almost as old as phylogenetic methods themselves. In fact, Platnick and Cameron argue that similar methods were accepted as standard practice in both fields before biologists came to embrace them (1977). There has also been a recent resurgence of interest in phylogenetic approaches to non-biological problems partly due to computational advances in bioinformatics, which not only allow for faster and easier computation, but also support the use of molecular clocks to root known speciation times (sometimes called divergence points) in ways that were previously difficult or impossible (Mace and Holden, 2005; Mace, Holden, and Shennan 2005).

To further emphasize the shift between biotic and abiotic studies of evolution, Howe and Windram coin the phrase phylomemetics in lieu of phylogenetics, given the use of the word meme to refer to a non-genetic principle that behaves in a genetic way (2011). Though the differences between memetic and genetic evolution may have bearing on the models and algorithms used to study these processes, in this work, we use methods and software developed explicitly for phylogenetic work, and refer to our study as such. Previous work in linguistics and textual criticism also borrows heavily from biogeography in coupling an analysis of linguistic divergence how dialects differ from one generation to the next with analysis of human migration routes (Rexov, Frynta, Zrzavy, 2003). Similarly, phylogenies of historical texts have been constructed for literary works such as Chaucers Canterbury Tales (Barbook, 1998) and Little Red Riding Hood (Tehrani, 2013). These approaches typically focus on finding divergence points to estimate when texts were altered, replicated or significantly changed by different authors or cultural groups.

Most immediately applicable to this study are phylogenetic applications by archaeologists and anthropologists who conceptualize artifacts as, "complex systems comprising any number of parts that act in concert to produce a functional unit," in which the "changes that occur over generations are highly constrained, meaning that new structures and functions almost always arise through modification of existing structures and functions as opposed to arising de novo "(OBrien, Lymen Darwent, 2002). This system view of artifacts is particularly applicable to digital objects, which may also be viewed as complex systems comprising any number of interactions between layers of information content and representation (Wickett et al., 2012). Bit sequences, encoded information content and information systems work together to produce a functional unit, and the changes that occur over generations of use are constrained by the practices and sociotechnical contexts of the groups using them.

Qualitative phylogeny & the biography of artifacts

Just as quantitative phylogenetics has a long history of application to the study of material and textual artifacts, so does the qualitative study of evolution as cultural diffusion. Anthropologists, economists and sociologists have each noted the importance of tracking the social "markings" of mundane objects that personalize, and make a given object individual to a period of time (Appadurai, 1986). In this vein, Igor Kopytoff proposed that tracking the movement of an artifact between different contexts of use required a biographical approach that could see "a culturally constructed entity, endowed with culturally specific meanings, and classified and reclassified into culturally constituted categories "(1986, p. 68).

Similarly, Williams and Pollock describe a technique called the biography of artifacts, which takes a popular software platform as a unit of analysis (e.g. Microsoft Sharepoint), and attempts to trace the way it was modified, changed, and socially shaped by studying the different contexts in which it was used. The ambition of the biography of artifacts approach is to show the evolution of similar technical artifacts in different social contexts, including their adaptability (or evolutionary fitness) across diverse software ecosystems (Williams and Pollock, 2009). Dosi and Nelson similarly relate evolutionary concepts from biology to behavioral economics and organizational theory (2003). In doing so, they relate technological change within private firms to environmental pressures in an ecology, effectively equating these externalities as selection mechanisms for evolutionary processes. Dosi and Nelson attempt to study links between organizational economics and evolutionary biology through qualitative observations of the practices, policies and technological adaptations of a firm.

A quantitative phylogenetic approach can add another dimension to each of these types of analysis. Though it cannot answer the same types of questions about how context or culture has shaped technical artifacts as used in different social settings, it can more rigidly answer when and to what extent an artifact has changed between cultures, and visualize those changes over time.

Operationalized Study

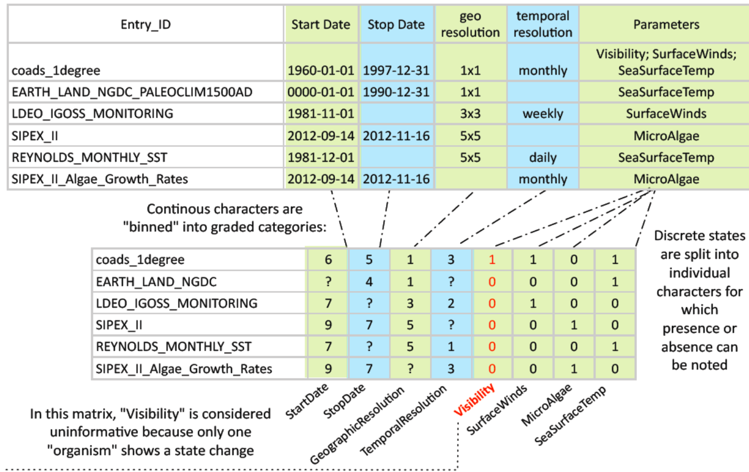
Our work on an ICOADS genealogy somewhat diverges from these previous studies in that we aren't making an evolutionary metaphor or analogy, we are directly borrowing techniques and software from phylogenetics (Page, 2009) in trying to quantify and subsequently visualize the relatedness of climate datasets. Part of the ambition in taking a biological approach to studying material aspects of ICOADS was to leverage the explanations and theories that this field offers for cooperation (Novak, 2006). If ICOADS data products did have a traceable evolution then it might be possible to use concepts like kin selection, fitness, group selection, and reciprocity (direct, indirect and networked) to shed light on how data have been reused, shared, developed, or refined. We were not trying to create a direct mapping between biotic reproduction and abiotic data reuse, but we were taking seriously the ecological metaphor that is often invoked in discussing the complexity of software / data intensive enterprises (Weber, Thomer, and Twidale, 2013).

Data Collected

To trace the genealogy of ICOADS, we searched for and harvested metadata records from NASA's Global Change Master Directory (GCMD) using the ISO 19115 standard. We used queries related to ICOADS such as "International Comprehensive Ocean and Atmosphere Dataset" or "ICOADS" and "COADS" to locate as many related ICOADS records as possible. In total, we discovered 99 records, of which only 23 represented different versions or subsets of the ICOADS project. The remaining (n =76) records represented derivatives or offspring of ICOADS.

Next, we identified properties (or characteristics) of climate datasets which we thought would be unique and important to signifying change from one generation of ICOADS data to the next. This included things like format, encoding characteristics, or even the parameters of the data set (e.g. Sea-Surface Temperature, Sea-Level Pressure, Wind Direction, Swell Height, etc.). We then extracted fields containing these properties from the metadata records. The fields harvested included: Entry Title, Entry ID, Summary, Geographic Coverage, Start Date, End Date, Geographic Resolution, Temporal Resolution, Scientific Keywords (often dataset parameters), Geographic Keywords, Sources (platform of data collection), and Instruments.

Next, we converted each field into binary codes for "presence" or "absence" of individual keywords (Figure 1.2). In some cases we coded additional

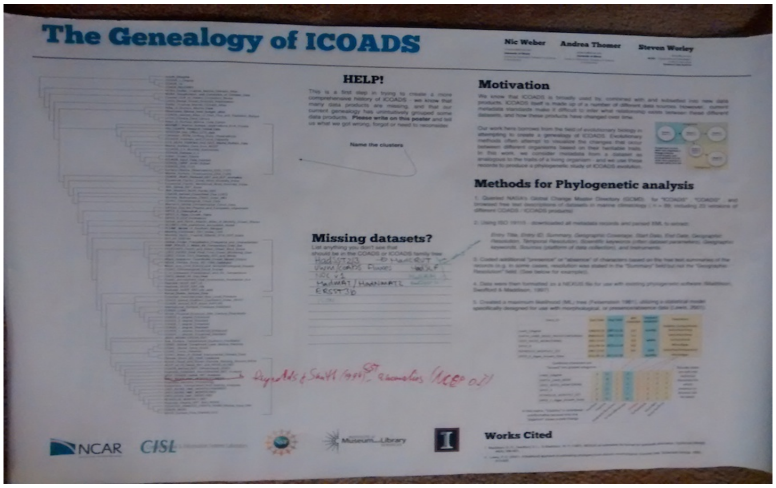


"presence" or "absence" of characters based on the free text summaries of the records (for instance, in some cases, resolution was stated in the free text "Summary" field but not the "Geographic Resolution" field).

Outcomes

With the coded data, we then produced a maximum likelihood (ML) tree (Felsenstein, 1981), by utilizing a statistical model specifically designed for use with morphological, or presence/absence data Ȧgain, the assumption that we were making in this process is that significant properties or characteristics of the metadata records could be clustered - such that data products that shared similar traits could be grouped together in the same way that similar species are grouped together in a phylogenetic analysis (for a complete discussion of this work see (Thomer and Weber, 2014; Weber, Thomer, and Worley, 2014.)

Our preliminary efforts were successful in creating a tree that chronologically resembled the release schedule of ICOADS (e.g. the root node was the earliest release of ICOADS and furthest nodes were latest releases). To evaluate the accuracy of this work, we presented a poster at the Fourth JCOMM Workshop on Advances in Marine Climatology, which is a major event in the ICOADS community. We asked participants to annotate our tree, and give us feedback on:



Poster presented at CLIMAR IV meeting, with participants annotations.

1. The accuracy of the clustering of related datasets, and
2. Datasets related to or derived by ICOADS that were missing from our tree.

The feedback from the ICOADS community verified the accuracy of our work. In particular, many scientists recognized the data products that were clustered based on related features, such as derivative sea-surface temperature (SST) or arctic sea ice data products. However, a number of curators from other data repositories were able to identify missing datasets, and logical relationships between clusters that were not well represented by the tree structure (See Figure 1.3 for annotations). We then manually added these datasets and their coded characteristics, and re-produced a maximum likelihood tree.

3.5.2 Interpretive Interactionism

The final method of data collection was a series of interviews with members of the ICOADS Steering Committee (ISC), a governance mechanism adopted by the community late in my research process (year three). In total, I interviewed six of the seven ISC members.

Semi-structured Interviews

Methodological Intention

The approach taken for conducting ISC interviews differed from ethnographic work for two specific reasons:

1. The ethnographic interviews were informal, and opportunistic. Using a semi-structured format would allow me to ask similar questions of different individuals and compare their perspectives.

2. With the permission of my participants, these interactions were recorded, partially transcribed, and analyzed for themes - allowing me to conduct further analysis of key ICOADS members and their experiences in governance roles.

Background Literature

The interpretive interactionist approach to qualitative research has two main objectives :

1. To capture the thoughts and ideas of an individual in their own words, recognizing that language is highly personal and that individuals place meaning on the words that they choose, the descriptive metaphors that they employ, and the narrative structures that they use in reference to the world around them.

2. To abstract from individual level analysis in trying to understand how personal meanings are shaped by interactions with people, and technologies in real-world contexts. (e.g. Prust, 1996)

Operationalized Study

I interviewed six of the seven ISC members via teleconference. I used a set of standard questions about ICOADS governance that I asked every participant, but modified these based on the number of years they had been involved with ICOADS, and their professional background. The conversations lasted anywhere from 38 to 80 minutes. I began each conversation with an overview of my research goals, and by obtaining permission to record the conversation. I had already presented research in front of these individuals, and was attending regular ISC teleconferences, so they were already quite familiar with me and my work.

Data Collected

Each interview was recorded, and partially transcribed for analysis. I also created a profile of each individual which included their relevant ICOADS publications, a CV, and a biographical sketch based on their personal websites or my previous interaction with them as participants of my ethnographic study.

3.6 Analysis of Case Study Data

The data collected through various methods in this dissertation has been aggregated in a research repository which hosts the software, protocols, interview tapes, and field memos written during ethnographic data collection. With the exception of taped interviews and field notes, which may reveal personally identifiable information about participants, these materials are available at:

After aggregating the data in a repository, I then used the higher level categories of the knowledge commons framework (Background Environment, Attributes, Governance, and Patterns & Outcomes) to code the various data. This consisted of re-reading material, listening to transcripts and assigning pieces of data to particular categories from the framework.

In coding this data, I attempted to answer questions posed by each category. For instance, using the *Background Environment* category I tried to answer questions posed by Madisson, Frischman and Strandburg (2014, such as "What is the background context (legal, cultural, etc.) of this particular commons?"; "What is the default status of the resources involved in the commons (patented,copyrighted, open, or other)?"; as well as questions guiding my own case study, such as"How was this default established?"; "How well do participants understand this default?"; and "Has the default changed?" I then iterated through each category of the framework until I felt that I could reliably answer each set of questions. In some cases I returned to participants to ask follow up questions and clarify their statements or interpretation of events that I observed. Below, I further describe these categories and their application for my analysis.

I then generated a set of secondary categories, and third level categories and repeated this process until I had exhausted all of the data sources. A similar iterative and approach was taken by Strandburg et al. who gathered quantitative and qualitative data while studying resource sharing in biomedical clinics (2014, p. 159-160), as well as Contreras' study of the genome commons (2014a).

3.6.1 Analytical Frameworks Used for Organizing Analysis

IAD

The Institutional Analysis and Development (IAD) framework was designed as method-agnostic tool that could simplify the analytical task confronting anyone trying to understand institutions in their full complexity. (McGinnis, 2011). The settings for much of this work were institutions that operated with a mix of governance structures, and had a need for longevity or sustainability of a single shared resource. Various versions of this framework can be found in the work of Kiser and Ostrom (1982), Ostrom (1990), Ostrom, Gardner, and Walker (1994), McGinnis (2000), and Ostrom (1998, 2005, 2007b, 2010, 2011a). This was to be achieved by integrating data collected by sociologists, lawyers, politicians, economists, and political scientists in trying to understand "how institutions affect the incentives confronting individuals and their resultant behavior" (Ostrom, 2005 p. 8). Below, I sketch out the most basic features of the framework.

The goals of the IAD framework can be broken down into it's two parts: 1. Analysis, and 2. Development.

The analytical goals of the framework are to guide a systematic study of formal and informal institutions, allowing for similar variables to be gathered and compared across cases.

The development goals are both diagnostic and design based; through design the IAD can account for how institutions are established, maintained, and transformed; and through diagnosis can identify missing institutions - or components of an institution - that are the a source of dysfunctional performance.

The IAD framework is divided into three levels of analysis for studying underlying factors of an intuition's success or failure: exogenous variables, action arenas, and outcomes and evaluations.

1. Exogenous variables include biophysical attributes of the shared resources, community attributes, and structural governance policies (formal or informal).

   Exogenous variables are also considered the input to a complex system. In an overview of the framework Frischmann gives a simple example of this group, from a lobster fishing grounds

   > "...attributes might include the relevant biological characteristics of lobsters, such as the rates at which they age and reproduce; attributes of the community of fishermen, such as the proximity in which they live to others, the existence of familial relationships, and the skill sets needed for lobster fishing; and the rules explicit or informal that govern fishing." (Frischmann, 2013 p . 8).

   The community attributes, using the same lobstering example above, could include the owners of boats and docks, the distance fishers live from one another, the

   The most crucial aspect of this level is that while individual attributes or variables are identifiable and observable, they need not be considered fully decomposable (Ostrom, 2005). This means that the salient features of a commons are often thought to be emergent properties where the mutual constitution of people and objects create phenomena which are indivisible. The use of multiple research methods for the same institutional setting are often justified on the grounds of needing to understand these emergent properties (Ostrom, 2005).

2. An Action Arena space where exogenous variables and actors interact, cooperate, and conflicts emerge.

   This level of the IAD studies the stakeholders and the resources of the commons as inventoried by the Exogenous Variables in a particular period of interaction such as the harvest of a shared common field, or the fishing season of a particular coastal community. Using the understanding gained from cataloging the Exogenous variables, the analyst observes this process and then identifies tangible outcomes, objective results, and the way that conflicts from the past, or the present are voiced, contested, settled or deferred. This analysis can take place over long periods of time or short site visits in a single time frame. These observations can also be supplemented with statistical analysis of quantitative data, such as the crop yields of a commons studied over a certain period of time, or even through agent-based models or multi-modal network analysis used to simulate these interactions.

3. Outcomes and Evaluation the results and feedback of action situations

   In the final stage of the framework an institution can be evaluated based on outcomes of the action arena; that is, how well an institution did or did not solve a collective action problem. Criteria for evaluation will vary by institutional setting, but usually includes calculation of costs, such as (1) information costs (2) coordination costs; and, (3) strategic costs (Ostrom et al. 1993). At a higher level of analysis, such as that aimed at developing policy, the sustainability of the shared resource can be analyzed in terms of (1) efficiency in access and use (2) equity of distributed wealth or costs, (3) accountability for resource management, or (4) adaptability of the institution in light of these outcomes (Ostrom et al. 1993; Imperial and Yandle, 2007).

   Over time the use of the IAD to study socioecological systems has proven useful in comparing the outcomes of real world scenarios with those presumed to outcomes addressed through public policy.

Knowledge Commons Framework

Initially described in 2010 as the “constructed cultural commons, ”(Madison, Frischman and Strandburg, 2010) the knowledge commons is both a concept, and a framework to describe institutional arrangements that govern the community provisioning and production of resource systems made up of “cultural ”artifacts - ranging from complex sociotechnical systems to simple cassette tape trading communities. The ambition of adapting the IAD to cultural sphere is to systematize investigations, provide guidance for a more rigorous set of evaluation criteria - especially as it relates to the matching of observation with existing theories and models- and to enable the integration of findings cases across different domains of knowledge production. (2014)

The proposed framework achieves this through three major innovations with the IAD:

1. It emphasizes among distributed actors, and the constructed nature of the resources themselves (purposefully built objects as opposed to natural resources)

2. Where the IAD separates outcomes (level 2) and patterns of interactions (level 3) the constructed cultural commons framework treats these as iterative, mutually constitutive processes.

3. As a result of the more complex relationships between resources, participants, and governance structures the relevant attributes of each may not neatly divide into categories. In short, this is the acknowledgment of a mutually constitutive socio-technical perspective.

The quintessential case for the Knowledge Commons are free and open source software projects such as Apache or Linux where there is a mix of proprietary rights, resource producers, users, and on-going tensions amongst these participants in the governing both the project as well as versions or distributions (Schweick and English, 2012). However, a number of diverse case studies beyond open source software have emerged using this framework, including the study of patent pools amongst technology firms (Madison, 2012), the Associated Press (Muarry, 2014), Wikipedia (Madison, Frischmann, and Strandburg, 2014), and notably, scientific infrastructures supporting large-scale collaboration in biology (Contreras, 2014, astronomy (Madisson, 2014) and biomedicine (Strandburg, Cui and Frischmann, 2014). These latter studies of knowledge commons in scientific collaboration further collapse a distinction made in the original IAD framework between the "patterns of interaction" and outcomes following an Action Arena. The argument for this innovation is that patterns of interaction generated by the formal and informal rule systmes are often inseperable from the outcomes it produces, "How people interact with rules, resources, and each other, in other words, is itself an outcome that is inextricably linked with and determinative of the form and content of the knowledge or informational output of the commons. "(Frischmann, Madison, and Strandburg, 2014)

In using the Knowledge Commons framework to organize my data analysis, I make two important modifications to previous works:

1. The first section titled ''Background Environment' is expanded in scope to include historic antecedents to the ICOADS project. Doing so makes the overall narrative more intelligible for a reader unfamiliar with climate science, marine climatology, or marine surface data. It also allows my analysis to demonstrate ways that historical events, and international policies governing meteorology impact the arrangement of contemporary institutions for data production and exchange.

2. I describe regimes that govern an "Action Arenas" rather than focusing on specific sets of events. Through the different releases of ICOADS data products, I demonstrate a co-evolution of norms and rules rather than a causal relationship between, for instance, behaviors and sanctions. This also allows for an investigation of governance issues over a longer period of time, and overall I believe this lends to a more thorough analysis of ICOADS governance. To achieve this, I draw upon a systematic approach to coding variables related to governance of a shared resource system. I describe this process in detail below.

3.6.2 Criteria for Interpreting Results & Answering Research Questions

To review - the Knowledge Commons Framework (KCF) provides four high level categories - Background Environment, Attributes, Action Arenas, and Patterns & Outcomes - for analyzing data gathered through a case study design. The ambition of this work is provide a way to systematically gather data, and compare findings from various different domains of knowledge production, looking for similarities and differences in the ways that institutional arrangements are able to overcome social dilemmas in sustaining a commons. As the designers themselves note, "...there is much work to refine and systematize this approach" (Frischman, Madison, and Strandburg, 2014). One way that the IAD has been systematized in the domain of socioecoligcal systems is through the design of project specific databases, where case study data can be uploaded and compared by a large number of collaborators. Out

Variable Name	Data Type
Begin date	Text
End Date	Text
Governance scale	Categorical
Multiple levels	Categorical
Institutional diversity	Ordinal
Centralization	Ordinal
Description	Text
Causal level	Categorical
Selective Pressure	Categorical
Governance trigger	Categorical
Knowledge type	Categorical
External recognition	Ordinal
Metric diversity	Ordinal
Policy instrument	Categorical
Scale match	Binary

Lists the variables to be coded for ICOADS governance.

Type	Whether the variable is measured at the (1) interval, (2) ordinal, (3) categorical, or (4) open-text level.
Component Type	What type of component in the Knowledge Commons (e.g. Background Environment, Attribute, Action Arena, etc) the variable describes.
Question	What question(s) is posed to the user when they are measuring the variable for a case study
Select options	The range of values that the variable can have
Unit	For interval variables, the unit of measurement.
Role	Records whether the variable describes components with a particular role (e.g. users, producers, provisioners, etc)
Importance	Describes the theoretical implications, as well as the meaningful interpretations of the variable selections
Definition	Provides a basic definition of the variable and what it is meant to achieve
Domain	Records what sector(s) (e.g. scientifc; cultural; etc) the variable is associated with, if it is specific to a particular sector.

Defines the properties of each variable.

of the necessity of creating a database structure, the process is systematized, creating variables, properties, and resulting definitions for how to properly code case study data.

I draw on this work to create a set of definitions, variables and their properties for the governance of ICOADS. In particular, I draw on the Social-Ecological Systems Meta-Analysis Database (SESMAD) project's definition of governance variables (Cox et al, 2014). Below I offer two tables:

1. Lists the variables to be coded for ICOADS governance.

2. Defines the properties of each variable.

Each variable is further defined by the properties for that particular variable[5].

Properties consistent with each variable are defined below. Not all properties are required for a variable to be coded correctly. For instance, domain and range categories may be unnecessary for some variables.

3.6.3 Validity: External and Internal

External

External validity was achieved through my presentation and sharing of findings from this work with members of the ICOADS community, and in particular two key informants who have regularly read drafts of this work, made suggestions for improvement, and clarified my writing for historical accuracy. Additionally, my key informant worked with me in coding of variables related to governance. This included exchanges about the proper definition of a variable, as well as what values the variables should contain.

Internal

Internal validity is addressed through triangulation of data collection, and by comparing the results of my work to case studies that have used the same framework in different settings - including the genome commons in biology (Contreras, 2014), the Urea Cycle Research Network in biomedicine (Strandburg, Cui, and Frischmann, 2014), and the Galaxy Zoo project in Astronomy (Madison, 2014). Synthesizing results across these diverse domains of knowledge production is what allows me to address the stated research questions of the dissertation:

1. What are the effective institutional arrangements (governance) for sustainable community based science infrastructures? and,
2. How and why do they differ between domains of knowledge production?

Answers to these questions are presented in Chapter 5.

3.6.4 Limitations

The case study design described above has a number of limitations, including;

[5]This information can be found at http://git.io/vJnPX

- This is a single case and the results will be difficult to generalize from, especially for large concepts like sociotechnical systems, or knowledge commons.

- The informetric studies are heuristic, and imperfect from traditional Informaiton Science standards. Instead of judging their signicance with p-values, I would argue that they were useful tools for collaboration with my participants.

- A focus on governance is only one component of a sustainability study. This will limit the ability of this work to provide comrephenisve guidance on sociotechnical systems sustainability.

3.7 Summary

In this Chapter, I have described the range of empirical research methods used to collect and analyze data; including the case study design, process of data collection, as well as the analysis and organization of the results into a modified knowledge commons framework, and the coding of variables related to evolution of ICOADS governance which draws upon the SESMAD framework. Results from this analysis will be presented in Chapter 4 (following). In Chapter 5, I use these findings to answer the stated research questions of the case study. I then compare these results to previously completed case studies, and answer the stated research questions of the dissertation overall.

CHAPTER 4

4.1 Introduction

This chapter presents an analysis of ICOADS governance using the knowledge commons framework. I begin with an overview of concepts from previous chapters. I then use data collected from the case study of ICOADS to fill in the major framework categories - Background Environment, Attributes, Governance & Rules-in-Use, and Outcomes. Further synthesis of these findings appear in Chapter Five.

4.2 Knowledge Commons Framework: Overview

To briefly summarize the discussion from previous chapters:

1. In socioecological systems sustainability, the commons has proven to be an effective institutional arrangement for sharing existing resource sets amongst diverse peoples. In sociotechnical systems, commons governance can also be used in creating, modifying, and remixing new resources; the latter is often described as a knowledge commons.

2. Resource sets making up a knowledge commons tend to resemble public goods - meaning they are non-rivalrus and non-exclusive. Governance in knowledge commons often rely upon loose interpretations of jurisdictional law (patents, copyright, etc) or unspoken norms and rules that are learned over time.

3. Both socioecological and sociotechnical systems overcome collective action dilemmas related to sustainability (e.g. free riders, tragedy of the commons, etc) through these rules and norms. However, as a result

of differences in the way the two types of systems are produced, and provisioned they often face different types of social dilemmas. Therefore governance effective in one domain may, or may not be effective in another. Categorizing and identifying the unique social dilemmas of sociotechnical commons is a first step in systematizing the study of these institutional arrangements. The following case study is focused largely on understanding how governance impacts sustainability.

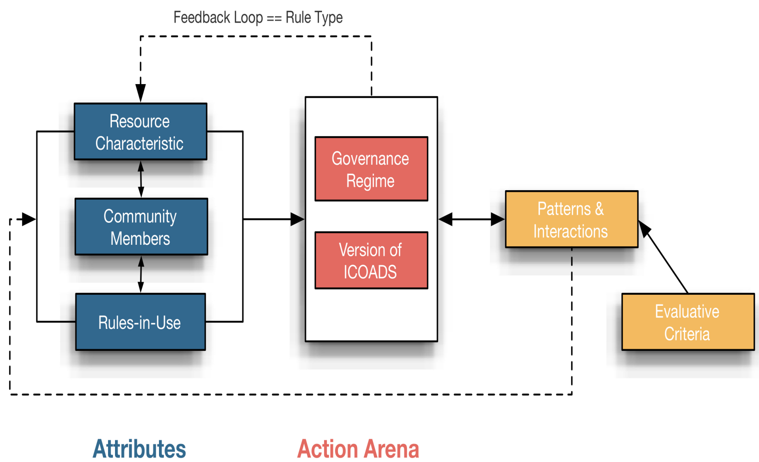
4. Unlike socioecological systems, sociotechnical systems are intentionally created. Normative issues related to the planning, design, and deployment of sociotechnical systems is critical to understanding their current and past states. History, narrative, and shared culture are recognized as important factors for studying the evolution of knowledge commons, but are difficult to access via observation-based research methods. Therefore, archival resources, and historical approaches are emphasized for developing a comprehensive case study.

The Knowledge Commons Framework (KCF) is a recent attempt to use socioecological frameworks for studying sociotechnical systems. In part, the KCF is meant to help reduce the complexity of studying institutional arrangements in shared resource systems (Frischmann, Madisson, and Strandburg, 2014). KCF makes two major modifications to frameworks used to study sustainability in socioecological systems, notably the Institutional Analysis and Development (IAD) framework:

1. Interactions between the attributes of the community, it's rules, and the constructed nature of the resources being pooled are seen as mutually constitutive. This collapses a distnction made by the IAD between these three components. At the crux of this modification is a difference between systems which are designed and built versus those that naturally occur.[1]

2. The KCF also collapses a distinction between "patterns of interactions" and the outcomes that follow. This is to recognize that these interactions are not finite, but instead iterate through cycles of production, consumption, use, collapse, re-design, and eventually reproduction.

[1]This point will be further emphasized in the following case study.



The Knowledge Commons Framework

In the following sections I use the Knowledge Commons Framework to analyze ICOADS sustainability across three governance regimes. I occasionally draw upon interview data from the "interactionist" studies described in Chapter 3. Where this is the case, I distinguish between participants with the following convention - ICOADS Steering Committee (ISC) and randomly assigned number (e.g. [ISC-06]). Participants are not identifiable through this coding nor the selection of transcribed interviews used in this chapter.

4.3 Background Environment

Knowledge Production in Climate Science

To better understand the historical precedence of ICOADS it is important to first explore some of the unique aspects of the background environment from which this project emerged. This includes some preliminary words about differences between weather and climate data, and the types of research that each facilitate.

An overly simplistic explanation might be that climate data are just weather data aggregated, and then averaged over long periods of time. This is partially true in that a basic understanding of the climate is necessarily dependent on past weather related observations, but the coverage and completeness of those weather records are, for reasons I describe below, highly variable, and the calculations of past events hardly average.

Understanding and studying climate is based on inference, calculation and interpretation of past weather observations which are heterogeneous in type, of variable quality, and often subject to different, and sometimes conflicting property rights regimes. Joseph Fletcher, the first PI of COADS, put the problem of data access for the study of climate this way:

> "In laboratory science we can often formulate the question ourselves and design experiments to test possible answers. For geophysical systems nature defines the behavior we seek to understand and we learn what that behavior is by observing nature. We must, also, test our hypotheses against this observed behavior. This is not as neat as a laboratory experiment but is unavoidable. The problem of understanding climate change is further exacerbated by the long time scale. We have no choice but to look backward in time because we cannot afford to wait for the future to unfold. " (Fletcher, 1992)

To bound this overview, I'll first answer two basic questions related to the enterprise of climate science research:

1. Where do historical weather records come from?, and
2. How is knowledge about the climate produced from these records?

Following this section, I then turn to the historical events leading up to ICOADS.

Where do historical weather records come from?

As Fletcher described above, the earth and environmental sciences depend on observational data to produce new knowledge. Contemporary sources of observational data include platforms such as networked censors, land-based observing stations and data loggers, radiosondes and aircrafts capable of sampling in the upper-atmosphere, satellites, drifting and moored buoys, and many manual forms of data collection such as a graduate student sitting in a field digging up fossils as evidence of a species occurrence event.

Climate science, like many of the natural sciences, depends on historical records that have been preserved, archived, and made available for reuse in doing large-scale analysis. Because we cannot travel backwards in time to re-sample weather events of the past, the enterprise of climate science has developed a number of ingenious techniques and strategies to overcome limitations of spatial and temporal gaps in historical records. These innovations include finding new sources of data, such as geophysical indicators of weather events taken from the cores of trees, or coral reefs; the digitization of analog resources such as weather logs kept aboard sailing vessels from past centuries; or, through sophisticated computational techniques that stretch existing data to cover geographic regions and time periods with sparse data coverage.

When it comes to data the field of climate science is entrepreneurial - it takes whatever resources are available, and where there are none, develops new techniques to produce a best estimate of climatological history.

How is knowledge produced from historical weather data?

Climate science is a slow moving, and increasingly conservative domain of knowledge production. This is due in part to it's recent politicization, but also due to the fact that it relies upon the steady improvement of a pre-existing set of observations. Like historians who rummage through the archival records of man, climate scientists are the archival scavengers of the earth's record. As Edwards describes this view, '

> "Their work is never done. Their discipline compels every generation of climate scientists to revisit the same data, the same events digging through the archives to ferret out new evidence, correct some previous interpretation, or find some new way to deduce the story behind the numbers. Just as with human history, we will never get a single, unshakable narrative of the global climates past. Instead we get versions of the atmosphere, a shimmering mass of proliferating data images, convergent yet never identical." (2011, p. 431)

The "convergent yet, never identical" nature of climate science described by Edwards is partially what makes comparison, sharing, and cooperation so incredibly important to the generation of reliable knowledge in this domain. As such, the sharing of datasets, the creation of centers of excellence around one particular type of model, or physical system are commonplace.

Unlike other domains of knowledge production, climate research results are regularly summarized in national and international reports that act as "consensus" documents about the state of climate knowledge. These high-level documents serve as the basis of policymaking and international political and economic action related to climate change. This includes highly publicized documents such as as Assessment Reports (AR) issued by the International Panel on Climate Change (IPCC), as well as national level reports that are more regionally focused on extreme weather events and their societal or economic impact.

4.3.1 History and Policy of Meteorological Data

In this section I'll cover two important aspects of ICOADS background environment:

1. Early international cooperation and standardization efforts by Lt. Matthew Fontaine Maury, and

2. A set of international resolutions, treaties, and bills in the US Congress that address the commercialization of meteorological data.

4.3.2 Early International Cooperation

Early meterological efforts to improve the reliability of weather data included efforts to both increase the speed at which information was exchanged, and coverage of observations. For instance, the foundation of the US Postal Office in 1792 included expedited services for weather related information (John, 1998), and as early as 1849 the Smithsonian began issuing meteorological devices to telegraph companies in hopes of increasing the speed at which weather observations could be collected and exchanged (NOAA, 2011).

At this time, most data sources were land-based observations, or lower atmosphere measurements taken through rudimentary meteorological instruments. A critical dimension of this expanding network were the weather records of the sea - which posed a considerably more difficult challenge for data collection . In the middle of the 19th century a U.S. Naval lieutenant, Matthew Fontaine Maury, would attempt to organize an international network of ocean-based weather observation, and in the process set the stage for what is now a robust marine climatology enterprise.

4.3.3 Maury's Rubbish Data

Like many tales or heroism Lt. Matthew Fontaine Maury may have fallen into greatness rather than chosen it for himself. On his honeymoon in 1839, Maury was thrown from a stage coach and badly fractured his leg. The injury prevented him from returning to a regular post at sea for twelve months. Maury spent that year writing a series of anonymous publications about strategic military directions for the US Navy, and agitating for new and improved marine charts. His letters garnered the attention of many naval officers and congressmen, eventually leading to his identity being revealed, and his appointment to a new post, the director of the NAVY Depot of Charts and Instruments.

While in this post Maury sought to improve the reliability of naval weather records. In doing so, he operationalized a massive effort to create charts of winds and sea currents from old log-books that "had been stored away in the Hydrographic Department as rubbish" (Corbin, 1888). The processes he put in place for creating these charts, described in many accounts of this period, are similar to contemporary data curation; similar resources are aggregated, their measurements are normalized and recordings are quality controlled, and summary documents are produced for reuse.

Having circulated and promoted the resulting charts widely, word began to rapidly spread of their usefulness. By 1854, the charts were used regularly by shipping vessels the world over (Lewis, 1996). A merchant's magazine of the time calculated that the total amount of money that the shipping industry could save using Maury's charts was on the order of $2.25 million annually (Corbin, 1888 p. 56). A British report issued the following year estimated the savings to all British vessels at around £10 million[2].

After reports of the economic impact of his work, the Secretary of the Navy wrote to Congress pleading for renumeration on Maury's behalf. One letter draws particular attention to the free and open exchange that he "un-

[2]Both calculations are in dollar amounts of 1853, and 1854.

selfishly" initiated. The secretary goes on to say that "...Mr. Maury might have secured a copyright" for the charts which were "sitting unused," but instead Maury created a free resource that would allow ships to "sail more safely and quickly than ever before". The letter emphasizes that Maury had, through salvaging rubbish data, created a "common property of the world." (VMI as summarized by Corbin, 1888)

Along with the "Charts and Winds" publication Maury also distributed blank logs, and in the margins he prescribed a method of recording measurements such as wind direction, temperature, etc.

The impact of Maury's approach to data aggregation and analysis was further cemented by a publication entitled "The Physical Geography of the Sea and its Meteorology" which went through over twenty printed editions garnering Maury some small, but international fame.

In 1853 Maury coordinated an international congress on meterological observations in Brussels, Belgium. Participating nations included Germany, Austria, UK, Netherlands, Sweden, Norway Spain, as well as Brazil, Chile and Prussia (Corbin, 1888 p. 72).

At the 1853 Congress Maury made two proposals:

1. "...all maritime nations should cooperate and make these meteorological observations in such a manner and with such means and implements, that the system might be uniform and the observations made on board the public ship be readily referred to and compared with the observations made on board all other public ships, in whatever part of the world."

2. "...it becomes not only proper, but politic, that the forms of the abstract log to be used, with the description of the instruments to be employed, the things to be observed, with the manipulation of the instruments and the methods and modes of operation should be the joint work of the principal parties concerned."

Each participating nation agreed to carry out these specific terms, and to broad and sustained meteorological cooperation more generally. In a letter to USA Congress, Maury notes his success as well as the precedence of this meeting:

> "Rarely has there been such a sublime spectacle presented to the scientific world before all nations agreeing to unite and cooperate in carrying out according to the same plan one system of philosophical research with regard to the sea. Though they may be enemies in all else, here they are friends. Every ship that navigates the high seas with these charts and blank abstractlogs on board may henceforth be regarded as a floating observatory a temple of science. "(Corbin, 1888. P. 275)

Commenting on this foresight 150 years later, the director of the World Meteorological Organization (WMO) saw Maury's contribution as providing a normative framework for the field, lying down principles that were "...so basic to our understanding of how meteorology should be done; (a) all nations should cooperate; (b) observations should be standardized; (c) the enterprise should be global; and (d) the parameters measured, the data recording and exchange, and the instruments and methods of observation should follow an agreed plan." (Rasmussen, 2003)

In summary, Maury's legend teaches us two important things about contemporary work in climate science:

1. He demonstrated the immense value in aggregating and summarizing previously existing data, and in the process developed schemes to standardize the recording and observations of data, greatly reducing normalization and aggregation efforts in the future.

2. Maury established the principal of international cooperation for the free and open exchange of marine data. Twenty years after this congress, the International Meteorological Committee (today the World Meteorological Organization) was founded to institutionalize this international cooperation. It would take another century before the principal of free and open data exchange would be cemented in policy.

I turn now to the historical moments in which the WMO was formed, and the continued creep of commercial interest in this domain.

4.3.4 WMO and the Principle of Free and Unrestricted Exchange

In the wake of the cooperative agreements made in Brussels during the 1853 congress, many nations began to establish national meteorological offices and weather services. But it wasn't until 1950 that a World Meterological Organization (WMO) was founded. In 1953, exactly 100 years from the first international congress in Belgium, the WMO articulated the following conventions:

- To facilitate worldwide cooperation in the establishment of networks of stations for the making of meteorological observations or other geophysical observations related to meteorology and to promote the establishment and maintenance of meteorological centers charged with the provision of meteorological services;
- To promote the establishment of systems for the rapid exchange of weather information;
- To promote standardization of meteorological observations and to ensure the uniform publication of observations and statistics;
- To further the application of meteorology to aviation, shipping, agriculture and other human activities; and
- To encourage research and training in meteorology and to assist in coordinating the international aspects of such research and training. (WMO, 1953))

In the following years, two critical resolutions from the WMO were passed - both of which hold implications for the sustainability issues facing contemporary climate science research infrastructures.

Resolution 35 - 1963

In 1963, WMO Resolution 35 established the open exchange of data among WMO members, including maritime log-books for processing historical weather observations. The resolution was non-binding, but set out recommendations for how data *should* be formatted for exchange, including International Maritime Meteorological (IMM) punched card and tape (IMMPC and IMMT) formats. This resolution was significant, because it sets a precedence for what is referred to as a "principle of free and unrestricted exchange of meteorological data "between members of the WMO. Like precedence in judiciary law, this principle will be continually referenced by WMO members advocating for the WMO to enforce open exchange of data through rigid policy language.

Resolution 40 -1994

In 1994 there was a strong lobby from a WMO working group on the commercialization of weather data to revise Resolution 35, along with a number of other previous resolutions guaranteeing the open and free exchange of data between all WMO member nations. The resolution draft was titled "WMO policy and new practice for the exchange of meteorological and related data and products including guidelines on relationships in commercial meteorological activities." The proposal outlined criteria for two tiers of WMO member data:

- The first tier would include all data required to carry out WMO programs - including extreme weather event monitoring, etc.
- The second tier, included all non-essential data. This was to be exchanged at the discretion of individual nations. (Co-Data, 1997)

By creating a broad "non-essential" tier of resources the proposal would allow commercial vendors to reach distribution agreements with individual nations, and in turn privatize the loosely defined non-essential data products. As a result, Resolution 40 became highly contested and a rift within WMO opened between free market advocates, and those who would defend a "principle of free and unrestricted exchange of meteorological data "(Flemming, 2013).

The resolution that the WMO eventually voted on, and passed in 1995 states the following:

"As a fundamental principle of the World Meteorological Organization (WMO), and in consonance with the expanding requirements for its scientific and technical expertise, WMO commits itself to broadening and enhancing the free and unrestricted international exchange of meteorological and related data and products" (WMO, 1995)

The two separate tiers of resource access were eliminated, but there remained an ambiguity in the resolution's language which read that "data (and related products) required for carrying out WMO program "shall" be made available by all members", and all other data ""should" be made available" (WMO, 1995). This caused a considerable amount of confusion about what constituted data required for WMO programs, including data that could have an intellectual property claim by individual scientists. On this matter, minutes from a CO-DATA working group at ICSU are telling - summarizing the somewhat strange wording they note "Group members argued that it was inconsistent to urge full and open access and yet establish a policy whereby scientists can have proprietary rights for a period of time." (1997) These distinctions also did little to clarify how WMO members from developing countries could develop a cost-recovery model that charged for access to data.

Although Resolution 40 avoided creating a commercialization of "non essential" meteorological data it did create ambiguity in what was considered to be a viable method of cost recovery. What's more, it did little to quell the potential of enclosure for the meteorological data commons. The impact of this ambiguity is also important for understanding why meteorological projects so often speak in collective terms, and why historical events, such as Maury's 1853 Congress, are so often described in the formal literature of marine climatology: these events are a subtle reminder that although policy creates the opportunity to act otherwise, it is the historical precedence of cooperation that *should* guide contemporary actions.

4.3.5 Continued fight over intellectual property

In the broader legal and political arena of the 1990's similar debates over copyright and intellectual property of data were taking place. I'll summarize three of these events relevant to ICOADS: 1. Language from a rejected WIPO treaty on intellectual property protection for database creators; 2. The failed bill H.R. 3531; and 3. A second failed bill, H.R. 2652, both in the

US Congress.

In 1996, an international treaty was developed and put up for consideration by the World Intellectual Property Organization (WIPO) to make databases a patentable intellectual property (including data held in weather and meteorological data centers). The proposal included consideration for the following key points related to climate data, and environmental information more generally:

- Prohibit unauthorized extraction, use, or reuse of any database, or any substantial portion of a database (as defined by the database vendor), and effectively establish the basis for a pay-for-use system,
- Apply to all privately generated data or repackaged U.S. government data, and
- Establish strong civil and criminal penalties, including penalties for third-party liability (liability incurred by an unwitting intermediary or disseminator). (NCAR, 1997)

After much debate, the WIPO rejected the language of this proposal, in part, because it lacked a firm definition of what could be considered "intellectual effort" in assembling a collection of facts into a database. Instead, the "Berne Convention" which creates international copyright of digital materials, not including scientific data, includes the following language "protection of intellectual property rights may extend to compilations of data or other material (databases), in any form, which, by reason of the selection or arrangement of their contents, constitute intellectual creations. (Where a database does not constitute such a creation, it is outside the scope of this Treaty.)" (WIPO, 1996).

In response to the both the WIPO's decision and Resolution 40, NCAR's then president Richard Anthes wrote the following,

> " Despite the recent successes with WMO and WIPO, few people believe the debate is over for good. The issues involved in using the Web for dissemination of data, and in balancing the rights of the commercial sector with the legitimate needs of the research and educational communities, are pervasive and challenging enough to ensure continued discussion. Perhaps the biggest

victory in the recent bout is that the arena for this discussion has expanded significantly. The atmospheric sciences community, which has written so persuasively in recent months, will remain engaged as the debate unfolds." (1997)

And indeed, these debates continue up to present day WMO activity. For instance, in 2013 a panel of experts was commissioned by the WMO Executive Council to draft a report outlining the need for Resolution 40 to be revised. One of the panel's main questions was whether or not the "principle of open and free exchange" was being honored in the Global Framework for Climate Services (GCFS), a UN-led initiative to guide the development and application of science-based climate information and services in support of decision-making (WMO, 2015)[3].

Two other important attempts at privatizing meteorological data are worth noting:

1. In 1996, H.R. 3531 "Database Investment and Intellectual Property Anti-piracy Act" was introduced to the US House of Representatives. This bill sought intellectual property protection for digital technologies, including the patenting of scientific databases. Debate about the impact of the bill included extended discussion about how patents would have affect the USA's ability to participate in the exchange of scientific data supporting international research agendas. The bill never made it out of committee and expired with the 104th Congress (Sarewitz, 1998).

2. The 105th Congress would try again. H.R. 2652, titled "The Collections of Information Anti-piracy Act, " was a broader bill that included a number of different ways for data and software to be patented. This bill was eventually passed by the House Representatives, but its language was folded into the Senate's "Digital Millennium Copyright Act "(DMCA). Like its predecessor H.R. 2652 expired with the 105th Congress. Through Senate negotiations around DMCA, language including the protection of scientific databases and data was eventually dropped. (Sarewitz, 1998)

[3]As of May, 2015 this panel has yet to report its findings.

Conclusions from Policy & History

The review of relevant policy and historical events sketched above provides important background for the development of ICOADS - both from an ethical, as well as a legal prospective. The default settings of cooperation are made clear with two conceptual turns of phrase: 1. The description of Maury's standardization work as providing a "common property of the world", and 2. The commons sentiment being further codified in WMO policy that protects a "principle of free and unrestricted data exchange"; and the continued political defeat of bills attempting to make meteorological databases patentable.

4.3.6 Goals and Objectives

At a broad level, the WMO resolutions described above establish a basis for the governance of exchange and access to marine surface data. However, to focus on formal policy alone would be to overlook a wealth of innovation and ingenuity in the design of cooperative institutions for the production and provision of these resources. The collaborative communities responsible for producing these resources are nested within a number of different organizational settings, and bound to different disciplinary and cultural norms. An overlap of institutional commitments has required each new climate or meteorological endeavor to reflect upon the established norms of the WMO, Maury, and other their commitment to principles of free and open data. Few of these institutions are bound, legally, to cooperate or share resource with one another.

The initial goal of the COADS project was to fill a gap in the historical climate record through creating a systematic and high quality archive of marine data. An early publication notes that the impetus for COADS is a well understood, but under-appreciated fact: "The world ocean covers over 70 percent of the earth's surface. The history and future of global climate therefore cannot be understood without ocean weather observations." (Woodruff et al., 1987)

COADS ambition then was to be a "...readily accessible archive of global climate information" not simply a dataset, as the name implies (Woodruff et al., 1987). This is an important distinction as it has ramifications for the definition of the resource set, and the property rights claimed by ICOADS users, developers, and provisioners.

Critically important to this development was the inclusive nature of the project. A more contemporary publication from the core team of ICOADS developers summarizes the sentiment of this project nicely,"This is an open community and new participants are welcomed." (Woodruff et al., 2011) In an interview with a longtime international partner of the project, I asked about early work in developing a public good, and she replied this way, "...its not, this isn't something that we specifically talk about, but it underpins everything we do. It is this idea of creating a public good, making everything accessible. Making it easy for people really underpins everything we do. I don't think we ever talk about it... but we're in the business of creating public goods." [ISC-02]

4.4 Attributes

This section includes an overview of characteristics defining the COADS resource set, and community members.

4.4.1 Resource Characteristics

In the most basic sense, ICOADS is a set of computational infrastructures and databases containing historical marine surface records from the 17th century to the present day, available for download from two data centers (NCAR's Research Data Archive, and NOAA's National Climate Data Center). The resource set - that is, the various resources needed to make meaning of these records - is considerably more complex; including, documentation and metadata records about the data, software for accessing and sub-setting the data, and services for curating and provisioning the databases (as well as producing new versions of the databases).

Surface Marine Data

Surface marine data include physical phenomena observed and captured at the Ocean and Atmosphere interface - typically considered to be 15 meters into the atmosphere from the sea surface. Data points might include any of the following phenomena: Air Temperature, Cloud Amount/Frequency, Cloud Height, Cloud Types, Dew Point Temperature, Humidity, Ice Edges, Precipitation Amount, Pressure Tendency, Sea Ice Concentration, Sea Level Pressure (SLP), Sea Surface Temperature (SST), Surface Winds, Swell Visibility, Wave Frequency, Wave Height, and Wave Speed/Direction.

Marine surface data have been collected using a variety of measurement techniques, including on-board ship instrument, moored and drifting buoys, radiosonde, and satellite. The processes of data collection in each of these platforms has changed significantly over a period of coverage which reaches back to the middle of the 17th century. For this reason the quality, reliability, coverage, accuracy, and accessibility of these historical records are, to say the least, variable.

As such, doing climate research with marine surface data requires coverage that is both spatially and temporally significant. Spatially significant means that data points are sampled equally across the ocean's surface. Temporally significant means that data are recorded at regular intervals of time across those spaces. The infrequent and unsystematic nature of how data were collected by sea-faring vessels, which are the overwhelmingly dominant observing platform for marine data until the 1970's, are an artifact of the rudimentary practices of sailing in the seventeenth and eighteenth century: ships took indirect routes between ports, and often recorded weather data infrequently or haphazardly as a result.

To overcome limitations in the historical record, data are often interpolated or "gridded "(converting scattered individual data points from a single observed surface into a regular grid or raster of derived values). Gridded data are produced in monthly summaries so as to make large volumes of data accessible to end users.

Data centers or individual research teams may each take a different approach to correcting errors, resolving gaps in spatial and temporal coverage, or quality controlling marine surface data. This means that although the same historical records can be used as inputs by different data centers, they can and do develop appreciably different blended data products. Taking different approaches to quality control also means that one archive, or research team, can specialize in a particularly difficult aspect of the historical climate record, and even build a reputation internationally for producing high quality, reliable products based on a single variable.

Unlike the laboratory sciences, datasets in meteorology are not built from a single source, but instead drawn from a common, openly accessible pool of resources that are beyond the capabilities of any one institution alone to collect. These derivative products require substantive intellectual efforts to improve. This often makes claims on property rights related to marine surface data unclear, and occasionally contested. The "common property of the world" sentiment espoused by Maury largely carries over to the sharing of data products in this domain, however, as described in detail below, these issues are a form of social dilemma which threatens the sustainability of these resource sets.

ICOADS

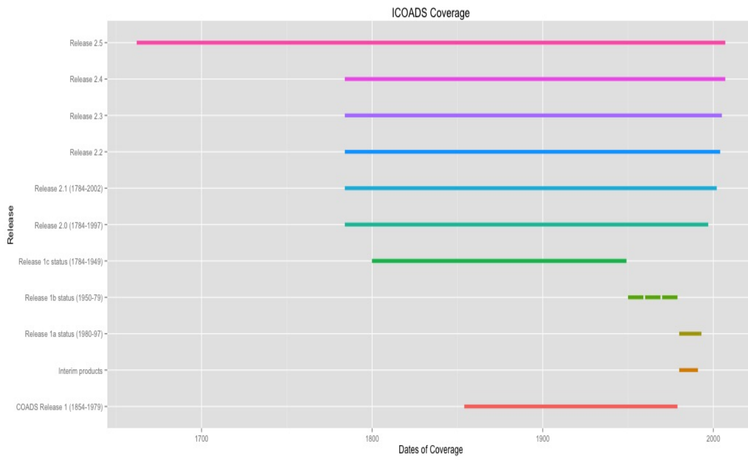
As described above, ICOADS was meant to be a comprehensive archive of historical marine surface weather information. In an early project meeting, the lead PI of COADS put this directly, "When we look backward in time we find the observational record frustratingly incomplete. We have satellite records for only a couple of decades and very incomplete radiosonde coverage for about four decades. To document longer term behavior we have only surface observations and for reasonable spatial coverage of this water planet they must include data from the ocean domain." (Fletcher, 1992)

The first release of COADS in 1983 included 11 different data products and their processing history (Jenne and Woodruff, 1986), a Fortran 77 program on magnetic tape used to read the packed binary-data (Woodruff et al 1987), as well as a number of publications describing the statistical trimming methods used to create derived variables (monthly summaries) of the aggregated data (Slutz et al., 1985; Fletcher, 1983).

Below I describe four key components of the ICOADS resource set: 1. Data, 2. Metadata, 3. Standards, Software & Statistics, and 4. Computational Infrastructure.

Data Products

ICOADS data products have increased in size, and sophistication over time. The first release of COADS in 1983 included 71.9 million records. Release



The expanding coverage of ICOADS over each new data release.

2.5 of ICOADS includes over 250 million records. As new resources are discovered, digitized, or reanalyzed for improvements, the coverage of ICOADS data products continues to improve in terms of historical coverage. The first release of COADS included records dating back to the 1850's, but continued international cooperation, as well as recent contributions from projects like the RECovery of Logbooks And International Marine data (RECLAIM), and citizen science initiatives such as the Old Weather project have extended the records back until the middle of the 17th century (1662).

Metadata

Use of surface marine data requires extensive documentation and metadata. This includes formal manuals developed by the WMO to code ship, instrument, and data types, but also less formal documentation, especially for historical data, such as the marginalia in sailing logs, or even the journals kept by crew members that may have noted irregularities in weather during particular days.

The accuracy of some observing platforms are well understood - such as drifting buoys which are rarely recovered and repaired and so their measurement accuracy may drift undetected and result in decreased accuracy over time (Woodruff et al. 2011, p. 17). Other platforms, such as data collected aboard early sailing vessels, may have limited metadata and are not well understood in terms of accuracy or quality. This variability in platform quality and metadata, as well as the discovery of new data sources (i.e. logbooks that have not been digitized, contextual documentation that is found in an archive, etc) requires a slow, but nearly constant revision of the historical record.

Additionally, blending these different platforms into one cohesive data product is a non-trivial exercise that requires consistent applications of quality control, and a robust source of computing power to process and homogenize these records. For instance, in announcing the first release of COADS, the project partners make this iterative process clear, warning that,

> "Any conclusion drawn from the historical record should be qualified by the fact that the observation, reporting, collection, and digitization of these data have been subject to a great deal of methodological change. Besides introducing more or less unknown inhomogeneities into many variables, these changes have sometimes been processed incorrectly. The resulting errors, as well as simple recording or transmission errors, occur very frequently. While a major effort has been made to indicate reports containing errors, some kinds of errors cannot be trapped by statistical methods. A very common error in the original data was incorrect representation of latitude and longitude, and only in extreme cases were these identified. Thus it must be remembered that while millions of errors have been identified and eliminated from the trimmed summaries, the resulting data are still far from clean." (Slutz et al., 1985)

Some project partners have made their entire career around improving metadata to marine surface data, and contributing those corrections and improvements back to international projects such as ICOADS (e.g. Kent et al., 2007).

Standards, Statistics Software

Finally, two other products important to the ICOADS resource set are the standards it has developed for the documentation and exchange of marine climate data, and the statistical approaches it has created to summarize these types of data. Additionally, the software written and traded amongst community members for accessing and analyzing subsets of ICOADS data are becoming more important to the future work of this project. I'll briefly describe each of these products below.

- As marine surface records became more valued in climate reconstruction projects there also emerged a need to standardize the format of exchange for historical marine data. Until 2000 no uniform, internationally agreed upon format existed. Through a series of proposals, the International Maritime Meteorological Archive (IMMA), was recommended for a standard format for exchanging marine meteorological data, developed and used extensively for the 2.5 release of ICOADS. In short, IMMA is "ASCII-based format with a fixed core set of the most commonly reported meteorological variables sufficient for most users, also including the time, location, and individual platform identification (e.g. ship call sign or WMO buoy number, if available). (Woodruff et al., 2011)" As a standard IMMA, and its broader adoption, facilitates the exchange of many international projects that model their work after ICOADs, such as the Climatological database for the World's Oceans (CLIWOC). More broadly, it is a contribution endorsed by JCoMM, and was one of the most successful innovations of the project beyond the provisioning of data.

- The ICOADS community, like much of contemporary science, produces and uses a number of unique software packages for accessing and making use of subsets of ICOADS data. These include anything from simple scripts to fetch data from shared servers, to large Fortran libraries for reading and analyzing data in the IMMA format.

- The summary statistics developed to grid ICOADS data are a non-trivial aspect of the innovation of this project. Originally meant to reduce the space that the data would take up on tape reels, the process of calculating monthly summaries, and gridding data at 1 x 1 resolution remains in use today largely for convenience of the community of end users - innovating with this scheme would be both inconvenient and costly with uncertain payoffs.

4.4.2 Community Members

Data Users

Users of ICOADS data are diffuse. In the climate science community, ICOADS data have been used to produce a variety of new data products, and used within a number of different international climate assesment reports. This includes staff scientists where ICOADS is archived (NOAA's NCDC, and NCAR's RDA), as well as scientists and graduate students at universities around the world.

ICOADS data are occasionally used outside of climate science; examples include a digital historian who traced naval operations during World War I (Mullens, 2013), and a silicon valley entrepreneur giving a commencement address at the Naval Academy (Ondrejka, 2013).

Data Producers

Data producers in the ICOADS community are of three types:

1. Historical Curators, including individuals whose research agenda are based around bias corrections in the historical climate record.

2. Contemporary Collectors, including institutions that contribute data to the GTS through moored buoy censor networks, as well as ships which participate in the WMO's Voluntary Ship Observation (VSO) program.

3. Contemporary Curators, including software engineers that both develop new and provision existing data. This includes normalizing new data from the GTS, correcting minor errors in existing data, and providing improved software for subsetting ICOADS data products. Curation also takes place during the production of new ICOADS releases, including the cleaning (normalization and quality control), preprocessing, and scheduling of tasks on NCAR's supercomputers. ICOADS data curators are also the individuals that update and make improvements to publicly facing metadata, send reports or emails to current users to announce known errors, and fulfill individual subset requests.

Data Provisioners

In this section I focus on two institutions which serve ICOADS data, documentation, and software to end-users.

NOAA/ Earth Systems Research Laboratory (ESRL) Physical Sciences Division (PSD), Boulder, Colorado, USA (and) National Climate Data Center, Asheville, North Carolina, USA

NCDC is a relatively new partners in the ICOADS project, and has taken on the service ICOADS data to expert end-users in the federal government (for reasons described below). As a result, Few user services are available, and downloading is only possible through FTP servers. As a result, NCDC serves mostly expert users and federal employees of NOAA.

A major contribution to the provisioning of ICOADS data from the ESRL's PSD is made through the project primary investigator, Scott Woodruff. Until the winter of 2012, he and one other FTE are responsible for coordinating the documentation, updates, and software development for all on-going ICOADS work. The responsibilities of these individuals also includes handling much of the design and implementation of new releases or enhancements to existing data products served by NCDC.

The Research Data Archive (RDA) National Center for Atmospheric Research, Boulder, Colorado, USA

The RDA is a repository of atmospheric and oceanographic observational data, weather prediction model output, gridded analyses and reanalyses, climate model output, and satellite derived data that has been curated by staff in the Computational and Information Systems Laboratory (CISL) at NCAR since the mid 1970's (Jacobs and Worley, 2009). The RDA is unique amongst earth science data archives in that it serves almost exactly the same amount of data as it stores- meaning that the RDA last year contained about 1.3 Petabytes of data, and in total it also served users, located all over the world, about 1 Petabyte of data.

A major component of NCAR's mission is to serve the entire atmospheric science research community - from government scientists to undergraduates.

As a result, the RDA's provisioning of ICOADS has focused on serving end-users of varying skills and technical competencies. For instance, over the last thirty years, the RDA's staff has developed a robust graphical user interface (GUI) allowing for unique subsets of ICOADS data - and also offer curation services to create unique subsets for users who cannot manage the GUI. The RDA has developed a number of metadata elements that are meant to help users discover the dataset more easily. The RDA also provides a suite of Fortran software that allows for observations made in the IMMA (International Maritime Meteorological Archive) format to be read and used in a variety of applications, and they also provide a global archive of year-month observations. In short, access to a number of different ICOADS products, and additional services surrounding the dataset leads to the RDA serving about 90% of all ICOADS users on an annual basis (the rest being served through NCDC).

4.5 Governance & Rules-in-Use

Governance and rules, as argued throughout Chapter 1 and 2 of this dissertation, are critical aspects of effective institutions for collective action. Governance and rules are studied through what the IAD and KCF call "action arenas," which are analyzed by observation, participation, or archival works about how"individuals (acting on their own or as agents of organizations) observe information, select actions, engage in patterns of interaction, and realize outcomes from their interaction." (McGinnis, 2011; Contreras, 2014).

In the sections that follow I divide ICOADS governance, and the evolution of its rules into three distinct regimes of governance; the events around the formation and dissolution of a governance regime can be seen as an "action arenas," Within these "action arenas" I identify action situations where rules are proposed, enacted, or negotiated amongst participants in both the provisioning and production of ICOADS.

I draw on the ethnographic, and interpretive interventionist data in constructing the narratives below. In particular, these sections attempt to tie events of the past to contemporary issues through the actions, words, and writings ICOADS community members. I characterize the outcomes and patterns of interaction for each separate regime, but note that a full synthesis of the findings across all three regimes is found in Chapter 5.

4.5.1 Three Regimes of ICOADS

When I began interviewing ICOADS Steering Committee members, I asked each participant a variation of the following question: "If you were to look back at the history of ICOADS, what would meaningfully separate one period of the project's development from any other? "The first four interviewees all offered the same time periods - COADS (1980 - 2000), ICOADS (2000-2011), and the later (often difficult to name) period that resulted from funding problems beginning at NOAA (2012 - present). These periods were also named because they align, somewhat crudely, with each new data release (version) of ICOADS. In further interviews I corroborated these periods with other ISC and ICOADS community members who, by and large, agreed that these were meaningful divisions to make in the evolution of the project.

4.5.2 Regime 1: 1981-2001

Three important contextual factors that led to the initial realization of COADS:

1. In the late 1970's, many earth science disciplines had seized upon the opportunity of increased computing power and decreased cost of data storage to further a climate research agenda. Climate models were just beginning to couple separate physical systems (i.e. ocean and atmosphere, land and atmosphere, etc.) and the potential to create earth system models was brining with it the need for more authoritative, and more complete historical weather data resources (Trenberth et al. 2002). In the case of early COADS work, innovations with "blending" different data sources, and with "trimming" statistics to create monthly summaries made the potential for a historical data-set a reality.

2. Similarly, the geo-political landscape was becoming more peaceful, and cooperative than it had been in nearly half a century. Riding the success of the International Geophysical Year (1957-58), the exchange of archival data among cooperating nations became common place. A minor WMO resolu-

tion, 35 (Cg-IV) passed in 1963 further normalized this practice. During this period NOAA was able to obtain a number of international collections of punched card decks from major maritime counties - including archives from the Dutch, British, Japanese, Norwegian national meteorological offices, as well as a collection of rare Russian whaling logs from the 19th century. By the late 1960's efforts to combine the different logs had resulted in a series of "Tape Deck Families" (TDF) that could be shared amongst cooperating nations (Woodruff et al., 1987). Recognizing the potential value in this exchange, the WMO then initiated an international effort to improve bias in these records through the Historical Sea-Surface Temperature (HSST) Data Project; with the NSF later supporting the USA's participation through a funding program titled, National Science International Decade of Ocean Exploration (Quayle, 1977; Woodruff, 1986). Because the records were so large, and processing so computationally intensive, the HSST divided the project up by Oceans - the USA responsible for Pacific, Germany for the Atlantic, and the Netherlands for the Indian Ocean. This initial effort generated a new set of SST data products, as well as new techniques for both processing and combing different data sources (buoys, ship logs, etc.). Notably, this was one of the first WMO projects to use magnetic tape reels to store data -making it considerably easier to access and use larger volumes of the archive.

3. Advances in supercomputing were enabling data processing tasks, in particular the calculation of floating point operations, at speeds that were previously unimaginable. Although still prohibitively expensive to most research projects this new calculative power brought with it the possibility to create comprehensive records by processing a large number of inhomogeneous datasets at a single location (Jaffe and Woodruff, 1983).

Pre-assembly

While each of these factors was important for COADS to be initially assembled, it was perhaps the last point, which proved most fortuitous to the ongoing HSST effort to create a comprehensive archive. In 1980, Wilmot Bill Hess left NOAAs Environmental Research Laboratory (ERL) to become the Executive Director of NCAR. Shortly after his appointment Hess secured funding from DOE for NCAR to purchase a second supercomputer, a used

Caption: NCAR's supercomputing history, including the early CRAY 1a

Cray-1A[4]. Since NCAR purchased the Cray 1A used, it came with a trial period that allowed NCAR engineers to calibrate and verify the machines capabilities making certain that it met predetermined performance benchmarks.

Before moving to NCAR, Hess had worked with J.O. Fletcher and was sympathetic to the need for a comprehensive marine based climate record. Having seen the value of such an archive during the HSST project, Hess decided to allocate a full third of the new Crays trial time to the processing TDF data, that could be made into a single, homogeneous dataset. In a previously recoded interview Fletcher described how difficult ICOADS data processing was,"just to read through it without doing anything to reformat, or anything else, at that time was about $ 100,000 worth of computer time, just to read through the goddamn thing once". (Shoemaker, 1997)

In total, the donation included over 1000 hours of Cray-equivalent CPU time (Woodruff et al., 1987). In the initial publication that announced COADS to the climate community, Slutz et al acknowledge this debt, saying "Throughout the effort, the support and encouragement of Dr. Wilmot N. Hess was crucial, as Director of ERL during the early stages and as Director of NCAR during the later stages. "(1985)

COADS versions 1, 1a, 1b, & 1c

Having collaborated widely on the initial assembly and processing of the data, NOAA, NSF (through NCAR) and the Cooperative Institute for Research in Environmental Sciences (CIRES) at the University of Colorado, began a formal cooperative agreement to maintain and distribute the Comprehensive Ocean and Atmosphere Dataset (COADS) into the future (Fletcher, 1983; Woodruff, 1985).

In a foreword to one of the first COADS related publications, Fletcher describes the process of creating the archive, saying:

[4]Besides having about 4.5 times the throughput of NCARs current computing infrastructure, the Cray 1A also included the first automatic vectorizing Fortran compiler (Enabling many GCMs to be used and tested by NCARs community while they were developing the Community Climate Model 1 (Bath et al., 1987)).

> "It has taken four years and much effort by many individuals and several institutions to obtain and process the hundreds of tapes containing the basic data input. All of this effort was provided from ongoing activities; there was no appropriation identified for the task. It is a tribute to the spirit of cooperation among the participating organizations that the task has been successfully completed." (Slutz et al., 1985)

The fact that no direct government appropriations were used for the initial creation of the COADS meant that there was also no expectation about how it was to be served to end-users, updated, or maintained. While having no funding meant that this work was somewhat sporadic in being completed, it also meant that there were no cost-recovery expectations by federal funding agencies.

It is important to note the historical context of this work; I've already covered the WMO resolutions which described a moral imperative to freely share marine data, but there were also a number of other legal and political pressures at the time to monetize scientific research. In particular the 1980 Bayh-Dole Act had amended existing patent law so that the intellectual property rights of inventions stemming from government funded research were no longer re-assigned to the federal government (Mirowski, 2011). Universities, small businesses, and non-profits, including FFRDC's, were allowed to retain ownership of inventions, and monetize these resources as they saw fit. While, it would have been a contentious move to try to capitalize on the international cooperation that resulted in COADS being assembled- the spirit of entrepreneurship around basic science research was pervasive in the early 1980's (Giddens et al, 2003). The decision of whether or not to create an artificial barrier to entry (i.e. a price or membership model for access to COADS) should be seen through this historical lens - the archive could have either been distributed for free; or NOAA, NCAR, and CIRES could attempt to recoup some expenses incurred in devoting labor and resources to creating COADS.

The UKs Meteorological Office had been simultaneously processing and developing a similar set of historical weather observations, the Main Marine Data Bank (MDB). The MDB was nearly as extensive in geographic coverage as COADS, and included a number of unique datasets that were not present in the initial release of COADS. At the time the UK Met Office was, like many nations in the 1980s, selling these records to recoup the costs of production and service. COADS partners felt the best way to achieve an authoritative record were to combine efforts of the two different projects (Woodruff, personal communication). The UK partners agreed, but asked COADS to purchase the MDB. COADS partners initially rebuffed the offer and decided to focus instead on creating a free and openly accessible archive in the spirit of Maury's "common property of the world" (Parker et al, 2001).

In announcing the initial availability of COADS, early contributors often described this decision on ethical grounds. For instance, an early publication that detailed the first release of COADS stated that the project would consistently attempt to be authorotative, and to further "make this record available to the individual investigator in a form that is reliable and easy to use for anyone interested" (Slutz et al, 1985).

COADS 1 (1979-1887)

The first release of COADS included a date range beginning at the development Maurys standard logbook (1854), up until NOAA began processing data from the set of Tape Deck Families described above (1979). In total, the tapes assembled for the initial release contained more than 71 million unique observations.Release 1 would included 11 different data products and their processing history (Jenne and Woodruff, 1986), a Fortran 77 program on magnetic tape used to read the packed binary-data (Woodruff et al 1986), as well as a number of publications describing the statistical trimming methods used to create derived variables (Slutz et al, 1985; Fletcher, 1983).

COADS 1a, 1b (1992-1999)

COADS Release 1a, and 1b simply update and improve the coverage of interim data products that were developed until 1992. This release also saw the adoption of systematic way to add newly collected data to the existing records - a practice that would continue until 1995. COADS 1a was also the first release to make electronic documentation available for download through an anonymous FTP server. Release 1b improved data quality issues discovered in records from 1950-1979, and included a set of Russian marine logs from Arctic explorations. These interim releases also marked a point of controversy in the early project. Biases discovered in the blended products were numerous, and this led to disagreements about how the records should be updated to reflect known errors. The re-processing of gridded data would be expensive, and time consuming, but establishing COADS as a credible source was one of the foremost early goals of the project.

Review of Governance in Regime 1

The blending different data sources to create an authoritative record of marine surface data was enabled by three factors:

- WMO Resolution 35 (cg XI) and the success of the IGY in 1957-1958 created a more cooperative international exchange of marine data sources.
- The HSST project serves as a proof of concept - a comprehensive, and authorotative historical record of marine weather was not only possible, but would be exceptionally valuable to the climatology community
- Supercomputing facilities, and statistical trimming techniques were available to make such a historical archive readily exchangeable and useful to ongoing research.

Governance Variables

The production of early versions of COADS follows a very traditional self-organized model of governance. The discretionary spending used to fund the initial project, and the appeal to WMO resolutions for governance principles (i.e. free and open exchange, no hierarchy of contribution) contributed to the creation of a resource set that had low-rivalry, and low degrees of excludability - which was in opposition to competitor products which were being sold, excluding a broader climate community from accessing these resources. In short, this regime establishes a marine data commons that is to be informally governed through provision and production.

4.5.3 Regime 2: 1999-2009

Three institutional features are prominent in COADS second governance regime:

1. The project is renamed to ICOADS in recognition of international contributions to release 2.
2. A governing body, The Joint Technical Commission for Oceanography and Marine Meteorology (JCOMM) is established by WMO, further nesting ICOADS work in the international climate and weather data enterprise, and providing a closer link to WMO as an international governing body for meteorological data.
3. A set of meetings, the JCOMM Workshop on Advances in Marine Climatology (CLIMAR) and the Workshop on Advances in the Use of Historical Marine Climate Data (MARCDAT), to exchange results from recent research in marine climatology, and to coordinate future improvements and enhancements to the historical marine climate data record more generally.

When asked to describe this period, a participant characterized this as a

> "...time when other users of ICOADS started collaborating and feeding back on the project. So, people like, big people in the field, started doing a lot of work with ICOADS and using it for various new products that they were doing and feeding all of that back, what they learned, to project PI's - that is really the start of the evolution of ICOADS." [ISC-04]

4.5.4 ICOADS Release 2

In 1997, after incremental improvements to COADS 1, 1a, and 1b were made publicly available planning began for the development of a second release. The attention and enthusiasm that a decade's worth of science had produced for COADS brought with it new sources of data and funding (Woodruff et al, 2001). Producing a new release of COADS included reprocessing, and reanalyzing all existing data, as well as blending new sources discovered in the interim period (since 1979). This is a somewhat unique aspect of the production process in ICOADS new releases; in order to properly compare duplicate records, and integrate new data sources in existing gridding schemes, each data source has to be completely reprocessed, using new algorithms and new quality control benchmarks. This requires a substantial amount of coordination, and dedicated computing resources over what is typically a two to three year process of producing a new release.

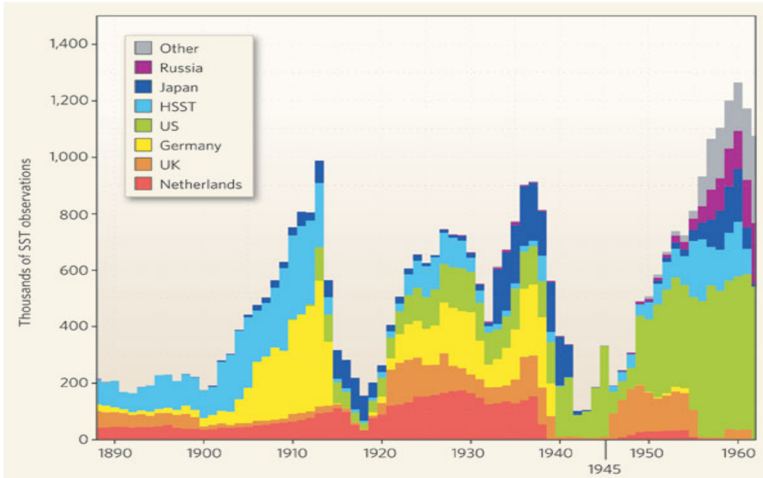
During this period of data processing, developers of COADS and the UK Meteorological Office's MDB negotiated terms upon which the two data products would merge - improving known errors in both datasets as well as moving towards a single, authoritative resource for the climatology community to use. In describing the negotiations around this period, every community member that I interacted with emphasized that this was not a natural progression or a necessarily cooperative one.

One participant explained the state of data access in the following way,

> "For whatever reasons, it is much harder to prohibit data access in the US. The UK, you know the MET, the Marine Data Bank, they have a lot of overhead to make data easy to use. And if they understood it they didn't need to create those extra services. They don't. They couldn't invest in it. When you look back at the literature, pre-COADS. People use whatever. Sometimes those TDFs, but there is nothing very authoritative there. The Germans used their archive. I couldn't use MDB because I didn't work at the Met. Nobody had to share. And they didn't. So why did they merge? They had to. The science wasn't getting better without it" [ISC02]

Although there was widespread agreement that the combination of the two data sources would create a better product, another participant described the process as contested,

> "It was heavily debated...COADS was a NOAA project, and it was US centric- I don't want to say insular, but it was very "Boulder" oriented. And the debate was... that is a lot of money, a lot of resources went into collecting this data through the Voluntary Observation Ship program, and that is supported by individual nations. Not from a pool of funds. Those are investments...So that debate really focused attention on the fact that COADS should be looked at as an International dataset, because you know a lot of other countries were putting money and data into it."



Caption: Records in ICOADS as shown by contributing Nation

A single graphic from Release 2.0 emphasizes the point made by this participant; In the figure below we see a breakdown of the nations that contributed archival data to the project. Looking backwards in time to the early 20th century, almost all of these records come from nations other than the USA and UK. Whats more, the HSST data makes up a bulk of this mid-century data products, and as was described above, that project was an international "proof of concept" on the cooperation required of producing comprehensive marine data products.

The proceedings of the first MARCDAT workshop, where the name change would be voted upon by JCoMM members (described below), records the process as follows,

> "In its final plenary session, the Workshop voted in support of the name International Comprehensive Ocean-Atmosphere Data Set (I-COADS) for the new blended observational database. This name recognizes the multinational input to the database while maintaining continuity of identity with COADS, which has been widely used and referenced." (Diaz et al, 2001).

The very next sentence in the workshop proceedings invokes the history of this enterprise, and is worth including here,

> "The Workshop was an appropriate lead-in to the conferences planned by JCOMM for September 2003 in Brussels, to commemorate the 150th anniversary of the conference convened in

Brussels in 1853 by US Navy Lt. Matthew Fontaine Maury to establish, inter alia ,the standardization of meteorological and oceanographic observations from ships at sea. Maurys work (see Lewis, 1996) remains the foundation of much operational and research maritime meteorology and oceanography"

The process of renaming the project is particularly useful example of the difference between rules-in-form and rules-in-use. From my participants perspective, the process of deciding to rename the project was contentious, but necessary to advance the science being done. And yet, not one of them remembered that the matter was put to a formal vote. In their minds, the decision although contentious politically, was about scientific progress - as my participant said "...the science wasn't getting better without it"

ICOADS Release 2.0 & Interim Products

In September of 2002, a second full release of COADS, and the first to be titled I-COADS, was made publicly available. The data in this release would initially include two archives, one of which was a real-time mode - that included a blending of data coming from the GTS, and the second was a delayed mode, which would apply stricter quality controls to newer data, adding it on to existing historical records (Worley et al, 2003). The delayed mode archive therefore encompassed Releases 1a (1980-97), 1b (1950-79), and 1c (1784-1949), but reprocessed and improved by the addition of new data sources, including the entirety of the Marine Data Bank.

Three additional innovations are important to this release period:

1. Documentation was put on an ICOADS dedicated website (eDocs). This signaled a further move to decentralizing the knowledge of COADS - users could also now submit bugs, or report errors through this website instead of via personal communication with developers. This is a subtle, but important innovation that creates an ICOADS identity, instead of a NOAA or NCAR identity.

2. Interim products would now be labeled like software releases, with a decimal point signifying a new incremental improvement (e.g. 2.1, 2.2, etc.). Interim products during this period occurred on regular intervals,

meant to show responsiveness to a new international community of participants.

3. Release 2.1 in 2003 would adopt the new ASCII IMMA format, an innovation which would standardize the exchange of marine data, and influence policy about data sharing at both the national (NOAA) level, and the international (WMO) level. This further demonstrates the influence of ICOADS as an international authority, and marks a certain reputation amongst international partners as a standards bearer.

Release 2.1 would also see the resolution of the ICOADS gridding scheme improve, with “2 latitude 2 longitude and 1 1 boxes beginning in 1800 and 1960 respectively.” (Worley et al, 2003). While this seems like a minor improvement, the change in resolution enabled easier integration with other environmental data available at the time (Woodruff et al, 2006).

JCoMM

Until 1999 the World Meteorological Organization coordinated data management and observing systems related to meteorology and oceanography through two different governing bodies - the Commission for Marine Meteorology (CMM), and UNESCO’s IOC, jointly with WMO, through the Committee for the Integrated Global Ocean Services System (IGOSS). Citing a need to combine overlapping expertise, the WMO wrote that “While enhancing safety at sea remained the primary objective of marine forecast and warning programmes, requirements for data and services steadily expanded in volume and breadth during the preceding decades. Moreover, many of other applications required observational data sets and prognostic products for both the oceans and the overlying atmosphere.” (2011). To better coordinate the collection and curation of data in these two domains, the Joint Technical Commission for Oceanography and Marine Meteorology (JCOMM) was formed. To reflect the shared responsibility of this commission, it is co-chaired by both a meteorologist and an oceanographer.

The combined domain expertise of JCoMM directly preceded the international relabeling of COADS, which blends oceanographic and atmospheric datapoints. The formalization of the governance of this domain in part helped the ICOADS project partners realize the international scope of their work, especially through the founding of marine climatology workshops described below.

CLIMAR & MARCDAT "Sustaining the International Community"

In parallel with the formation of JCoMM as an international governing body of the WMO, two related workshops were designed to bring together stakeholders of the marine climate data community. Advances in Marine Climatology (CLIMAR) brought together a community of developers and users of surface marine climate data, while the Advances in the Use of Historical Marine Climate Data (MARCDAT) workshop focused specifically on users of historical weather data for climate research. CLIMAR and MARCDAT are important to cooperation within those domain, and often result in concerted efforts to improve known errors, or in collaboration around on-going projects.

For example, through early international coordination at MARDAT, ICOADS project partners agreed to coordinate their work at reducing biases related to Sea Surface Temperature (SST) variables, and to increase (where possible) data coverage in order to contribute to planned international climate assessments. Diaz et al write about this process in the proceedings of the 2002 MARCDAT, "A staged timetable for implementation was agreed: firstly, a two-year period would lead to the third C20C Workshop around April 2004; and, secondly, a period of about five years would lead to the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC)." ICOADS partners successfully met these targets, and the contribution to AR4 was noted by lead authors of Working Group 1, who used these SST measurements extensively (AR4, 2007).

Review of Governance in Regime 2

ICOADS second regime developed an institutional capacity to collaborate more broadly, and to systematize a release schedule. During this period, the formalization of an international governing body for meteorology and oceanography (JCoMM) had a nesting effect for ICOADS own governance; it placed organizational pressure on ICOADS to conform to certain models of distribution promulgated by the WMO, but JCoMM also took on the responsibility of coordinating stakeholders through the MARCDAT and CLIMAR workshops. Additionally, the founding of JCoMM brought with it the opportunity to promote ICOADS own standardization work, such as the IMMA format for exchanging marine data, and to gather consensus on future directions of the archive so as to best meet broader climate assesment goals.

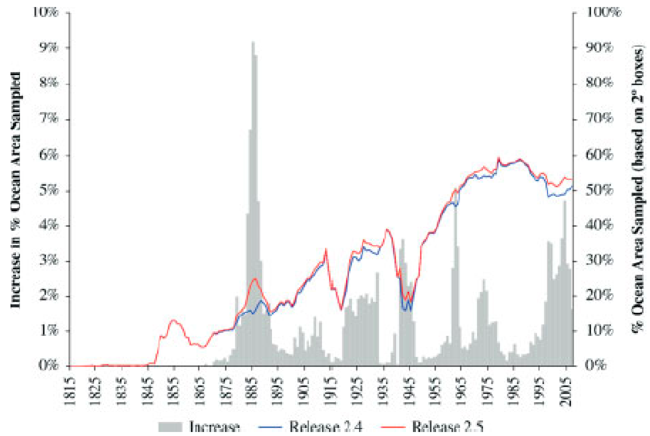
This regime marks a clear division from an initial COADS governance which was "self-organized " to a "polycentric"model which nests rule making and enforcement at different levels, and creates jurisdiction, or institutional overlap in promulgating new rules.

4.5.5 Regime 3: 2009- Present

The defining features of ICOADS third (on-going) regime is the release of a valuable interim product, ICOADS 2.5, and the continued formalization of its governance body. In early 2012, funding was partially eliminated for critical members of the ICOADS team at NOAA. This caused disruptions for planned enhancements to the archive, but overall the focus of the ICOADS community remains on: 1. Coordination of ICOADS release 3.0 and the ICOADS Value Added Database (IVAD) which brings with it a new contribution mechanism for ICOADS stakeholders; and 2. Making a formal application to the WMO for ICOADS to be recognized as a Center for Marine-Meteorological and Oceanapgraphic Climate Data (CMO).

ICOADS 2.5

In late 2009 data processing began for an interim ICOADS release 2.5. This product would further enhance real-time access to end users, substantially reduce the burden of sub-setting ICOADS, and included a number of improved bias adjustments. Previously released interim products (2.1 -2.4) had incorporated some new data, but a substantial amount of new digitized products from international partners was now available, including newly digitized data from the RECovery of Logbooks And International Marine data (RECLAIM). A journal article describing this new release indicates the further international cooperation in this process, "Data provision, collation, and distribution remain the responsibility of the founding partners, but other countries and international organisations including the Joint World Meteorological Organisation (WMO)Intergovernmental Oceanographic Commis-



Caption: From Woodruff et al. 2011 "Annual percentage ocean area sampled for SST for R2.5 (red curve) compared to R2.4 (blue curve) (right axis). Annual percentage increase in global ocean area sampled for R2.5, compared to R2.4 (bars, left axis)."

sion (IOC) Technical Commission for Oceanography and Marine Meteorology (JCOMM) now make noteworthy contributions." (Woodruff et al. 2011). As a result, release 2.5 made substantive improvements in the coverage and completeness of data sampling over the ocean as shown in the figure below.

ICOADS role in Climate Reanalysis

Climate reanalysis are relatively recent approach to data assimilation in earth systems science. The major goal of a reanalysis project is to give a numerical account of the very recent past (10, 20, or 40 year time steps) using a combination of archived climate and (very recent) weather data with whole earth models. (Dee et al, 2015) Because the "inputs" for reanalysis assimilations require broad coverage (both space and time), ICOADS has been used regularly in reanalysis projects that include an ocean component (e.g. Trenberth et al. 2001). Reanalysis datasets are some of the most important, and valuable resources that the climate science community has produced in the last 20 years (Edwards, 2011 p. 323-8).

One participant described her work on input datasets for reanalysis projects in the following way,

> "My career would have been a lot different if I had focuses on the exploitation of the data rather than the improvement of the

> data. And that is why citations are important. If you start out with my metadata publications, it gets tens of citations, ICOADS papers get hundreds, but the gridded datasets, and especially the reanalysis stuff those get tens of thousands of publications. And you do look upwards, and think tens of thousands of citations?"

This is to note that, the financial value of reanalysis datasets have been enormous, they've also paid dividends to those with their names attached. So, while individual scientists working to improve parts of ICOADS records fail to garner much attention, the project as a whole has been recognized throughout the climate community as a high-quality resource partially as a results of its continued use in climate reanalysis projects. This is an important outcome for the project, as developers of ICOADS 2.5 note in their description of enhancements to the archive,

> "...NOAA, the European Center for Medium-Range Weather Forecasts (ECMWF), the Japan Meteorological Agency (JMA), and the US National Aeronautics and Space Administration (NASA) have all taken advantage of ICOADS for their reanalysis. Communication between reanalysis centers and the ICOADS developers is excellent. Each new reanalysis is based on the most recent ICOADS update, thereby taking advantage of any improvements in data quality or quantity based on the efforts of marine data experts. ICOADS therefore reduces efforts required for data preparation and quality control for reanalysis projects. In a complimentary manner, reanalysis efforts (e.g. Compo et al., 2006) also uncover data problems in ICOADS that feedback to the developers and lead to future improvements." (Woodruff et al, 2011).

Defunding of NOAA Partners

Amid political turmoil around federal budget allocations for FY 2013, NOAA announced in the winter of 2012 that it would eliminate future support of ICOADS through the Climate Program Office (CPO). A public announcement made in February of that year reads:

> "For budgetary reasons, stemming from pending large cuts at the NOAA Climate Program Office (CPO), ESRL Directors have determined that it is no longer feasible for its Physical Science Division (PSD) to continue supporting any further ICOADS work effective immediately."(NOAA, 2012)

Even though ICOADS acts as an input to many downstream data products, it is viewed as a research project by federal agencies. In an extended conversation about this decision, a long-time ICOADS project partner explained this reasoning:

> "The number one thing to recognize from that period is, one of the drawbacks for most soft-money produced data products, whether its ICOADS or something similar, these are not sanctioned data products like the National Weather Service, these are done under research funding and most people think of research funding being time limited, and often times the folks that develop these products are not the best marketers, and they're not very adept at or even equipped to trace how they are used..." [ISC-04]

In response to this decision, ICOADS project partners were able to secure letters of support from prominent international partners. The initial impact of this defunding has been minimal to end-users who have experienced little disruption on their access of ICOADS, but this controversy greatly delayed work on a third release, and stunted the development of a new valued added database (IVAD).

When asked about gathering support from the international community, a long-time project partner offered the following description:

> "I think what happened was a funding cut came down from NOAA...and whoever was making that decision did not understand how important ICOADS is to the research community. How widely used it is ... because there was no statistics, there is no metrics of how many people would all of the sudden have their data products disrupted if ICOADS goes away, and that resulted in a very quick and fairly effective outpouring of letters of support, from major agencies around the world that used ICOADS...There

were directors of major research institutes and operational institutions around the world that said 'Hey, we need this product continued"' [ISC06]

In the late fall of 2012 NOAA agreed to restore partial funding of ICOADS, but shifted financial support from the Earth Systems Laboratory (ESRL) to the National Climate Data Center (NCDC). This event had four substantive impacts on ICOADS, and its governance model:

1. After the Winter of 2012 defunding announcement, project partners immediately began negotiating the terms of a "Letter of Intent (LOI) to Enhance Support for the International Comprehensive Ocean-Atmosphere Data Set (ICOADS) Program through International Partnership." The letter of intent is a non-binding document, similar to a memorandum of understanding, which has no legally enforceable obligations, but instead spells out the terms of cooperation between project partners in order to, "more formally recognize existing and planned international contributions that build on the ongoing investment in ICOADS." Signatories include: the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC), Cooperative Institute for Research in Environmental Sciences (CIRES) University of Colorado, National Center for Atmospheric Research (NCAR), Deutscher Wetterdienst (DWD), Center for Earth System Research and Sustainability - University of Hamburg, UK Met Office, Climatic Research Unit (CRU) - University of East Anglia (UEA), and the UK National Oceanography Centre (NOC).

2. An international steering committee (ISC) was formed in order to establish ICOADS future goals, including " new tasks and their prioritization; cooperation on identifying partner-specific experts for each task; and facilitation of expanded international cooperation." The ISC includes a representative from each of the signing institutions of the letter of intent. The steering committee is to have the following structure: led by a chair on a rotating basis between signatories, each member will hold a formal voting right (how voting is conducted, who calls a vote, etc. is not described), and institutional entrance or exit from the committee will require a 30 day notice to other members.

3. The letter also formalizes roles that each project partner should take on. These roles are defined in agreement with a set of ICOADS project goals, which include "cooperation and collaboration on the rescue, digitization, assembly, processing, quality control (QC), and archival of surface marine meteorological and oceanographic data, for the express purpose of making such data openly and freely available. This availability will be accomplished primarily through the joint completion and issuance of major new ICOADS Releases, and intermediate products, together with their accompanying technical documentation and journal publications." (2013) These activities could describe each ICOADS regime, but the importance of these formally stated goals is that individual partners are assigned specific tasks in completing or contributing to these development efforts.

4. The move from ESRL to NCDC greatly impacted the dynamic of the ICOADS's leadership. The longtime PI of the project, based at ESRL, moved to part-time, and has since announced he will retire in Spring 2015. Additionally, three key resources - a full time programmer, remote-access systems for NetCDF users, and a set of virtual servers - were not replaced in the shift of responsibilities from ESRL to NCDC. All of this greatly impacted the speed at which work on a third release could proceed.

CMOC

ICOADS partners, after signing the letter of intent, were encouraged by JCoMMto seek out status as a Centre for Marine-Meteorological and Oceanographic Climate Data (CMOCs) from the WMO. As a CMOC, ICOADS could be recognized formally by JCoMM and the WMO as an essential research infrastructure for the weather and climate data enterprise. This would encourage a greater number of international partners to formally sign ICOADS letter of intent for cooperation, and allow project partners to more easily appeal to national funding agencies for financial support to contribute to ICOADS development. In my interviews with ISC board members, status as a CMOC was widely endorsed but often came with caveats of the drawback to further entrenching the ICOADS in the bureaucracy of the WMO. A participant explained the tradeoffs in applying for CMOC status as follows:

> "...one of the drawbacks to getting quote unquote recognized by JCoMM and WMO is they have a very heavy bureaucracy, they have a very heavy bureaucratic structure because JCoMM and WMO are both a part of the United Nations... the diplomatic structures there, it takes a long time to get things done... a single document can take years to get approved. One of the pluses, at least this is the way we see it, if you are recognized by this international body as being an important part of the either the climate or weather infrastructure, that data infrastructure gives you leverage in your individual countries to get dedicated resources." [ISC-01]

Another drawback is that the WMO's Vision for a Marine Climate Data System (MCDS) is relatively new, and there is little administrative support for the initiative within JCoMM. As such, ICOADS would be the first CMOC recognized by the WMO. Many participants expressed concern that it is unclear what responsibilities the status will hold within the WMO, and what kinds of resources will be available to CMOCs for impacting broader WMO policy.

ICOADS Release 3.0

The letter of intent signed by project partners emphasizes the following point, "By signing, partners agree to work closely with the common goal of enhancing and internationalizing ICOADS, including the near-term goal of completing the next major delayed-mode update, Release 3.0."

ICOADS partners had been planning the next full data release since 2011 as a result of new sources of data had became available through the Atmospheric Circulation Reconstructions over the Earth (ACRE) project, and dedicated supercomputing time at NCAR would become feasible for records processing. The latter is the result of a new supercomputing center (Yellowstone) coming online, which in-turn would free up existing high-performance machines. The defunding of ICOADS came at an inopportune time as developers at ESRL were planning to improve IMMA software for formatting ICOADS 3 data. Work on blending contemporary GTS data continued, but the delay in a new software package for IMMA formatting greatly delayed the re-processing of older records. Originally scheduled for completion in December of 2014, ICOADS partners have now targeted early 2016 as a release date.

Release 3 will include a number of major improvements, but there are two important innovations to note:

1. For the first time, unique report identifiers (UID) system will be used to assign ids to individual records in ICOADS historical archive. This means that all 295 million records would be serialized, and will be more easily accessed by end users. Further, the scheme adopted for UIDs is also in use by other community projects (upper-atmosphere datasets), and is intended to greatly reduce the burden of integrating ICOADS in future reanalysis projects (Freeman et al, 2014).

2. In 2011, the design of a new product called the ICOADS Value-Added Database (IVAD) was funded by NOAAs Climate Observation and Monitoring Division. The concept is a novel design for aggregating different bias corrections, and improvements in the ICOADS archive so that users can "select" which corrections they want at a point of download. IVAD is an important scientific innovation, but this product also changes the nature of ICOADS division of labor in producing new resources.

   In the IVAD model, end-users will be able to follow an established protocol to submit bias corrections, and volunteer to do peer review for submitted corrections. Eventually these new datasets will be integrated into the existing archive such that an end user, for instance, looking for SST could pose a query, and get bias corrections with different features, select which features were important to her, and then download original and processed records side-by-side. This moves ICOADS from a real-time "develop and release"dataset, to a peer-produced resource that improves incrementally with direct user contributions.

   > "this process... I mean, this is very unique - this is a concept that has not been implemented for the in-situ observing systems. Most of the datasets out there - the data quality

> is asssesed by a single organization. And they put the quality flags on data points, they decide whats a good value and whats a bad value - not always very clear either. There is not a whole lot of feedback loop on these products." [ISC 03]

IVAD should sound similar to open-source software projects, as I was told a number of time that this was partially the inspiration for having contributors feedback their "patches" to the project. In some sense, IVAD will turn historical records of ICOADS releases from a primary resource to a "kernel" - one which can be developed around and on top of. Like the governance of the second regime, the impact of IVAD on the ICOADS project is that of nesting one system inside another. In this case, ICOADS will further establish itself as a platform on which others develop and innovate with applications that make it more accessible, and easier to use in different contexts, by different users. I return to the larger implications of this nesting process in the following chapter.

Review of Regime 3

Regime three includes the very recent past of ICOADS governance, and its on-going work to develop an innovative new release. Although much shorter than the two previous regimes, this period has two significant implications for an evolving ICOADS governance model:

1. A letter of intent, signed by seven international project partners, establishes a number of formal governance mechanisms, including policy instruments and positional rules.

2. The design and integration of IVAD will dramatically change the production process of ICOADS, and in turn will create a new set of benefits and costs for the governance model to balance.

4.6 Summary

In this chapter I have summarized the differences between the IAD and KFC, and justified my innovation with both of these frameworks to present an analysis of the ICOADS case study data. I then provided background information important to understanding the context in which ICOADS was developed, including a number of formal policies at the international level, and failed legislation at the national level. I then offered an overview of ICOADS various resource sets, and community member roles. Finally, I justified the division of the ICOADS case study into three separate governance regimes, offered an analysis of how rules, and governance variables were effected by the "action arenas" of each regime - including my participants (ICOADS Steering Committee members) own perspectives about these events. In the next Chapter I summarize the major findings of this analysis and answer the research questions formally stated in Chapter 2.

CHAPTER 5

5.1 Introduction

In this chapter I summarize the findings of the ICOADS case study, and answer three case study specific research questions. I then review three previously completed case studies using the Knowledge Commons Framework, and compare and contrast findings from all four cases. I use this cross-case comparison to answer the overall research questions of this dissertation.

5.2 Case Study: ICOADS

The case study data organized by the Knowledge Commons Framework in Chapter 4 included the following categories: Background Environment, Community Attributes, Governance and Rules-in-Use. The Governance sections were divided into three "action arenas," each consisting of a set of governance outcomes. In the following sections, I review this data in the form of coded state variables for each regime. State variables characterize a system at a particular period of time (Walker et al., 2005). When formally defined, they can guide the work of comparing between system states, allowing for an analysis that is both reproducible by other researchers and reusable in new contexts (Cox et al., 2014). The process of developing a coding scheme for these variables took the following steps:

1. Identifying relevant variables from the Socioecological Systems Meta-analysis Database (SESMAD) project (Cox et al., 2014) for a sociotechnical system.

2. Identifying relevant literature to generate variable definitions.

3. Coding each regime based on relevant data from the ICOADS case study.

4. Reviewing the codings with my key informant for the sake of external validity

5. Iteratively editing definitions for consistency .

The outcome of this work is summarized in the table below:

Variable Name	**Regime 1**	**Regime 2**	**Regime 3**
Begin date	1981	2001	2011
End Date	2000	2010	Present
Description	In Text	In Text	In Text
Multiple levels	Single-Level Governance	International Governance	International Governance
Institutional diversity	Low	Medium	Medium
Centralization	4 (Highly Centralized)	3 (Somewhat Centralized)	2 (Somewhat Decentralized)
Governance scale	State-based	International	International
Scale match	No-Match	Match	Match
Selective Pressures	Competition, Resource Availability	Competition, Technological Innovation	Obselescence, Resource Availability
Articulation	Medium	Low	Low
Capital	High	High	High
Coordination	Low	High	High
Governance trigger	Slow Continuous Change	Slow Continuous Change	Sudden Disturbance
Diagnostic Capacity	Neither	Treating proximate causes (symptoms)	Treating underlying drivers
Autonomy	High- complete autonomy	Moderate- some autonomy	Moderate - some autonomy
Metric diversity	Low- no metrics for success	Medium- some metrics for success	Medium- some metrics for success
Rules-In-Use	Position, Authority	Information, Payoff	Boundary, Aggregation, Scope
Policy instrument	Incentive Based Insturments	Incentive Based Instruments	Input and Output Based Standards

Table 5.1: Coded State Variables from the each regime of the ICAODS case study.

Next, I summarize observations from the coded state variables. Formal definitions are offered before each explanation of how the variable was coded for the ICOADS case study[1] Just as the previous chapter, I occasionally draw upon interview data from “interactionist” studies. Where this is the case, I distinguish between participants with the following convention - ICOADS Steering Committee (ISC) and randomly assigned number (e.g. [ISC-06] Participants are not identifiable through this coding or the selection of interviews used in this chapter.

5.2.1 Discussion of Coded State Variables

Institutional diversity

[1]The properties that correspond with each state variable can be found at http://git.io/vJnPX.

Definition: "Ostrom (2005) has argued that institutional diversity is important for the same reason that biological diversity is important: that different institutional arrangements are frequently a response to local conditions and thus a diversity of arrangements are needed in order to adapt to a diversity of conditions." (Cox et al., 2014) This is particularly true of conditions in a sociotechnical system which couple diverse technologies and people. A diversity of arrangements used to sustain a sociotechnical system can also be seen as a way to avoid path dependence

From Regime 1 to 3 ICOADS went from having low institutional diversity to a medium. Low institutional diversity implies that that a homogenous set of institutional arrangements were applied to a diverse set of resources. Originally, COADS was governed by just two institutions (both federally supported research laboratories). Over time, this resulted in social conflict which led to the introduction of new institutional arrangements - including ICOADS, JCoMM, and the ICOADS Steering Committee - all of which can be seen as evidence of the project becoming more heterogeneous in its governing structure.

Centralization

Definition: A centralized governance system has few actors/actor groups that hold a disproportionate amount of authority of over actors or parts of a commons. More decentralized governance systems have flatter hierarchies.

Similarly, ICOADS governance became less centralized over time - Regime 1 was "highly centralized", and Regime 2 and 3 became a "somewhat decentralized" as it continued to decentralize its power and decision making authority. This is in line with institutional arrangements becoming more diverse, as described above, and the scale of the governance model shifting from state to international, as discussed next.

Governance Scale

Definition: This variable defines the scale at which a governance system operates, including:

- *International regime: more than one country is involved;*
- *State-based (national) policy: at the level of one country;*
- *Sub-national policy: within a country (e.g., states, provinces, regions);*
- *Local: a finer scale than subnational (e.g., community-level, several laboratories, etc.)*

The governance scale variable is meant to code what jurisdiction the governance system has authority over. For instance, can the governance model apply rules, sanctions, bans, or assign tasks at state, national, or international levels? ICOADS moved from a state-based national scale to an international scale; first with granting partners voting rights in the renaming of the project (in 2001), and then further formalizing the internationalization of the project with the signing of a letter of intent to cooperate (in 2013).

Scale Match

Definition: Scale match explains the relationship between a governance scale, and a resource scale. For instance, spatial mismatches occur when the spatial scales of management and the spatial scales of a knowledge commons do not align appropriately. For example, if a local community is attempting to manage an infectious disease reagent, there is a mismatch in the scale of management and the technologies range of use.

The scale of a governance can be coded, within different states of a system, as having a binary "match" or "no-match." In ICOADS first regime, a set of internationally archived records are used to produce resources that are provisioned with a state-based governance model. This sets up a tension where the broader ICOADS community felt there was not a match between the source of the inputs (international weather records) and the governance of the outputs (a state-based model). Regime 1 therefore does not provide a match between stakeholders and governance. In Regime 2 and 3, governance expands through the formation of JCoMM and this results in an international governance structure that now matches the scale of the resources and their users, producers, and provisioners.

Selective Pressures

Definition: Selective pressures are understood to cause governance adaptation. As Smith, Stirling, and Berkhout write, "Without at least some form of internal or external pressure...it is unlikely that substantive change to the developmental trajectory of a regime will result. "(2005)

The types of pressures exerted on ICOADS governance relate to its shifting between regimes. These variables are meant to define and provide context for those pressures, but these categorizations should not be read as directly causing a regime shift. They are one of many stresses, disturbances, and motivations for a regime shit.

For example, the first shift between regime 1 and 2 is a result of "competition" and "resource availability" pressures. Competition for users, and for satisfying the need for a comprehensive marine dataset influenced ICOADS leadership to formally recognize the contribution of its international partners, as well as merge with its chief competitor, the Marine Data Bank. By expanding the scope of legitimate stakeholders for producing and provisioning the commons, ICOADS governance also necessarily had to become less "Centralized", increase its "Institutional Diversity", and adjust its "Governance Scale". These three variables (Centralization, Institutional Diversity, and Governance Scale) correlate strongly with "Selective Pressures" when a regime shift occurs.

Selective pressures also have three sub-variables which are meant to further categorize responses from the commons governance.

Sub-variables of Selective Pressures

Articulation, Capital, and Coordination

- Articulation explains the extent to which a governance model is capable of recognizing, describing, or defining selective pressures.
- Capital, both social and technical, explains the extent to which a governance model can access and extract resources from a network of stakeholders.

- Coordination explains the extent to which a governance model is able to organize a response to selective pressures

In Regime 1, ICOADS stakeholders had a partial (medium) ability to articulate problems that were introduced by "competition" and "resource availability". This regime also had high social and technical capital , as was evidenced by the allotment of supercomputing time to develop COADS 1, and the ability to coordinate trading of archived marine data amongst international partners. The coordination of these partners, beyond simply contributing data, was somewhat low - as evidenced by early literature that described the project as a "national" effort (e.g. Slutz et al., 1983)

In Regime 2, selective pressures include "technological change", seen through innovations in the assembly of reanalysis data products, and "competition for resources", which were due to the proliferation of high quality subsets of ICOADS. Articulation of these problems remained low; the number of sub-sets were so numerous and so diffuse that they could no longer be included in ICOADS documentation. Both social and technical capital, as well as coordination within the second regime were high as ICOADS was able to coordinate development for integration with reanalysis projects, and position the resources to be used in major international climate assessment reports.

In Regime 3, the most current regime, ICOADS suffered initially from the selective pressure"resource availability," but seems to have responded to some of these problems by coordinating international partners for support. However, ICOADS continues to face selective pressure to innovate, which it has not been able to respond to, and thus risks being branded "obsolete" in relation to already existing high quality subsets and reanalysis products.

In the overall state of regime 3, articulation was low. This is evidenced by the fact that the ESRL defunding seemed to take even most senior NOAA project partners by surprise. But, this also demonstrated that levels of social capital were high in the third regime, as NOAA consequently acknowledged that restoration of partial funding for ICOADS was due to an outpouring of support from the international community. Technical capital was low, as the interruption of work on IMMA software greatly delayed progress on an anticipated and improved ICOADS third release[2]. Coordination was high in

[2]This is an admitted flaw to the way that the "Capital" variable is currently coded:

Regime 3, as demonstrated by the ability of ICOADS to respond with letter of intent to cooperate, and the formal assignment of tasks to project partners that agreed to work towards a third release.

Selective pressures, along with governance triggers and diagnostics are all highly related variables and should have a consistent relationship, which I describe in the next two sections.

Governance Triggers

Definition: This variable is meant to provide information about the accumulation of selective pressures. Governance triggers are characterized by the notion of disturbance from socioecological systems management. A disturbance is some event, or set of selective pressures which cause a change in state variables. Governance triggers can be either "sudden disturbances", such as a defunding event from NOAA that shifts state variables; or, a governance trigger can consist of "slow continuous change", such as the obsolescence of a technology over time as ICOADS is currently experiencing in its third regime.

Regime 1 and 2 of ICOADS can be characterized as having slow continuous changes. Disturbances during these periods are related to organizing, creating rules, and fostering cooperation amongst diverse partners. Even during periods of scientific controversy, such as the discovery of anomalies in sea-surface temperature measurements in a historical period of the 1940-1960s (Rayner et al., 2005), there was a relatively stable ICOADS response and a cooperative effort to improve these errors (Woodruff et al., 2011).

Regime 3, which is marked by an unanticipated defunding event, is the first time that sudden disturbance causes a shift in state variables. The ability to diagnose this problem, and coordinate responses to this event can be seen as important characteristics of a sustainable knowledge commons.

Diagnostics

Definition: Characterization of the scale at which a regime responds to a selective pressure, or disturbance. A diagnostic capacity can be to treat un-

When social and technical capital do not align the results are a compromised "medium" coding, when in fact it should be "high" and "low" *derlying drivers of a distrubance, treat the symptoms of a driver, or neither.*

The diagnostic variable attempts to categorize the response of a regime to selective pressures. This variable is related to governance scale and scale match, as well as the selective pressures sub-variables:

- A governance with a strong scale match and ability to articulate, draw on capital, and coordiante responses will be able to successfully transition between regimes.
- Without a scale match, or effective articulation of selective pressures, the governance regime will lack the ability to make a diagnosis of why a selective pressure is occurring. Subsequently, the governance regime may respond to only the symptoms of the problem, and not address issues related to the underlying causes of that problem.

Regime 1's governance had no scale match, a medium level of articulation, and low coordination to respond to selective pressures. The governance model failed to respond to issues of competition and resource availability; and, it neither treated underlying causes of the selective pressures, nor its symptoms until a regime shift was required to do so. The ability to shift governance models smoothly, was partially a result of the high social and technical capital that

Regime 2 had a match between governance and resource sets, but a low ability to articulate the underlying causes of its selective pressures. Failing to recognize underlying problems related to measurement and evaluation of the project resulted in a sudden disturbance event that nearly ended the project. In treating the symptoms related to selective pressures like technological change and competition for resources, Regime 2 was able to coordinate a new method of scheduling releases, a new method of coordinating work (CLIMAR and MARCDAT), and begin development of a new database model through IVAD.

Regime 3 of ICOADS finally begins to treat underlying drivers of its disturbances, and social dilemmas - notably by formalizing commitments from international partners, and beginning to diversify the metrics that it optimizes its work for (e.g. Weber et al., 2011)

Metric Diversity

Definition: The range of evaluative criteria that a project optimizes for, or is judged upon.

All three regimes suffer from evaluation problems that stem, at least in part, from a lack of metrics that can clearly communicate ICOADS' impact to organizations that bear the costs of provisioning a shared set of resources. In discussing the first and second regime, metrics were rarely mentioned by my participants, nor in the literature of ICOADS. More broadly, there was no recognition throughout regimes one and two that an underlying problem in sustaining ICOADS was a lack of formative evaluation. Metric diversity, in this case, is strongly tied to diagnostics; an indicator of healthy or robust knowledge commons may indeed be a diversity of metrics for evaluating success, attainment of goals, or more broadly conveying the importance of resources to a broad set of stakeholders.

Policy Instruments

Definition: A taxonomy of the basic types of policies and institutions that a governance system uses in order to effect actor behavior and achieve commons outcomes. These include: outcome, input, and usage-based standards; incentive-based instruments; and, information and insurance provisions.

Policy instruments can be broadly categorized as either regulatory or incentive based. Regulatory policy instruments are aimed at effecting behavior of commons stakeholders through coercion or punishment. Incentive-based policy instruments are meant to elicit desired behavior through reward. In ICOADS, incentive-based instruments were used throughout all three regimes. Regime 1 and 2 relied exclusively on incentive-based policy instruments, as these regimes had little authority in the way of sanctions or enforcement of rules to establish any kind of regulatory policy about what kind of actions were, or were not accepted. By formalizing partner responsibilities in regime 3, including those related to the production of a third release, ICOADS governance has adopted some minor "input and output standards". In general, the lack of diverse tool uses has been problematic for

ICOADS governance, which has not

Rules-in-Use

To reiterate from previous chapters, rules here are understood to "specify the values of the working components of an action situation; each rule has emerged as the outcome of interactions in an adjacent action situation at a different level of analysis or arena of choice" (Ostrom et al., 1994, p. 4142) Rule types inlcud: Position, Boundary, Authority, Scope, Aggregation, Information and Payoff. In the following section, I describe the rules-in-use for each of the three ICOADS governance regimes.

	Position	**Boundary**	**Authority**	**Scope**	**Aggregation**	**Information**	**Payoffs**
Regime 1							
Regime 2							
Regime 3							

Table 5.2: Accumulation of rules across three regimes of ICOADS (1983-2015)

Regime 1

Rules in Regime 1 can generally be classified as fitting two types:

- **Position Rules** specify a set of positions that actors can play. Each position has a unique combination of resources, opportunities, preferences, and responsibilities.
- **Authority Rules** specify which set of actions is assigned to which position. (McGinnis, 2011)

In COADS initial governance - which was ad-hoc, and loosely coordinated - position rules are established by data provisioners whom assigned responsibilities to one another, and to end-users. Users were free to obtain data, and encouraged to report errors and participate in improving the archive (Woodruff et al., 1987). Data producers who contributed statistical techniques, algorithms, software, or data processing were assigned rights and responsibilities to stewardship of these resources, including any version of the release one that they obtained archived as part of this work.

In an interview about this early process, one participant characterized the early coordination in the following way: "...the problem is that, ICOADS has a single point of failure. I mean (PI) was ICOADS - he was the identity. And while we didn't have a communication problem through language or responsiveness, we bumped up against competing interests - it was a much bigger project and scope than even (PI) realized when we first started putting data out...before the CLIMAR meetings started." [ISC- 06] This demonstrates the centralization of power during this regime, and as such, requires each new regime to revisit authority rules.

Regime 2

The second regime introduces three new rule types to ICOADS governance.

- **Information rules** specify channels of communication among actors and what information should, must, or must not be shared.
- **Payoff Rules** specify how benefits and costs are required, permitted, or forbidden to stakeholders.
- **Scope Rules** specify a set of outcomes, such as what should be produced, what levels of maintenance a resource set requires. Scope rules are closely related to payoff and information rules. (McGinnis, 2011)

Information rules developed during this regime are seen in the specifications of formatting data for exchange, and recommendations from the second CLIMAR meeting in 2004 which reads, "There were seen to be shortcomings in the access to ICOADS data. There are many, overlapping sources of data and products, and the problem of optimising data provision is complex. Many users are working with outdated versions of COADS. Often data are available, but it is difficult for the uninitiated to discover what is there. There should be a Web-based route map to the best available data which should be widely advertised to all the various user communities" (Parker et al., 2004). The "route map" took the form of website developed to host specification documents, and link to relevant existing WMO standards (Woodruff et al., 2006). The process of creating this resource moved to formalize what information should be available, when, and to whom[3]

Payoff rules are established in granting international partners formal recognition through a name change. Recall that when asked about the process, one of my participants replied "...that debate really focused attention on the fact that COADS should be looked at as an International dataset, because you know a lot of other countries were putting money and data into it." [ISC-01] The settlement of the name change debate brought with it a larger set of responsibilities for international project partners, and their engagement in provisioning and producing new interim products throughout the regime.

Scope rules are another artifact of the CLIMAR and MARDAT meetings, where strategic directions are set, and future goals are established. Establishing scope rules also represents a significant shift in the governance model, as it becomes increasingly less centralized.

Regime 3

Two new types of rules are added, and one rule type is further developed in this regime:

- **Aggregation rules** (such as majority or unanimity rules) specify how the decisions of actors at a node are to be mapped to intermediate or final outcomes.
- **Boundary rules** specify how participants leave or enter positions of authority. (McGinnis, 2011)

Aggregation and boundary rules are established by mechanisms introduced by the letter of intent to cooperate. When asked about how these rules are put into practice, a participant described this situation as follows, "The governance is still in flux- the committee are new, the terms of references are still new, and we still have to formalize a bit more how members serve, how long do they come and go, etc." [ISC 04]

[3] The development of a rule type doesn't imply that a social dilemma has been solved. In this case, blending of different data products and an overall growth of the archive leads to unmanageable complexity for end users. Introducing informational rules is a way to treat symptoms of a larger problem in managing a sustainable growth of the project, but rules aren't themselves symptoms of a solution.

In March of 2015 when a discussion about the appoint a permanent chair for the steering committee was raised, the interim chair reminded the steering committee that he was retiring. and expressed no desire to remain in the role during retirement. Tellingly, no one understood nor knew the procedure to make a nomination for a new chair, or whether on-going questions about future support of ICOADS at NOAA should impact this choice[4].

Additionally, the *Authority rules* of previous regimes are both formalized and enacted in new ways in the third regime. This includes formally dividing responsibility for the future provisioning of ICOADS resources, and the creation of a set of core institutions - beyond the original USA partners of NOAA, NCAR, and CIRES - that formally acknowledge their intention to sustain ICOADS over a five year period (beginning, December 2013).

Summary of Rules and State Variables

The systematic approach to characterizing state variables shows a number of important relationships between these variables and their impact on governance.

- Governance scale, scale match, and the sub-variable articulation all correlated strongly with governance triggers. These are not causes for, but contexts of state changes such as a regime shift.

- Selective pressures, and the sub-variables articulation, capital, and coordination are illustrative of both the ways in which governance states experience, and react to disturbances. In ICOADS case, these variables are strongly related to the rules-in-use, and policy instruments used by each regime.

- Diagnostic capabilities help explain the effectiveness of a governance model in understanding the underlying causes of governance triggers, and in turn, provide helpful context for understanding the impact of selective pressures.

- The analysis of rules-in-use shows how rule types accumulate over time, and that as a governance system is formalized, more bureaucracy is added to specify not just who holds power or authority, but how that

[4] As it stands, no decision has been made, and existing by-laws offer little guidance on how to proceed.

power is transfered (boundary rules), and how responsibilities assigned by an authority are directly mapped to outcomes (aggregation rules).

- I caution against judging the addition of new rules, and the expansion of existing rule types (e.g. authority rules in Regime 3) as either positive or negative evolution of the governance system. The more important variable related to sustainability is the match between governance, stakeholders, and resource sets.

Drawing on the data analysis above, I now answer the formally stated case study research questions. In each section, I re-state the research question, offer a summary of the findings, and then explain this answer with a supporting narrative.

5.2.2 Case Study Research Question 1

What are the different governance models that ICOADS has effectively used to manage shared resources over its thirty year existence ?

Summary
ICOADS has moved from a monocentric governance model, with a single center of authority, to a polycentric model, which nests rule making at various institutional levels and scales. The polycentric model has a comparable effect as the "comedy of the commons" where public goods are efficiently provisioned by encouraging actors to sustain a resource set through co-production of its resources. However, polycentric models are not a panacea. The nesting of authority at different levels may create interdependencies among institutional networks that prove costly in the long-term.

Supporting narrative
ICOADS governance evolved steadily over its thirty year period, with only three definitive shifts in models. Each shift is marked, as a I described in Chapter 4, by the need to produce new releases, but also by unique forms of selective pressures, disturbances, and requisite responses.

ICOADS first regime used what appears to be a mono-centric system of governance. This model assigned rules and coordinated actions between only the partners that *produced* COADS release 1; namely NOAA, NCAR, and CIRES. A monocentric system can be defined as a "... single decision structure that has an ultimate monopoly over the legitimate exercise of coercive capabilities." (Ostrom 1972 as quoted in McGinnis 1999, p. 5556) This is evidenced by the types of rules that are established; Position and authority rules are each aimed at establishing a basis of power: Who is in charge? What responsibilities do those positions then hold for the governance of the resource set? In regime 1 the power of authority was not distributed evenly, the lack of governance and resource scale match, and the policy instruments (which were incentive-based) each led to the monocentric governance model failing.

ICOADS second regime was marked by the introduction of a parent organization, JCoMM, and a set of formal workshops, MARCDAT and CLIMAR. In effect, these semi-formalisms shifted the model from monocentric to self-governance, which is defined through the "capacity of communities to organize themselves so they can actively participate in all (or at least the most important) decision processes relating to their own governance." (McGinnis, 2011) This is evidenced by the monocentric governance model being, at least symbolically, voted upon by stakeholders at CLIMAR 1 (1999). Introducing a new name and a new system of releasing data, communicating with end-users, and coordinating the tasks of improving ICOADS data for use in large-scale climate efforts was done, almost exclusively through informal participation of community members, but served nonetheless to shift the power structure from a centralized to a decentralized form.

The shift between ICOADS second and third regime includes a nesting of its self-organization into a polycentric model. Although ICOADS governance model formalized it did not move towards a command control, or centralized model; Instead, ICOADS nested itself within a number of national, international, organizational, and disciplinary jurisdictions of authority.

Polycentricity is defined by Aligica and Tarko as "structural feature of social systems of many decision centers having limited and autonomous prerogatives and operating under an overarching set of rules." (2011). The idea of a polycentric order was first offered by Michael Polanyi in the "The Logic of Liberty" (1951) and later used by Vincent Ostrom in the 1970's during debates about the efficiency of administrative structures in American municipalities (Aligica and Boettke, 2009). Critical to the study at hand, is the fact that Polyani originally developed the idea of polycentricy to describe the governance of science - which pursued knowledge in a variety of different disciplines and was hence subject to a variety of different value systems. But, what Polanyi described was the oddity that such a large enterprise could effectively function with no appeal to an ultimate power authority. Instead, informal rules and norms are learned through initiation, apprenticeship, and the evolution of social orderings. He was, in some sense, identifying this mode of knowledge production as a form of commons.

A passage from Aligica and Tarko clarifies an important point about Polanyi's writing of polycentric systems in science, "...an abstract and underoperationalized ideal cannot be imposed on the participants by an overarching authority. Thus, the authority structure has to allow a multitude of opinions to exist, and to allow them not just as hypotheticals but as ideas actually implemented into practice. The attempt to impose progress toward an abstract ideal is doomed to failure, as progress is the outcome of a trial-and-error evolutionary process of many agents interacting freely." (2011)

Polycentricity then is the absence of a centralized power, combined with a nested structure which allows the "many centers of authority" to cooperate effectively through trial and error. The emergence of cooperation through interactions is the crux of both Ostrom's work, and that of David Axelrod in his seminal work on game theory (1982).

In contemporary science settings, such as ICOADS, the division of labor is becoming increasingly depends upon collective action, and thus the emergence of peering, including peer production, is itself a stylized form of a polycentric model of governance - a "trial and erorr process, with many agents interacting freely" but organizing themselves at different levels of authority and rule-making, to produce complex goods. In ICOADS, this nesting of authority takes the following shape:

At the uppermost level is the World Meteorological Organization[5] *which promulgates resolutions and standards for all of meteorology to follow. Nested within the WMO is the Joint WMO-IOC Technical Commission for Oceanography and Marine Meteorology (JCOMM), an intergovernmental body that promotes standards, and coordinated marine observation systems, as well as data management. Nested within JCoMM are a series of workshops, CLI-*

[5]The WMO is, itself, nested within the United Nations

MAR and MARCDAT, which coordinate the research activities and agendas of the field of marine climatology. As Kent et al. put it, the "backbone of this work" is ICOADS (2001), a knowledge commons whose governance is now shaped by a steering committee with representatives from seven different international research centers and universities. Each representative is further nested within laboratories, research groups, and the various sub-fields that their research contributes to.

ICOADS polycentric system of governance operates efficiently because there is no one authority that transcends any one level of governance. Each level is able to appeal to a different level for directions on standards, operating procedures, or guidance. There is no polycentric police, judge, or jury. There is no legal or financial obligation, for example, if the organizations that signed a "letter of intent to cooperate" choose to leave ICOADS. But there are rules, and there are social repercussions for violating norms related to the historical background from which a commons emerges. In the case of ICOADS, the polycentric governance model creates a freedom to operate within one's own nation, discipline, or technical competence, but at the same time provides a rule set, and and enforcement strategies at various levels, types, and sectors of governance.

The polycentric model of governance has two important implications for sustainability within knowledge commons:

1. As Carole Rose explained in the "comedy of the commons" (1986), public goods are successfully provisioned, in part, because the users of these resource system are made complicit in their maintenance, and upkeep. Such a complex system of duty and responsibility is impossible to design without some amount of rules, authority, and norms being nested within a community structure. Similarly, the polycentric model of governance in ICOADS has become successful because it enables loosely connected stakeholders, even those that are not formally part of the governance body to contribute through letters of support, attendance and engagement in CLIMAR and MARCDAT, and through contributions to improving biases in the data -either formally by making submissions to IVAD, or informally through filing bug reports. Like the comedy of the commons description, the efficiency of such a complex institution would be difficult to predict, and even harder to purpose-

fully design. By creating a symmetry to operate within the commons, the polycentric model is able to achieve a high level of efficiency across scales and levels of a knowledge commons.

2. Nesting governance at different levels creates a certain amount of protection from disruptive disturbances by the very fact that the commons is coordinated across different levels and scales of governance. But, these interdependencies are also dangerous for the very fact that failure at a high level has the potential to cascade negative effects down to each subsequent level. For instance, federal budget disagreements about funding for basic science research cascaded down to the ICOADS project level in a rather dramatic fashion.

5.2.3 Case Study Research Question 2

What causes a governance system to shift from one regime to another?

Summary: Regime shifts are described through the ability of a knowledge commons to coordinate responses to selection pressures. A regime shift is related to selection pressures, diagnostics, and a number of other state variables - but ultimately, a regime will shift either out of opportunity (to improve upon the current model) or our of necessity (to avoid collapse). The ability of any system to respond to pressures and return to a functioning state can be described as resilience. ICOADS, in responding to selective pressures has demonstrated an ability to reconfigure its governance model in order to adapt to changing selection pressures. This is due, in part, to its flexibility in adopting new rules as well as its ability to formalize its governance when faced with pressures related to financial support. Transitions between governance regimes, and adaptivity are suggested as a future direction for sociotechnical systems research.

Supporting Narrative

Following Smith, Stirling, and Berkhout (2005), I will explain a regime shift as a combination of:

1. Selection pressures bearing on the regime, and
2. The coordination of the current governance model to adapt to these pressures.

In the following sections, I characterize the regime shift in ICOADS governance between the first and second, and the second and third regimes. I then describe the implications of these processes on ICOADS sustainability.

Regime Shift 1 to 2

In Regime 1, selection pressures where characterized as "competition" and "resource availability."

Niche competition within the climate science community for an ocean-atmosphere dataset resulted in significant conflict between the UK Meteorological Office, and the USA partners that were producing COADS. While both offered high quality data products, COADS could not make up for the temporal and spatial coverage of records that MDB held, and MDB conversely did not have the data processing capabilities, nor a sufficient cost-recovery model to significantly expand MDB.

COADS partners originally drew upon a principle of "free and unrestricted exchange" from the WMO Resolution 35, as well as the legacy of Lt. Matthew Maury in producing an open access resource. The cooperating COADS institutions forefitted their intellectual property claim to these early resource sets, and over time model of distribution helped COADS establish itself as a high quality data source within the climate science community. However, it was not a principle of openness, a resolution from a governing body, nor a naval legacy that would cause COADS to merge with its main rival, the MDB; but, instead, as my participant put it, "So, why did they merge? They had to. The science wasn't getting better without it" [ISC - 02].

The shift from a monocentric governance model to a self-governance model was a result of a collective desire to improve upon the current state of knowledge. This manifested itself in selective pressures like competition and resource availability, and the development of rules regarding what information must be shared amongst partners (information rules), and payoff rules, which specified who was to receive what benefits, and when. For example, the UK MET office would receive an advance copy of ICOADS processed data, with enhanced statistical subsets related to SST. This gave the UK

MET the ability to capitalize on their shared data, and begin working towards valuable secondary ICOADS data, such as the HadSST products.

Regime Shift 2 to 3

A shift between regime 2 and 3 was precipitated exclusively by the loss of financial support to NOAA partners. This sudden disturbance created a threat that ICOADS would no longer be a reliable source of marine surface data. The major selection pressure in this regime is "obsolesence", as ICOADS is forced to coordinate a broad network of actors in response to mounting pressure that it will not be able to serve end users, and produce a valued third release.

In coordinating a response, ICOADS project partners in the USA formalized agreements with international partners through a letter of intent to cooperate, which had the effect of further decentralizing the governance, and adding a new set of boundary, authority, and scope rules that would assign decision making power, establish voting mechanisms, and introduce new policy instruments for use by specific actors. This regime shift, like the one previous, had a nesting effect and contributed largely to ICOADS shifting from self-governance to a polycentric model.

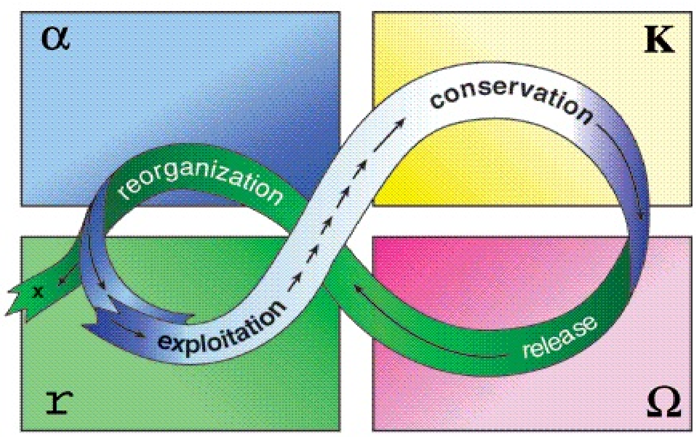
Causation

The explanations above make use of descriptive state variables, which I argue can help reduce the complexity of understanding a sociotechnical system's response to pressures. However, these descriptions are not pointing out direct cause and effect mechanisms, even where relationships between variables are observed. With this in mind, two broad conclusions can be drawn about the process of regime shifts in ICOADS:

1. As a sociotechnical system, ICOADS was able to absorb the effects of these pressures, and to coordinate responses which allowed for successful regime transitions. In this sense, ICOADS is a robust sociotechnical system by the very fact that it was able to effectively respond to these pressures. However, not all selective pressures and effects of a shift are equal. In the shift from regime 2 to 3 it is evident that there continue to be difficulties in producing a third release, and in fully realizing the potential of planned enhancements, such as the ICOADS value-added database (IVAD). This may mean that "obsolescence" is a more disrup-

tive pressure than others, such as competition or resource availability, which likely effect all sociotechnical systems at some point.

2. Cycles of adaptivity, and system's level resilience are an important future direction for knowledge commons scholars interested in sustainability. In Chapter 1, I offered a definition of resilience as the "measure of the persistence of systems and of their ability to absorb change and disturbance and still maintain the same relationships between populations or state variables." (Holling, 1974) But, the linear description of moving from one state to another is likely to have intermediate steps glossed over in this initial analysis. A move towards an adaptive cycle model, as pictured below, may be more helpful in understanding the ways in which sociotechnical systems, and knowledge commons in particular, respond to selective pressures; and more broadly, how sociotechnical systems successfully transition between governance regimes. I return to this proposal in Chapter 6.



Caption: A system - whether biological, ecological, organizational, etc. - is said to go through an adaptive cycle, where the ability to move between a back loop (e.g. $\alpha to \Omega)toaforeloop(e.g.rtoK)isafunctionofasystem'sitsresilience.$

5.2.4 Case Study Research Question 3

Which types of disturbances are ICOADS resilient or vulnerable to?

Summary: ICOADS current vulnerabilities are due to a lack of institutional diversity, and metric diversity. Over time, ICOADS has proven an *ability to increase its resiliency through a governance model that accommodates regenerative disturbances. Future work is suggested for the identification and further clarification of disturbance types, and responses.*

Supporting Narrative

Disturbances are simply changes to a system state. Recall that a definition of a scientific cyberinfrastructure is that, in part, they allow for the "smooth operation of scientific work at a distnace." (Edwards et al., 2006) Disturbances in a sociotechnical system like ICOADS can then be seen as phenomena that effect the "smooth operation" of their functioning.

In this analysis, I'll focus on two types of disturbances:

1. Regenerative disturbances build the capacity of a system to withstand future change.

2. Disruptive disturbances cause damage to a system, some of which may be irreparable.

Regenerative Disturbances

As ICOADS moved from a monocentric to a self-organized governance model it became less centralized, and matched its governance model with its resource scale. This in turn allowed ICOADS to develop a capacity to respond to selective pressures such as competition, resource availability and technological innovation. Controversies over name changes, infighting over future directions of the project, the introduction of new international governing bodies, and the discovery of known errors in the dataset (Woodruff, et al., 2006), caused initial changes to state variables, but ultimately helped ICOADS to refine the types of rules it was adopting to meet these needs. The responses and transitions between states due to regenerative distrubances should be seen as ICOADS building capacity to withstand change.

Disruptive Disturbances

In shifting to a polycentric model, ICOADS faced a number of stresses, including a partial loss of funding, the use of servers hosting a valuable subset of ICOADS, and three dedicated full time employees to the development of software crucial to a third release. An interesting correlation between variables when there is a sudden distrubances is that incentive-based policy instruments were replaced with input and output standards. This might imply that instead of simply trying to entice cooperation from partners, boundary and aggregation rules were put into place to assign responsibilities for producing a third release, restrict what institutional privileges came along with signing a letter of intent to cooperate, specify who was entitled to vote on future directions of the projects, and more generally, specify expectations for future contributions.

One of ICOADS vulnerabilities across all three regimes is a lack of metric diversity. As Cox et al., note, this leaves sociecological commons particularly vulnerable to change because it "optimizes system performance around criteria which may be indicators of short term success, but poor predictors of long term sustainability."(2014) The same is likely true of sociotechnical systems, which may optimize around certain metrics of success (e.g. citations in academia) which indicate, but do not predict a system's robustness.

When asked about the loss of funding, one participant explained:

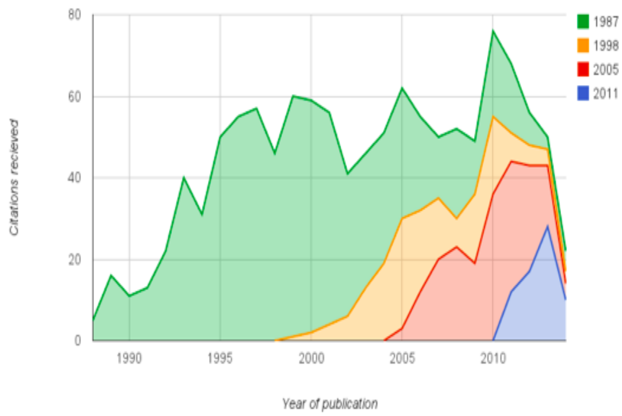
"...whoever was making that decision did not understand how important ICOADS is to the research community. How widely used it is ... because there was no statistics, there is no metrics of how many people would all of the sudden have their data products disrupted if ICOADS goes away" [ISC 06]

What bibliometric indicators that are available to quantify ICOADS impact are incomplete, and often misleading. For example, in a small sample of the documents that cite ICOADS[6], we found that 24% (n=79) had cited the wrong ICOADS release paper for the data that the study was purporting to use (Weber, Worley, and Mayernik, 2014). This gives the illusion that COADS release 1 remains widely used in contemporary research[7], when in fact that particular data release hasn't been available to end users since the late 1990's.

Additionally, when we look at the total number of citations to these data papers, they are disproportionate to the number of unique users that down-

[6]For clarification, to "cite ICOADS" simply means that a paper which announced a new release, such as ICOADS 2.5 is cited. This data comes from the citation content analysis described in Chapter 3

[7]We controlled for the fact that these studies may have indeed been doing historical studies which used release 1 data.c

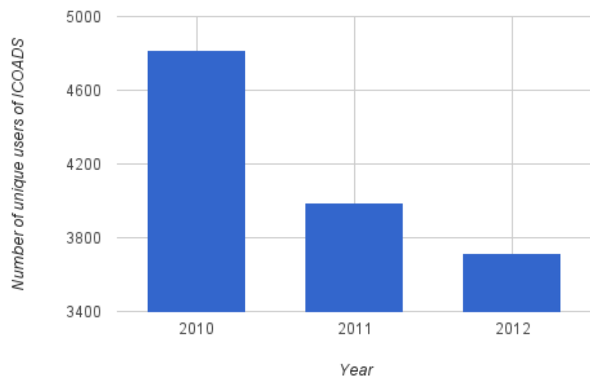


Caption:Distribution of citations to four major releases of ICOADS (COADS Release 1: 1987; ICOADS Release 2 1998; ICOADS Release 2.1: 2005; ICOADS Release 2.5: 2011)

load ICOADS data. For instance, ICOADS data release papers received a total of 201 citations from 2010-2012. During this same time period, 12,519 unique users downloaded at least 10 mb of ICOADS data from NCAR's Research Data Archive. This gives ICOADS a unique use download to citation rate of 0.016.

A final vulnerability that ICOADS faces, which fits neither the regenerative nor disruptive disturbance typology, is related to the selective pressure of obsolescence. In the sociology of science, Merton coined the phrase "obliteration by incorporation" (1948, p. 2728) to refer to processes where an original idea gains rapidly in popularity and use, and quickly becomes incorporated into a common stock of domain knowledge. Occasionally, this results in a failure to recognize the idea producer, because the idea becomes so widely understood that acknowledging its creator would seem redundant. The most immediate example is "Einstein's theory or relativity", but less iconic scientists have created lasting ideas only to be forgotten by way "obliteration" of credit through rapid "incorporation" of their work in the stock of common scientific knowledge. Garfield refined this idea with citations in describing how a novel idea, technique, or concept is occasionally so broadly and quickly adopted that it is fails to be formally cited at a comparable rate (Garfield, 1975).

ICOADS continues to be used as an input to large-scale reanalysis projects, and simultaneously is subset into innumerable smaller-scale products, many



Caption: Number of Unique Users that Downloaded $10mbofICOADSdata, 2010 - 12$

of which receive more citations and funding than ICOADS. It suffers from obliteration by incorporation into the climate record - failing to be recognized properly at either scale. Frischmann has similarly described the process of public infrastructures being subsumed into public good categorizations - they provide the valuable inputs into downstream innovation, but receive little of the upstream attention of investment firms or private sector support (2005). And indeed, we see ICOADS take on many infrastructure-like features, including the salient fact that it is rarely paid attention to until it breaks down (Star and Ruhleder, 1997).

5.2.5 Limitations of Case Study

The case study presented above has answered three case study specific research questions, largely by data collected during ethnographic and informetric studies of ICOADS community. There are a number of limitations to the results presented here, including the small sample size (n = 1), the scale of the informetric work, and the ability to demonstrate causative effects of changes in state variables. Establishing causation for regime shifts, or disturbance response should not be seen as the goal of the approach taken in this case study. Systematically coding state variables simply allows for the complexity of a system to be reduced, the different components of the system to be accurately described, and the findings of this work to be meaningfully compared to other cases. To overcome some of these limitations, I next compare the findings of this case study with those from three additional knowledge commons.

5.3 Cross-case Comparison

In this section I synthesize findings from across four different case studies of the knowledge commons, their evolution, and the background environments out of which they emerged. I begin first with an overview of three previously completed studies, and then describe the ways in which these different domains produce knowledge. I use a set of state variables developed in the ICOADS case study to compare the current governance models of all four commons. Finally, I use these comparisons to answer the dissertation's research questions.

5.3.1 Knowledge-Commons Framework

Madison, Frischmann, Strandburg have over the last five years developed a research agenda, and framework around the idea of a knowledge commons. Their argument turns on the assumption that "innovation and creativity are matters of governance of a highly social cultural environment." (2010 p. 669) That is to say that a state vs. market model choice creates a false binary for governance in the cultural sphere, and further, these two approaches are incompatible with innovation required of collective action institutions.

Their response is to adapt and extend the Institutional Analysis and Development (IAD) framework as it has been used in environmental studies. The result, a "knowledge commons framework", provides a method to empirically challenge the accepted wisdom that intellectual property rights and state governance provide the best options for sustaining a knowledge commons. The following case studies are focused exclusively on knowledge commons in scientific research and development settings.

5.3.2 The Genome Commons

Contreras contribution to the knowledge commons literature is a rich and detailed history of what he calls the genome commons (2014); databases

	Genome Commons	Urea Cycle Net.	Galaxy Zoo	ICOADS
Domain	Biology, Life Sciences	BioMedical, Public Health	Astronomy	Climate Science
Year Established	1996*	2003	2007	1983
Funding Mechanism	Mixed Portfolio	NIH	Sloan, Oxford University, Johns Hopkins University	NOAA, NSF, National Ocenaographi-Center (NOC - UK)
Method of Data Collection (for KCF study)	Document, Policy Analysis	Document, Policy Analysis, Survey, Interview, Observation	Document, Policy Analysis, Interview	Document, Policy Analysis, Interview, Observation, Participation, Informetric
Commons Resource Set(s)	Data, Databases	Data, Patients, Reagents, Literature, Infrastructure	Data, Software Platforms	Data, Metadata, Software, Stat. Techniques

Table 5.3

and digital archives of genomic data (DNA nucleotide base pairs) that have emerged thanks to a lowered price of DNA sequencing technologies, and large-scale international collaborative efforts to "map" entire species genomes. Through policy analysis, Contreras demonstrates, "the evolution of the genome commons from what was initially a public domain vehicle established to deter the proprietary designs of emerging biotechnology companies, into a unique polycentric governance institution for the growth, management, and stewardship of a massively shared public resource." (2014, p. 102)

Contreras draws our attention to early controversies over the patenting of databases, and the private sector's attempts to compete with large public sector projects like the Human-Genome Project (HGP) (Collins et al., 1998). This conflict, in part, leads to the generation of thousands of pages of policy written about the rules that data producers must follow, how archives are to guarantee access for potential users, and what scientific societies should do to create institutions for provisioning these resource sets over the long-term. The major contribution of Contreras work is twofold:

1. The genome commons demonstrate how, in an international setting with fierce competition for semi-rivalrous resources, a polycentric model of governance can emerge to nest institutions within a highly functional framework of cooperation. This is due largely to what Contreras identifies as the Mertonian norms upon which science, its actors, and the design of its institutions are based.

2. In practice, the genome commons are effective because of policy instruments like "embargo periods" for archived data which grant "right of first access" to genomic data contributors. In previous works, Contreras describes the need for a "latency analysis" to better understand how

the rapid dissemination of research results are balanced with limited exclusive rights to access (Contreras 2010a, 2010b). These dimensions, he suggests, are critical to the design of future policy governing knowledge commons.

5.3.3 Galaxy Zoo, and the Astronomy Commons

Madisson uses the knowledge commons framework to explore the intersection intellectual property rights, crowdsourcing, and "big data." (2014) He pursues a comparative project asking "How do the outcomes produced by commons governance differ from outcomes that might have been available if alternative governance had been employed?" (2014 p. 231) To do so, Madisson's main contribution is a case study of Galaxy Zoo and its model of sharing credit amongst a large number of crowd workers, and astronomers. He uses findings from this work to compare the institutional arrangements of Galaxy Zoo with the Nearby Supernova Factory project, whose organizational structure resembles that of a traditional firm.

Like Contreras, Madisson suggests that the Galaxy Zoo project succeeds in part because of a broad appeal to the Mertonian norms of science. Madisson also claims that this knowledge commons exemplifies Hagstrom's claim that science is a "gift culture" (1968) trading in ideas instead of material gains. This is a claim supported by an observation that the project's designers, exemplified by Chris Lintott, forefitted personal claims to data transcribed by crowdworkers, as well as the Galaxy Zoo's commitment to the credit and acknowledgment of these contributions in resulting journal publications (2014, p. 235).

Madisson makes an important point that the openness of a resource set, both those produced and provisioned by an entity like Galaxy Zoo, is not a binary distinction; it is instead a continuum. The data, and underlying software that produces Galaxy Zoo is free, and openly accessible, but in the case of the data, it was only released to the public after an initial round of project specific publications were in press[8] (2014, p. 215).

Another major conclusion from this work is that "...institutional order and knowledge governance, such as commons, are mutually constitutive. One

[8] Describing how these kinds of distinctions in "open" and "free" access effect long-term sustainability is described a major open question for knowledge commons scholars.

largely exists because of the other." (Madisson, 2014 p. 211) Madisson argues that the emergence of crowdsourcing, peer production, and big data management in science are creating new orders of power, and democratizing access for many different (and new) knowledge commons stakeholders. In turn, the social dilemmas arising from broad collective action are poorly understood and deserve continued attention of those concerned with knowledge commons in science and technology settings.

5.3.4 Urea Cycle Disorder Research Network Commons

Strandbrug, Cui and Frischmann apply the knowledge commons framework to a rare disease research context, which consists of patients, caretakers, biomedical researchers and life scientists, as well as public policy experts across a number of cooperating institutions in the USA (2014). As they note, the need to share resources such as patient data, outreach materials, study protocols, reagents, diagnosis, and even patients themselves creates the need for a commons approach to open exchange and resource pooling. This however, does not mean the research coordinating networks that participate in a commons are absent of rivalry; there is intense competition for funding from the NIH, for locating and managing patients to participate in a cohort study, and for longitudinal data collection that one institution may bear the majority of costs to produce and provision.

Although embedded in the a biomedical field which often has strong private sector interests, the authors note that there is a strong incentive to openly share resources in this domain,"'Open' approaches are particularly attractive in the rare disease context, given the small numbers and geographical dispersion of potential research subjects and the inapplicability of the blockbuster drug business model." (2014, p. 155) Pharmaceutical companies may have little financial incentive to produce drugs for rare diseases *because* they are rare and represent a relatively small market. As a result, rare disease research and treatment comes with the need to conduct clinical trials, rapidly share results, and work closely with private industry to produce effective drugs. The rare disease research context therefore nests clinical, academic, pharmaceutical / private industry, as well as the NIH and a number of other non-profit organizations.

5.4 Research Questions

This dissertation has pursued answers to two broad research questions:

What are the effective institutional arrangements (governance) for sustainable scientific knowledge commons? And, how do these arrangements differ between domains of knowledge production?

In the following section, I compare and contrast the previously described case studies based on the contexts in which they produce and provision a knowledge commons. This analysis provides answers to how knowledge commons differ between domains of knowledge production. In the section following this one, I characterize the effective governance of sustainable commons across these different domains using coded state variables.

The following table summarizes the characteristics of knowledge commons that will be compared.

p3cm

	Genome Commons	**Urea Cycle Disorder Consortium**	**Galaxy Zoo**	**ICOADS**
Potential for Commercialization	High	Med / High	Low	Low
Resource Type (s)	Common Pool Resource	Common-Pool Resource / Public Goods	Public Good	Public Good
Governance Regime	Polycentric	Self-organized	Self-organized	Polycentric
Discount Rate	High	Low	High	Low
Model of Peer Production (Haythornthwaite)	Heavyweight	Heavyweight	Lightweight	Heavyweight
Task Uncertainty (Whitley)	Low	Low	Low	Low
Mutual Dependence (Whitley)	High	High	High	High
Latency Requirements	High	Medium	Low	Low

Table 5.4: Comparing knowledge production and provisioning in different knowledge commons

Mutual Dependence and Task Uncertainty

One way to initially compare across these different knowledge commons is to ask what the background environment supplies in the way of a "default" assumption about the production and sharing of resources. Since each case study is embedded in a scientific domain of knowledge production, the discipline or field specific to this context is likely to have some influence on these practices. Whitley's work on the differences of knowledge production practices in a" field" of science provides a helpful framework for understanding these issues. In this work, Whitley has described "the nature" of intellectual fields through variations in two dimensions: mutual dependence and task uncertainty (2000).

1. *Mutual Dependence* describes the degree to which one field relies upon another for validation of findings or for inputs to new work. Fry and Talja point out mutual dependence also accounts for "...the extent to which a field adopts evaluation criteria and standards from other fields for the assessment of work externally produced, rather than developing its own criteria." (2007, p. 118).

2. *Task Uncertainty* is the degree to which members of a field understand and articulate problems to be solved, and agree on methods or techniques needed to solve those problems. For instance, high task uncertainty means there is little coordination, results are based on loosely constructed criteria for acceptance or standards for significance.

For a commons to form in the first place there should be a low degree of task uncertainty. Producing a shared resource, in some ways, is agreeing upon the validity or meaning of that resource. And, it should be expected that each domain invested in commons-based work has a high degree of mutual dependence, as commons are often formed for the express purpose of collective action.

All four case studies exhibit high mutual dependence and low task uncertainty. This classification, which is the only agreement across all four cases, is helpful because it makes clear the initial conditions that a field should have in order to most effectively use a commons arrangement[9].

Peer Production

Each of these case studies demonstrates that the division of labor needed to reliably sustain a knowledge commons depends, in part, on peer production models. Increasingly, peering occurs in not just the production of new resources, but also in the provisioning of shared infrastructures, and the preservation of existing resource sets. To differentiate between models of peer production, I'll use Haythorthwaite's notion of "heavyweight" and "lightweight"

[9] A small qualification: Other combinations may result in an effective commons, but these are not as readily obvious as the high mutual dependence and low task uncertainty combination

peer production described in Chapter 2: The Genome Commons, The Urea Cycle Disorder Network, and ICOADS all use a form of heavyweight peer production, where strongly connected and highly committed experts make contributions to a common pool of resources which require peer review before acceptance. Galaxy Zoo on the other hand uses a crowdsourced form of data annotation which resembles the lightweight peer production model; a large number of weakly connected novices perform simple tasks that are aggregated and quality controlled in producing complex resources.

These characteristics of peer production effect a number of other ways in which the resources and governance of a knowledge commons can be characterized; including, the resource type, the governance model that results, the latency requirements for protecting producers / innovators, and the discount rate that users of a commons adopt.

Resource Types, Commercialization, and Latency Analysis Resources produced by these knowledge commons are largely the results of scientific research processes, and as such they have some degree of rivalry. This means that categorizing the type of good being pooled will be inexact; each of these commons has a shared resource system made of goods that fit somewhere between common-pool resources and public goods. Only the Genome and Urea Cycle disorder commons have the realistic capability of being commercially exploited, and this leads to their having a need for latency in the release of these data resources to a broader community of users. In Galaxy Zoo, and ICOADS[10] limited monopolies are not likely to be institutionalized, although both have at times informally delayed immediate data release for this purpose. In the future, this may require Galaxy Zoo and ICOADS to adopt new rules, or design new approaches to data access.

Discount Rate In Chapter 2, I described characteristics of successful commons in the socioecological realm. Ostrom (1999, 2000) argued that one of the most important factors in sustaining a commons is those who are most dependent on the resource have a "low discount rate" in approaching the common resource; in other words, they have a "willingness to sacrifice current payoffs for higher

[10]This was observed in major data contributors receiving early access to new releases

payoffs in the future." (Acheson, 2011) In the knowledge commons, discount rates can be observed in the willingness of individuals to cooperate, contribute, or responsibly consume resources so that they exist for future generations to use.

Contreras discussion of the Genome commons shows that a high potential for short-term profit - either through commercialization or through credit of discovery - creates the need for formal strict regulatory policies and normative enforcement of rules in the genome commons. Another dimension of discounting important to the genome commons is related to the rapid technological advancement of the field; differences between first, second, and third generation sequencing data quality means that these datasets decay in value rapidly. This is the direct opposite of ICOADS, where historical weather records mature in value over time. Actors within these two domains have different approaches to valuing the future of a resource set: Genome commons actors will have a high discount rate, while ICOADS actors will have a low discount rate.

The Urea Cycle Disorder network curiously mixes high and low discount rates; the limited nature of its funding (5-year NIH renewable grants) requires the rapid production of research results, but also requires a sustained effort to keeping the network effectively functioning for future rounds of funding. This tension leaves actors sometimes discounting a future which might not exist, and others carefully preserving resources for long-term use (see Strandburg, Cui, and Frischmann 2014, p. 165 for further discussion).

The Galaxy Zoo, and crowdsourcing in general, coordinate crowd workers with high discount rates - not only will they likely not be involved in the future of the research project, but the strength of ties between actors are very low. This is one of the steep costs of using a crowdsourcing approach to data transcription, and is likely to be important in address in sustainability planning. Astronomers involved in the project often have a low discount rate, as the data sets resulting from this work are highly valued by the research community and likely to be inputs to many future research projects.

Governance Model

Given the characterizations described above, which models of governance do these knowledge commons actually use? The Genome Commons and ICOADS are both nested within a number of different institutions of decision making power and authority. This is partially due to the fact that they are both international in scope, have existed for a long period of time, and receive funding or infrastructure support from a number of different institutions. They can both be seen as adopting a polycentric governance model, with nested levels of authority and rule enforcement.

The Urea Cycle Disorder network most closely resembles a self-organized governance model; it receives only partial direction from the NIH for governing its shared resources, and all institutions directly involved in the network have equal authority in establishing and enforcing rules. Similarly, Galaxy Zoo receives partial direction from its project partners at Oxford University and the Sloan foundation as a funding agent, but by and large the project remains self-organized; reporting informally to a board of directors assembled to vote and establish by-laws. All project partners having a relatively equal role in shaping future directions of the project, but the overwhelming power of authority is held by a few individuals.

5.5 Governance Variables

In this section I further characterize the effective governance of sustainable knowledge commons using a set of state variables developed in the case study of ICOADS. These states are summarized in the table below, but my analysis focuses on only a subset of these state variables.

	Genome Commons	Urea Cycle Dis.	Galaxy Zoo	ICOADS
Multiple levels	Coordination among multiple levels	Coordination among multiplelevels	Single-level governance	Coordination among multiple levels
Institutional diversity	High	High	Low	Medium
Centralization	2- Somewhat Decentralized	3-somewhat centralized	4- Highly Centralized	2-somewhat decentralized
Governance scale	International	Sub-national	State	International
Scale match	Match	No-match	No Match	Match
Metric diversity	Medium	Medium	High	Low
Rules-In-Use	Information, Aggregation, Scope, Payoff	Boundary, Information, Authority, Aggregation, Scope, Payoff	Information, Payoffs	Boundary, Position, Information, Authority, Aggregation, Scope, Payoff

Table 5.5

5.5.1 Centralization

The degree of centralization in any governance system impacts how its decisions are made, how outcomes are achieved, and how the commons is managed. Both ICOADS and the genome commons are decentralized (e.g. their administration and points of service are distributed across numerous institutions) including rule-making structures which extend to a number of national and international scientific agencies. The Urea Cycle Disorder network is somewhat centralized, as its decision making focuses largely on PI's named in its founding NIH grant. Similarly the Galaxy Zoo is highly centralized, with governance located almost exclusively within its founding institution, and small board of directors.

5.5.2 Governance Scale and Scale Match

ICOADS and the genome commons are in a state of governance match, where the institutions coordinated, and the resources produced or provisioned are international in scope as is the governance scale. The Urea Cycle Disorder network governance is exclusively national in its governance model, but the patients they manage and the resources they produce are global in scope. Similarly, Galaxy Zoo has a state-level governance model, but coordinates the work of a large international crowd, whose interests and use of data includes experts and non-experts. Both Galaxy Zoo, and the Urea Cycle Disorder network governance do not have a scale match, which may impact their growth and stability. However, it is worth noting, that early in ICOADS tenure, it too lacked a scale match. The Urea Cycle Disorder network and the Galaxy Zoo are the two youngest knowledge commons being analyzed; a mismatch in scale and governance may simply be an artifact of the evolution of governance rather than a serious fault of either commons.

5.5.3 Metric Diversity

Galaxy Zoo is the only knowledge commons with a high diversity of metrics. This is partially a result of the broad success of the project, which has been able to redefine it's goals and mission a number of times as the platform for crowdsourcing rapidly evolved. As a knowledge commons, Galaxy Zoo has also proved to be capable of hosting and designing novel transcription projects outside of astronomy. Thus, the resource set expanded from data, to software, to even the community of crowd-workers themselves that migrate from project to project. Each of these shifts has allowed Galaxy Zoo to set new goals, and consequently describe the success of those project's on terms that are appealing to potential funding agencies. A non-trivial spillover effect, is that this diversity of metrics allows Galaxy Zoo's spokesperson to quote convincing statistics of success throughout his many speaking engagements around the world - adding substantively to the reputation of the project (Madisson, 2014 p. 255)

ICOADS, the genome commons, and the Urea Cycle Disorder network all have medium to low diversity of metrics. A strong tie to funding agencies at national scales, and the academic nature of these research communities locks them into bibliographic and informetric (usage-based analytics) evaluation tools which are, as described throughout the ICOADS case study, limited indicators of a commons' success.

5.5.4 Rules-in-Use

All four knowledge commons share information and payoff rules in common. At a broad level, these are likely the most essential rules for a successful knowledge commons to adopt, as they specify not only what information must be shared, and with whom, but also how resources are to be distributed amongst stakeholders. In this sense, entitlements (either through production or provisioning) are a prerequisite for engaging in a commons, both at an institutional and individual level.

The Urea Cycle Disorder network, the genome commons, and ICOADS also share aggregation and scope rules in common. These may be rule types of a mature commons, and may be an artifact of those commons having faced controversies in the past, such as the genome commons enclosure, or the defunding of ICOADS, which exert pressures related to governing outcomes.

ICOADS is the oldest institution, and is also the only knowledge commons that, in its current state, has some version of all seven rules as a part of its governance structure. As the case study demonstrated, these rules have accreted over time; each new regime added a new set of rules, while retaining some or all of the functions of the previous rule set.

5.5.5 Summary of Answer to Research Questions

Based on a cross-case comparison, effective governance models for sustainable scientific knowledge commons include polycentric and self-governance models. These two methods of governance allow for what commons scholars have long recognized as a key to effectively functioning shared resources systems; they allow "...a group of principals to organize themselves voluntarily to retain the residuals of their own efforts." (Ostrom, 1990) In ICOADS and the genome commons, we see that as the projects matured they became embedded in networks of rule making authorities, effectively nesting institutions in overarching policies that provide directions for particular actions, while still allowing individual actors and organizations to "retain the residuals of their own efforts."

The Urea Cycle Disorder network is a complex and valuable study of the limit of commons governance at a national scale. Quite obviously, the disease is international, so how this group effectively cooperates in a local context, and maximizes their ability to self-govern is an important case study for the NIH, as it approaches the design of a "big data" commons. And yet, the centralization of its rule-making, and the lack of scale match presents potential dilemmas for future sustainability efforts. It should be noted that this particular network had its NIH funding renewed in late 2014, and will effectively be operating until 2020, providing the opportunity for a longitudinal case study of self-governing, centralized commons governance.

The Galaxy Zoo case study is unique to science commons, but provides a potentially valuable link to other lightweight peer production platforms governed as commons. The relatively centralized governance, which also lacks a scale match between its resources and rule-making, is nevertheless the most visible and diverse of the four case studies. Its has effectively ported its transcription model to annotation in a number of different fields, including the climate data from old log-books which are feeding into ICOADS Release 3.

What the diversity of case studies here indicate, is that commons governance is mutli-faceted. It requires a match between the resources being shared, the management structure of the provisioning institutions, and the production model of the resource contributors. And yet, as the Urea Cycle Disorder network and Galaxy Zoo point out, the effective combination of rules in use, and a diversity of metrics communicating success can overcome these initial criteria. This is especially important to Galaxy Zoo, as the discount rate of its potential contributors are quite high.

Traditional notions of "field" specific knowledge production, although not exhaustive, remain helpful to the analysis of knowledge production in commons-based institutional settings. Whitley's mutual dependence and task uncertainty works as kind precursor to success - without which it is unlikely that a commons could easily succeed. But, the case studies described above also demonstrate that there are a number of important variables to consider besides simply the default background assumptions of an intellectual field. The use of networked information communication technologies increase the ease at which collaboration can take place across space and time; this, coupled with a the decrease in available funding for basic science research creates an environment of necessary innovation in science and technology research. The knowledge commons frameworks described here points to important state variables, but there are a number of other dimensions, each of which may be equally as important as governance, that have not been addressed in this study. In the following chapter, I summarize these limitations and point to future directions in this research.

5.6 Summary

In this chapter I have summarized the research findings from a case study of ICOADS governance, and answered three research questions related to the effectiveness of ICOADS' governance. I then compared these findings with three previously completed case studies, and answered the overall research questions of this dissertation. In doing so, I demonstrated how a systematic approach to coding state variables could be adapted from socioecological to sociotechnical settings. In the final chapter, I describe limitations of this work, and the implications of this research for policy, practice, and theory of the commons. I conclude with future directions for this work.

CHAPTER 6

6.1 Introduction

In this chapter I restate the limitations of this work, and describe the implications of this dissertation's findings for policy, practice, and theory of the commons. I conclude with future directions for studies of sociotechnical systems sustainability.

6.2 Conclusions

6.2.1 Effective Governance for Knowledge Commons

This dissertation began by noting the differences between scientific knowledge, and the objects that result from knowledge production: On the one hand, scientific knowledge is assumed to be a public good, with low rivalry and costly excludability; On the other hand, the products of scientific research (data, software, journal publications, etc.) are assumed to have rivalry for access, which in turn requires state intervention, or privatization for managing sustainable access and use.

The analysis found in Chapters 4 and 5 has demonstrated the efficacy of a hybrid model, the commons, used by four domains of knowledge production in research and development settings. *Under the right circumstances*, the commons can be an effective model for balancing interests between private and public sectors (e.g. genome commons, and Urea Cycle Disorder network), creating long-enduring institutions for producing and provisioning valued databases (e.g. ICOADS, and genome commons), and sustaining order and cooperation between large numbers of individuals competing for access to resources (e.g. observations made across all four cases).

The case studies reviewed in Chapter 5 are an initial step in providing empirical findings about those *right circumstances*; including, how knowledge commons efficiently operate, how they evolve and change over time, and the ways that governance of sociotechnical systems differ in sustaining shared resource sets.

This research indicates sustainable knowledge commons in science settings have the ability to:

- Match scales between governance regimes, and resource sets.
- Find ways to diversify institutional arrangements as resource sets and stakeholders grow in scale.
- Articulate selection pressures, and diagnose underlying drivers and symptoms of those pressures.
- Draw upon social and technical capital, and coordinate relevant stakeholders in response to selection pressures.
- Define, and optimize activities for a diversity of metrics.
- Balance policy instruments that are regulatory, and incentive-based.
- Adapt quickly to sudden disturbances, and draw on social and technical capital when disturbances are slow and continuous.
- Seek out, and absorb "regnerative disturbances", and consequently build a capacity to withstand "disruptive disturbances."
- Accrue and revise rules-in-use. Information and Payoff rules were recognized as an essential baseline for succesful knowledge commons across all four cases.

Some of these results are self-evident, but without empirical studies to clarify precursory assumptions, sociotechnical systems will, like socioecological systems, fall prey to oversimplifications of what works when, where, and for whom. This work also helps to underst understand which dimmmensions of sustainability should be explored in more detail, such as how the design of effective policy can encourage a match in governance scales; or, how institutions might be encouraged to diversify institutional arrangements and metrics through incenstives. This is to say that, the type of research presented here may be preliminary, but it is also foundational: it picks out systems features, names them, and puts standardized categories on them so that they can be compared with other similar systems in order to build reliable knowledge.

In the next section, I extend some of these observations to the polycentric system of governance.

6.2.2 Polycentric Models of Governance

Research and development activities in the sciences can be divided into many axis of categorization: basic vs. applied, theoretical vs. practical, inductive vs. deductive, etc. In contemporary research settings, a distinction that cuts across many of categorizations is that of collective action; research and researchers increasingly collaborate, share resource sets, and divide up labor in pursuit of grand challenge science problems. (Atkins, 2003) Even mathematics has turned to peer production to solve difficult theorems. (Ball, 2014)

Governing these new arrangements for sustainability will be an important, but difficult challenge for their stakeholders. The case study of ICOADS, the genome commons, and the Urea Cycle Disorder network has demonstrated that as a project matures, it comes with a need to formalize certain governance issues through rule making procedures of self-organized groups. In the case of the Urea Cycle Disorder network, matching the scale of governance with the resource sets provisioned has proven difficult for an international phenomena that is funded exclusively as a USA based research project. The genome commons and ICOADS both present cases where international cooperation is required for both the provisioning, and the production of resource sets. Over time these two knowledge commons have found ways to match those scales, although each initially struggled to do so.

The polycentric governance model observed in ICOADS and the genome commons developed as a result of growth and evolution, rather than purposeful design. The regime shifts that mark ICOADS changing from a self-governance model to a polycentric model was in fact response to selective pressures - such as a loss of funding from a major sponsor and the need to formally embed itself in a number of different bureaucratic structures.

As these polycentric models themselves mature, an important next step is further investigating their features; What types of nested levels can be observed across polycentric models? When do transitions between self-governance and polycentric models occur? What are the negative externalities (the unintended consequences) of having an interdependent governance system like a polycentric system? How, and in what ways do the the negative (or positive) externalities of a polycentric system differ from that of Federalism?

Limitations of the Polycentric Model

While I have focused largely on the positive aspects of polycentric systems having multiple levels of jurisdiction, this also creates multiple sources of friction, and redundancy. This is especially evident in science and technology where a mix of self-governance and state-based hierarchical systems create policies which are too broad to be useful, or too specific to be followed with intended fidelity. This is the current case with data management planning, and data sharing initiatives promulgated by federal research funding agencies. Borgman characterizes the "conundrum of data sharing" this way:

> "An investigator may be part of multiple, overlapping communities of interest, each of which may have different notions of what are data and different data practices. The boundaries of communities of interest are neither clear nor stable. In the case of data management plans, an investigator is asked to identify the appropriate community for the purposes of a specific grant proposal and for the proposed duration of that award." (2011)

This characterization sounds much like a polycentric system of governance, but one in which there is a mismatch between resource sets, actors, and rule making procedures. Friction and confusion result from funding agency mandates aimed at governing data management at a project level, and practices around sharing data for reuse at a field or discipline level. The success of broad programs like a "data management" initiative will require a match between scales, rules in form and rules in use, as well as the exploration of polycentric models of governance. Otherwise the provisioning of important resource sets will continue to be "conundrums" rather than effective public policies for increasing the impact of science and technology on society.

6.3 Limitations

Research in a sociotechnical setting, such as a knowledge commons, will inevitably suffer from a limited view of what are complex systems, and this research project is no exception. This work has been limited by a number of factors, including:

- The small sample size (n=4) of this work limits its generalizabilty. Additionally, the partial description of what are large and complex international institutions is limited by my own knowledge of these systems.
- The frameworks applied here are relatively new, and require a broader set of case studies to test their efficacy for generating reliable, comparable results.
- A majority of the analysis in this document has focused on the governance of knowledge commons. Other components briefly described here, such as actor groups or the characteristics of a resource set, may be equally important to sustainability.
- The approach used to systematically coding state variables is also relatively untested in sociotechnical systems research. With a small sample size, the limits of this approach are not well understood.
- Sociotechnical systems include many more context than research and development in science, and this may well effect the applicability of this approach to new contexts.

6.4 Future Work

Future work in the study of sociotechnical systems requires a number of important modifications to this intial study, which I describe in detail below.

I focus in particular on two open research questions from the comparative case study found in Chapter 5.

- Both Contreras and Madison relied on an account of Mertonian norms, otherwise known as CUDOS - Communalism, Universalism, Disinterestedness, and Organize Skepticism (1973) - to explain their case study subjects being committed to openness and sharing. Although the concept of "communalism" seems a good fit for this context, Merton's norms in general are an awkward framework for a research agenda like the knowledge commons. Merton's normative theory is grounded in functionalism - he sees the actions, behaviors and adopted rules of a culture through latent and manifest functions that are, above all else, constructed for the sake of preserving a culture (1963). Functions explain how cultures organize and act, in order to remain the same (how they continue to *function*), not how and why cultures evolve. This seems to conflict with the major goal of the knowledge commons research agenda, which is to provide an alternative to the functionalist accounts of institutions, especially those that rely on intellectual property rights protections. (see, Madisson, Frischmann and Strandburg, 2010 p. 665-78)

  And yet, in the ICOADS case study I too described a "default" moral framework that my participants referenced in describing the historical precedence of their work, especially as it related to Lt. Matthew Maury. My findings also seem to conflict with the fact that, in practice, these systems resembled flexible evolution, and adaptation, rather than fierce guarding against change. An open research question then is which theories (political, sociological, or otherwise) can help explain the continuity of this normative dimension of communalism given the dynamic evolutionary account that occurs throughout this dissertation. To put it more simply, collective action requires dynamic change and readjustment for sustainability, so how is it that there remains a normative center upon which geneticists, rare disease researchers, astronomers, and climatologists all seem to adhere? This cannot be the case. Careful attention to future differences between domains of knowledge production, their norms, and their practices is needed within the knowledge commons framework.

- The institutional level case studies presented here, with the exception of the Galaxy Zoo, runs somewhat counter-intuitive to previously published literature about the culture of science being "gift giving "(Hagstrom, 1965) and collaboration having a propensity that is subject to bounded-rationally (Birnholtz, 2005), or a matter of individuals using a limited monopoly for rent extraction (Birnholtz and Bietz, 2003; Harvey, 2009). Traditional notions of rent are effective only when the costs and benefits of collaboration are symmetrically distributed amongst stakeholders. Collective action dilemmas, which emerge when resources are pooled to accomplish basic tasks, such as calculating historical weather trends or sequencing a species genome, pose a different set of problems than property-based concepts like "rent" or bounded-rationality can provide explanations for. And this is becuase institutions for collective action suffer from social dilemmas where costs are *not* symmetrically distributed. This is not to say that behavioral theories are inapplicable to the commons context; quite the opposite - there is a need to understand how those traditional notions of limited monopoly (rent) effect commons when nested in polycentric governance structures. Contreras offers "latency analysis" but this describes only one temporal aspect of the social dilemmas of sharing resources and receiving credit.

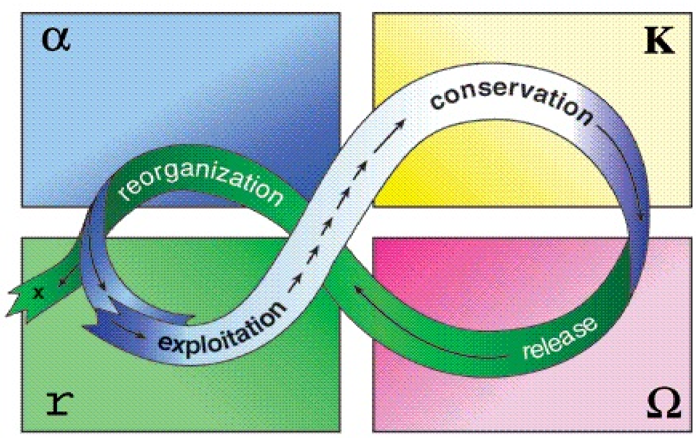
- The work presented here also suggests that beyond science commons, the sustainability of sociotechnical systems should be treated as a collective action dilemma - the groundwork for this approach in design based fields was laid by Paul Dourish (2010) in HCI. As he points out, design interventions for environmental sustainability are often focused on individual choice, which has (intentionally or not) "transformed the problems of sustainability into the cost benefit trade-offs of rational actor economics" (2010, p. 2). Again, abstracting from the environmental domain to the sociotechnical requires a number of modifications to work in HCI that focuses on ICTs for (environmental) sustainability. Developing new commons theory to explain institutional successes and failures, including case studies using the Knowledge Commons Framework, is an important first step.

To continue advancing a robust empirical account of sustainability in sociotechnical systems, I suggest that further work is needed on resilience and adaptive cycles.

6.4.1 Resilience, Vulnerability & Adaptive Cycles

The work presented here describes how a knowledge commons effectively transitions between different governance regimes over time. It suggests a need to turn away from functional integrity explanations of sustainability, and instead adopt a framework of analysis that is more congruent with studying knowledge production, and the governance of knowledge products through systems concepts like adaptivity and resilience. Future work can and should go in a number of directions. I will summarize two below:

- Adaptive cycles focus on two phases or transitions between system states: Phase one, the foreloop, is characterized by growth and accumulation, and phase two, the backloop, is a rapid response for reorganization and renewal (Holling, 1973).



Holling's adaptive cycle (1973), with four state changes and tranistions (loops) between different states

If a system doesn't develop a capacity to complete a backloop (from Omega to Alpha), it fails to renew and will collapse, hence it is not resilient to change and is unsustainable. The sociotechnical systems described here all had the ability to complete a backloop, although with varying degrees of success, and over different (longer and shorter) periods of reorganization and renewal. As Holling pointed out in socioecological systems, a path towards ecosystem collapse begins by extending foreloops (2007). Often times this means that capital is spent treating the symptoms of a problem, delaying the initial regeneration and increasing the likelihood that a system won't continue to respond quickly or efficiently to external pressures.

- To move towards a model of sociotechnical systems resilience and adaptivity, future work might begin by characterizing components of change, such as those from Holling's model, and the transitions between those components. In the state variable coding developed in this dissertation, I partially addressed adaptivity through the "diagnostic" variable, where either the symptoms or the underlying drivers of a disturbance are treated. Future work on resiliency and adaptive cycles for sociotechnical systems should also investigate this closer - trying to both characterize and better understand responses to selective pressures, disturbances, etc.

  Geels and Schot have developed a concept they call "pathways to transition between sociotechnical regimes." (2007) Like foreloops and backloops in the above socioecological model, the transitions approach conceives of regime change through four sustainability pathways called transformation, reconfiguration, technological substitution, and de-alignment re-alignment (Geels and Schot, 2007). These do not conceive of state changes as cyclical, and so are not meant to connect any one state to any other. Future work in this area should try to characterize how transitions, whether on Geels and Schot's terms or othwerise, connect to identifiable, general states of a sociotechnical system. It as at that point that collapse, or resilience can be better conceived in the sociotechnical realm.

Finally, much of this work has focused on differentiating between sociotechnical and socioecological systems, but as the work of Dourish (2010), Blevis (2007), and Tomlinson et al (2013) have shown the two are important compliments in a broader sustainability context. My work here demonstrates that commons as well as systems theory can aim to generalize to both sociotechnical and socioecological settings.

APPENDIX A

REFERENCES

Acheson, J. (n.d.). Ostrom for anthropologists. , *5*(2), 319–339. Retrieved 2015-05-01, from https://www.thecommonsjournal.org/index.php/ijc/article/view/245

Acheson, J. M. (n.d.). *The lobster gangs of maine*. Upne.

Aligica, P. D., & Boettke, P. J. (n.d.). *Challenging institutional analysis and development: the bloomington school*. Routledge.

Aligica, P. D., & Tarko, V. (n.d.). Polycentricity: from polanyi to ostrom, and beyond. , *25*(2), 237–262. Retrieved 2015-04-29, from http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0491.2011.01550.x/full

Anthes, R. (n.d.). *NCAR president's corner letter on free and open exchange of scientific data.* Retrieved 2015-04-06, from http://www.ucar.edu/communications/quarterly/spring97/president.html

Appadurai, A. (n.d.). *The social life of things: commodities in cultural perspective*. Cambridge University Press.

Atkins, D. (n.d.). Revolutionizing science and engineering through cyberinfrastructure: Report of the national science foundation blue-ribbon advisory panel on cyberinfrastructure. Retrieved 2015-05-01, from https://arizona.openrepository.com/arizona/handle/10150/106224

Ball, P. (n.d.). Crowd-sourcing: Strength in numbers. , *506*(7489), 422–423. Retrieved 2015-05-01, from http://www.nature.com/doifinder/10.1038/506422a doi: 10.1038/506422a

Barbrook, R. (n.d.). The hi-tech gift economy. , *3*(12). Retrieved 2015-05-01, from http://firstmonday.org/ojs/index.php/fm/article/viewArticle/631

BD2K, N. (n.d.). *Commons home page | NIH: ADDS portal.* Retrieved 2015-05-01, from https://datascience.nih.gov/commons

Benkler, Y. (n.d.-a). Between spanish huertas and the open road: A tale of two commons? , 69.

Benkler, Y. (n.d.-b). Between spanish huertas and the open road: A tale of two commons? , 69.

Benkler, Y. (n.d.-c). *The wealth of networks: How social production transforms markets and freedom.* Yale University Press.

Benkler, Y., Roberts, H., Faris, R., Solow-Niederman, A., & Etling, B. (n.d.). Social mobilization and the networked public sphere: Mapping the SOPA-PIPA debate. (2013). Retrieved 2015-05-01, from http://papers.ssrn.com/sol3/papers.cfm?abstract$_id = 2295953$

Berman, F., & Cerf, V. (n.d.). Who will pay for public access to research data? , *341*(6146), 616–617.

Bietz, M. J., Baumer, E. P., & Lee, C. P. (n.d.). Synergizing in cyberinfrastructure development. , *19*(3), 245–281. Retrieved 2015-05-01, from http://link.springer.com/article/10.1007/s10606-010-9114-y

Birnholtz, J. P. (2005). *When do researchers collaborate? toward a model of collaboration propensity in science and engineering research.* Unpublished doctoral dissertation, The University of Michigan.

Blevis, E. (2007). Sustainable interaction design: invention & disposal, renewal & reuse. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 503–512).

Bollen, J., Fox, G., & Singhal, P. R. (n.d.). How and where the TeraGrid supercomputing infrastructure benefits science. , *5*(1), 114–121. Retrieved 2015-05-01, from http://www.sciencedirect.com/science/article/pii/S1751157710000878

Burns, T. R., & Sthr, C. (n.d.). Power, knowledge, and conflict in the shaping of commons governance. the case of EU baltic fisheries. , *5*(2), 233–258. Retrieved 2015-05-01, from https://www.thecommonsjournal.org/index.php/ijc/article/view/260

Coase, R. H. (n.d.). The nature of the firm: influence. , 33–47. Retrieved 2015-05-01, from http://www.jstor.org/stable/765013

Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., Walters, L., & others. (n.d.). New goals for the US human genome project: 1998-2003. , *282*(5389), 682–689. Retrieved 2015-05-01, from http://www.sciencemag.org/content/282/5389/682.short

Collins, H. M. (n.d.). The meaning of data: Open and closed evidential cultures in the search for gravitational waves 1. , *104*(2), 293–338. Retrieved 2015-05-01, from http://www.jstor.org/stable/10.1086/210040

Commons, J. R. (n.d.). Problem of correlating law economics and ethics, the. , *8*, 3.

Contreras, J. L. (n.d.-a). Constructing the genome commons. Retrieved 2015-05-01, from http://papers.ssrn.com/sol3/papers.cfm?abstract$_i d$ = 2474405

Contreras, J. L. (n.d.-b). Data sharing, latency variables and science commons. , *25*, 1601. Retrieved 2015-04-29, from http://papers.ssrn.com/sol3/papers.cfm?abstract$_i d$ = 1501280

Contreras, J. L. (n.d.-c). Data sharing, latency variables and science commons. , *25*, 1601. Retrieved 2015-05-01, from http://papers.ssrn.com/sol3/papers.cfm?abstract$_i d$ = 1501280

Contreras, J. L. (n.d.-d). Prepublication data release, latency, and genome commons. , *329*(5990), 393–394. Retrieved 2015-04-29, from http://www.sciencemag.org/content/329/5990/393.short

Contreras, J. L. (n.d.-e). Prepublication data release, latency, and genome commons. , *329*(5990), 393–394. Retrieved 2015-05-01, from http://www.sciencemag.org/content/329/5990/393.short

Corbin, D. F. (n.d.). *A life of matthew fontaine maury.* S. Low, Marston, Searle, & Rivington.

Cox, M. (n.d.). Understanding large social-ecological systems: introducing the SESMAD project. , *8*(2), 265–276. Retrieved 2015-04-29, from https://www.thecommonsjournal.org/index.php/ijc/article/view/406

Cronin, B. (n.d.). The citation process. the role and significance of citations in scientific communication. , *1*. Retrieved 2015-05-01, from http://adsabs.harvard.edu/abs/1984cprs.book.....C

Dee, D. P., Balmaseda, M., Balsamo, G., Engelen, R., Simmons, A. J., & Thpaut, J.-N. (n.d.). Toward a consistent reanalysis of the climate system. , *95*(8), 1235–1248. Retrieved 2015-05-01, from http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-13-00043.1

Dietz, T., Ostrom, E., & Stern, P. C. (n.d.). The struggle to govern the commons. , *302*(5652), 1907–1912. Retrieved 2015-05-01, from `http://www.sciencemag.org/content/302/5652/1907.short`

Disco, N., & Kranakis, E. (n.d.). *Cosmopolitan commons: Sharing resources and risks across borders*. MIT Press.

Divoli, A., Nakov, P., & Hearst, M. A. (n.d.). Do peers see more in a paper than its authors? , *2012*. Retrieved 2015-05-01, from `http://www.hindawi.com/journals/abi/aip/750214/`

Dourish, P. (2010). Hci and environmental sustainability: the politics of design and the design of politics. In *Proceedings of the 8th acm conference on designing interactive systems* (pp. 1–10).

Edwards, P. N. (n.d.-a). Global climate science, uncertainty and politics: Data-laden models, model-filtered data. , *8*(4), 437–472. Retrieved 2015-05-01, from `http://www.tandfonline.com/doi/pdf/10.1080/09505439909526558`

Edwards, P. N. (n.d.-b). *A vast machine: Computer models, climate data, and the politics of global warming*. Mit Press.

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (n.d.-a). *Writing ethnographic fieldnotes*. University of Chicago Press.

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (n.d.-b). *Writing ethnographic fieldnotes*. University of Chicago Press.

Epstein, G., Prez, I., Schoon, M., & Meek, C. L. (n.d.). Governing the invisible commons: Ozone regulation and the montreal protocol. , *8*(2), 337–360. Retrieved 2015-05-01, from `http://www.thecommonsjournal.org/index.php/ijc/article/view/407`

Fegley, B. D., & Torvik, V. I. (n.d.). Has large-scale named-entity network analysis been resting on a flawed assumption? , *8*(7), e70299. Retrieved 2015-05-01, from `http://dx.plos.org/10.1371/journal.pone.0070299`

Fleming, J. R. (n.d.). *Fixing the sky: the checkered history of weather and climate control*. Columbia University Press.

Fletcher, J. O., Slutz, R. J., & Woodruff, S. D. (n.d.-a). Towards a comprehensive oceanatmosphere dataset. , *20*, 13–14.

Fletcher, J. O., Slutz, R. J., & Woodruff, S. D. (n.d.-b). Towards a comprehensive oceanatmosphere dataset. , *20*, 13–14.

Flyvbjerg, B. (n.d.). Five misunderstandings about case-study research. , *12*(2), 219–245. Retrieved 2015-05-01, from http://qix.sagepub.com/content/12/2/219.short

Frischmann, B. M. (n.d.). Infrastructure commons. , 121.

Fry, J., Schroeder, R., & den Besten, M. (n.d.). Open science in e-science: contingency or policy? , *65*(1), 6–32. Retrieved 2015-05-01, from http://www.emeraldinsight.com/doi/abs/10.1108/00220410910926103

Fry, J., & Talja, S. (n.d.-a). The cultural shaping of scholarly communication: Explaining e-journal use within and across academic fields. , *41*(1), 20–30. Retrieved 2015-05-01, from http://onlinelibrary.wiley.com/doi/10.1002/meet.1450410103/full

Fry, J., & Talja, S. (n.d.-b). The intellectual and social organization of academic fields and the shaping of digital resources. , *33*(2), 115–133. Retrieved 2015-03-25, from http://jis.sagepub.com/content/33/2/115 doi: 10.1177/0165551506068153

Garfield, E. (n.d.). *OBLITERATION PHENOMENON IN SCIENCE AND ADVANTAGE OF BEING OBLITERATED*. INST SCI INFORM INC 3501 MARKET ST, PHILADELPHIA, PA 19104.

Geels, F. W., & Schot, J. (2007). Typology of sociotechnical transition pathways. *Research policy*, *36*(3), 399–417.

Gibbons, M. (n.d.). Mode 2 society and the emergence of context-sensitive science. , *27*(3), 159–163. Retrieved 2015-05-01, from http://spp.oxfordjournals.org/content/27/3/159.short

Glnzel, W., & Schubert, A. (n.d.). Analysing scientific networks through co-authorship. In *Handbook of quantitative science and technology research* (pp. 257–276). Springer. Retrieved 2015-05-01, from http://link.springer.com/chapter/10.1007/1-4020-2755-9$_1$2

Hagstrom, W. (n.d.). 0. 1965. the scientific community. , *1968*, 1971.

Hardin, G. (n.d.). The tragedy of the commons. , *162*(3859), 1243–1248. Retrieved 2015-05-01, from http://www.sciencemag.org/content/162/3859/1243.full.pdfquot;gt;

Harvey, D. (2009). The art of rent: globalisation, monopoly and the commodification of culture. *Socialist Register*, *38*(38).

Haythornthwaite, C. (n.d.). Crowds and communities: Light and heavyweight models of peer production. In *System sciences, 2009. HICSS'09. 42nd hawaii international conference on* (pp. 1–10). IEEE. Retrieved 2015-05-01, from http://ieeexplore.ieee.org/xpls/abs$_all.jsp?arnumber = 4755627$

Heller, M. A., & Eisenberg, R. S. (n.d.). Can patents deter innovation? the anticommons in biomedical research. , *280*(5364), 698–701. Retrieved 2015-04-29, from http://www.sciencemag.org/content/280/5364/698.short

Hess, C., & Ostrom, E. (n.d.). A framework for analyzing the knowledge commons: a chapter from understanding knowledge as a commons: from theory to practice. Retrieved 2015-05-01, from http://surface.syr.edu/sul/21/

Holling, C. S. (n.d.-a). Resilience and stability of ecological systems. , 1–23. Retrieved 2015-04-29, from http://www.jstor.org/stable/2096802

Holling, C. S. (n.d.-b). Resilience and stability of ecological systems. , 1–23. Retrieved 2015-05-01, from http://www.jstor.org/stable/2096802

Howe, C. J., & Windram, H. F. (n.d.). Phylomemeticsevolutionary analysis beyond the gene. , *9*(5), e1001069. Retrieved 2015-05-01, from http://dx.plos.org/10.1371/journal.pbio.1001069

Imperial, M. T., & Yandle, T. (n.d.). Taking institutions seriously: using the IAD framework to analyze fisheries policy. , *18*(6), 493–509. Retrieved 2015-05-01, from http://www.tandfonline.com/doi/abs/10.1080/08941920590947922

Ingwersen, P., & Chavan, V. (n.d.). Indicators for the data usage index (DUI): an incentive for publishing primary biodiversity data through global information infrastructure. , *12*, S3. Retrieved 2015-05-01, from http://www.biomedcentral.com/1471-2105/12/S15/S3/

Jacobs, C. A., & Worley, S. J. (n.d.). Data curation in climate and weather: Transforming our ability to improve predictions through global knowledge sharing. , *4*(2), 68–79. Retrieved 2015-05-01, from http://sveetly.zapto.org/orange/pooxy4/nph-poxy.pl/jv/20/http/www.ijdc.net/index

Jansen, B. J. (n.d.). Search log analysis: What it is, what's been done, how to do it. , *28*(3), 407–432. Retrieved 2015-05-01, from http://www.sciencedirect.com/science/article/pii/S0740818806000673

Jirotka, M., Lee, C. P., & Olson, G. M. (n.d.). Supporting scientific collaboration: Methods, tools and concepts. , *22*(4), 667–715. Retrieved 2015-05-01, from `http://link.springer.com/article/10.1007/s10606-012-9184-0`

JOHN, R. R. (n.d.). *SPREADING THE NEWS*. Harvard University Press.

Kee, K. F., & Browning, L. D. (n.d.). The dialectical tensions in the funding infrastructure of cyberinfrastructure. , *19*(3), 283–308. Retrieved 2015-05-01, from `http://link.springer.com/article/10.1007/s10606-010-9116-9`

Kent, E. C., Woodruff, S. D., & Berry, D. I. (n.d.). Metadata from WMO publication no. 47 and an assessment of voluntary observing ship observation heights in ICOADS. , *24*(2), 214–234. Retrieved 2015-05-01, from `http://journals.ametsoc.org/doi/abs/10.1175/JTECH1949.1`

Kiser, E. O., Larry. (n.d.). The three worlds of action: A metatheoretical synthesis of institutional approaches.

Kittur, A., & Kraut, R. E. (n.d.). Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proceedings of the 2008 ACM conference on computer supported cooperative work* (pp. 37–46). ACM. Retrieved 2015-05-01, from `http://dl.acm.org/citation.cfm?id=1460572`

Knutti, R., Masson, D., & Gettelman, A. (n.d.). Climate model genealogy: Generation CMIP5 and how we got there. , *40*(6), 1194–1199. Retrieved 2015-05-01, from `http://onlinelibrary.wiley.com/doi/10.1002/grl.50256/full`

Lakhani, K. R., & Von Hippel, E. (n.d.). How open source software works:free user-to-user assistance. , *32*(6), 923–943. Retrieved 2015-05-01, from `http://www.sciencedirect.com/science/article/pii/S0048733302000951`

Lam, W. F. (n.d.). Governing irrigation systems in nepal. Retrieved 2015-05-01, from `http://agris.fao.org/agris-search/search.do?recordID=US201300029784`

Lessig, L. (n.d.). *Code and other laws of cyberspace*. Basic books.

Lewis, J. M. (n.d.). Winds over the world sea: Maury and kppen. , *77*(5), 935–952. Retrieved 2015-05-01, from

Leydesdorff, L. (n.d.). Theories of citation? , *43*(1), 5–25. Retrieved 2015-05-01, from `http://link.springer.com/article/10.1007/BF02458391`

Mace, R., & Holden, C. J. (n.d.). A phylogenetic approach to cultural evolution. , *20*(3), 116–121. Retrieved 2015-05-01, from http://www.sciencedirect.com/science/article/pii/S0169534704003933

Mace, R., Holden, C. J., & Shennan, S. (n.d.). *The evolution of cultural diversity: a phylogenetic approach*. Left Coast Press.

Madison, M. J. (n.d.). Commons at the intersection of peer production, citizen science, and big data: Galaxy zoo. Retrieved 2015-05-01, from http://papers.ssrn.com/sol3/papers.cfm?abstract$_i d = 2495404$

Madison, M. J., Frischmann, B. M., & Strandburg, K. J. (n.d.). Constructing commons in the cultural environment. , *95*, 657. Retrieved 2015-05-01, from http://papers.ssrn.com/sol3/papers.cfm?abstract$_i d = 1265793$

Mayernik, M. S., Callaghan, S., Leigh, R., Tedds, J., & Worley, S. (n.d.). Peer review of datasets: When, why, and how. Retrieved 2015-05-01, from http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-13-00083.1

McGinnis, M. D. (n.d.-a). An introduction to IAD and the language of the ostrom workshop: a simple guide to a complex framework. , *39*(1), 169–183. Retrieved 2015-05-01, from http://onlinelibrary.wiley.com/doi/10.1111/j.1541-0072.2010.00401.x/full

McGinnis, M. D. (n.d.-b). *Polycentric games and institutions: readings from the workshop in political theory and policy analysis*. University of Michigan Press.

McGinnis, M. D. (n.d.-c). *Polycentricity and local public economies: Readings from the workshop in political theory and policy analysis*. University of Michigan Press.

Merton, R. K. (n.d.-a). The self-fulfilling prophecy. , 193–210. Retrieved 2015-05-01, from http://www.jstor.org/stable/4609267

Merton, R. K. (n.d.-b). *The sociology of science: Theoretical and empirical investigations*. University of Chicago press.

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., ... others (n.d.). North american regional reanalysis. , *87*(3), 343–360. Retrieved 2015-05-01, from http://journals.ametsoc.org/doi/abs/10.1175/BAMS-87-3-343

Middleton, J. (n.d.). Merchants: An essay in historical ethnography*. , *9*(3), 509–526. Retrieved 2015-05-01, from http://onlinelibrary.wiley.com/doi/10.1111/1467-9655.00161/full

Miller, C. A., & Edwards, P. N. (n.d.). *Changing the atmosphere: expert knowledge and environmental governance.* MIT Press.

Mirowski, P. (n.d.). *Science-mart.* Harvard University Press.

Moore, S. A. (Ed.). (n.d.). *Pragmatic sustainability: Theoretical and practical tools* (1edition ed.). Routledge.

Moravcsik, M. J., & Murugesan, P. (n.d.). Some results on the function and quality of citations. , *5*(1), 86–92. Retrieved 2015-05-01, from http://sss.sagepub.com/content/5/1/86.full.pdf

Nakov, P. I., Schwartz, A. S., & Hearst, M. (n.d.). Citances: Citation sentences for semantic analysis of bioscience text. In *Proceedings of the SIGIR04 workshop on search and discovery in bioinformatics* (pp. 81–88). Retrieved 2015-05-01, from http://biotext.berkeley.edu/papers/citances-nlpbio04.pdf

NOAA. (n.d.). *Public notice: Termination of ICOADS development due to NOAA budget cuts.* Retrieved from http://www.esrl.noaa.gov/psd/data/gridded/pdf/ICOADS-notice-v4.pdf

O'Brien, M. J., Darwent, J., & Lyman, R. L. (n.d.). Cladistics is useful for reconstructing archaeological phylogenies: Palaeoindian points from the southeastern united states. , *28*(10), 1115–1136. Retrieved 2015-05-01, from http://www.sciencedirect.com/science/article/pii/S0305440301906818

Olson, M. (n.d.). The logic ofcollective action.

O'Mahony, S., & Ferraro, F. (n.d.). The emergence of governance in an open source community. , *50*(5), 1079–1106. Retrieved 2015-05-01, from http://amj.aom.org/content/50/5/1079.short

Ondrejka, C. (n.d.). *Maury, innovation and change.* Retrieved 2015-05-01, from http://ondrejka.net/history/2014/02/28/maury.html

on Science Policy, C.-D. W. G. (n.d.). CODATA newsletter no. 70. Retrieved 2015-04-06, from http://www.codata.info/resources/newsletters/nl72.html

Ortega, F. (n.d.). *Wikipedia: A quantitative analysis.* phdthesis. Retrieved 2015-05-01, from http://www.researchgate.net/profile/Felipe$_{O}rtega2/publication/200773248_{W}ikipedia_{A_q}u$

Ostrom, E. (n.d.-a). Beyond markets and states: polycentric governance of complex economic systems. , 641–672. Retrieved 2015-05-01, from http://www.jstor.org/stable/27871226

Ostrom, E. (n.d.-b). *Governing the commons: The evolution of institutions for collective action.* Cambridge university press.

Ostrom, E. (n.d.-c). *Unlocking public entrepreneurship and public economies* (No. 2005/01). WIDER Discussion Papers//World Institute for Development Economics (UNU-WIDER). Retrieved 2015-04-29, from http://www.econstor.eu/handle/10419/52899

Ostrom, E., Gardner, R., & Walker, J. (n.d.). *Rules, games, and common-pool resources.* University of Michigan Press.

Ostrom, E., Schroeder, L., Wynne, S., & others. (n.d.). *Institutional incentives and sustainable development: infrastructure policies in perspective.* Westview Press. Retrieved 2015-05-01, from http://www.cabdirect.org/abstracts/19936716128.html

OMahony, S., & Lakhani, K. R. (n.d.). Organizations in the shadow of communities. , *33*, 3–36. Retrieved 2015-05-01, from http://www.emeraldinsight.com/doi/abs/10.1108/S0733-558X(2011)0000033004

Page, R. D., & Holmes, E. C. (n.d.). *Molecular evolution: a phylogenetic approach.* John Wiley & Sons.

Parker, D. (n.d.). Second JCOMM workshop on advances in marine climatology. , *53*(2).

Peritz, B. C. (n.d.). A classification of citation roles for the social sciences and related fields. , *5*(5), 303–312. Retrieved 2015-05-01, from http://link.springer.com/article/10.1007/BF02147226

Peroni, S., & Shotton, D. (n.d.). FaBiO and CiTO: ontologies for describing bibliographic resources and citations. , *17*, 33–43. Retrieved 2015-05-01, from http://www.sciencedirect.com/science/article/pii/S1570826812000790

Polanyi, M. (n.d.). *The logic of liberty.* University of Chicago Press Chicago.

Rasmussen, J. (n.d.). Historical developmetn of the world weather watch. Retrieved 2015-04-05, from https://www.wmo.int/pages/themes/oceans/www$_{e}n.htmlone$

Rayner, N. A., Brohan, P., Parker, D. E., Folland, C. K., Kennedy, J. J., Vanicek, M., ... Tett, S. F. B. (n.d.). Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: the HadSST2 dataset. , *19*(3), 446–469. Retrieved 2015-05-01, from http://journals.ametsoc.org/doi/abs/10.1175/JCLI3637.1

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., ... Kaplan, A. (n.d.). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. , *108*. Retrieved 2015-05-01, from http://onlinelibrary.wiley.com/doi/10.1029/2002JD002670/full

Reichman, J. H., & Uhlir, P. F. (n.d.). A contractually reconstructed research commons for scientific data in a highly protectionist intellectual property environment. , 315–462. Retrieved 2015-05-01, from http://www.jstor.org/stable/20059181

Rexov, K., Frynta, D., & Zrzav, J. (n.d.). Cladistic analysis of languages: Indo-european classification based on lexicostatistical data. , *19*(2), 120–127. Retrieved 2015-05-01, from http://onlinelibrary.wiley.com/doi/10.1111/j.1096-0031.2003.tb00299.x/abstract

Ribes, D., & Finholt, T. A. (n.d.). The long now of technology infrastructure: articulating tensions in development. , *10*(5), 5. Retrieved 2015-05-01, from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1507context=jais

Rivers, C. M., Lofgren, E. T., Marathe, M., Eubank, S., & Lewis, B. L. (n.d.). Modeling the impact of interventions on an epidemic of ebola in sierra leone and liberia. Retrieved 2015-04-29, from http://arxiv.org/abs/1409.4607

Rose, C. (n.d.). The comedy of the commons: custom, commerce, and inherently public property. , 711–781. Retrieved 2015-05-01, from http://www.jstor.org/stable/1599583

Samuelson, P. A. (n.d.). The problem of integrability in utility theory. , 355–385. Retrieved 2015-05-01, from http://www.jstor.org/stable/2549499

Sarewitz, D. (n.d.). *Frontiers of illusion: Science, technology, and the politics of progress.* Temple University Press.

Schmidt, B. (n.d.). *Sapping attention: Reading digital sources: a case study in ship's logs.* Retrieved 2015-05-01, from http://sappingattention.blogspot.com/2012/11/reading-digital-sources-case-study-

Schweik, C. M., & English, R. C. (n.d.). *Internet success: a study of open-source software commons.* MIT Press.

Service, N. N. W. (n.d.). *Gateway history - the story on the growth of the main telecommunication switching system - NOAA's national weather service.* Retrieved 2015-05-01, from http://www.nws.noaa.gov/tg/histgate.php

Shaw, A., & Hill, B. M. (n.d.). Laboratories of oligarchy? how the iron law extends to peer production. , *64*(2), 215–238. Retrieved 2015-05-01, from http://onlinelibrary.wiley.com/doi/10.1111/jcom.12082/full

Shotton, D., & others. (n.d.). CiTO, the citation typing ontology. , *1*, S6. Retrieved 2015-05-01, from http://www.biomedcentral.com/content/pdf/2041-1480-1-S1-S6.pdf

Simon, H. A. (n.d.). *The sciences of the artificial, 1981.* MIT Press.

Slutz, R. J., Lubker, S. J., Hiscox, J. D., Woodruff, S. D., Jenne, R. L., Joseph, D. H., ... Elms, J. D. (n.d.). Comprehensive ocean-atmosphere data set: release 1.

Small, H. G. (n.d.). Cited documents as concept symbols. , *8*(3), 327–340. Retrieved 2015-05-01, from http://sss.sagepub.com/content/8/3/327.short

Smith, A., Stirling, A., & Berkhout, F. (n.d.). The governance of sustainable socio-technical transitions. , *34*(10), 1491–1510. Retrieved 2015-05-01, from http://www.sciencedirect.com/science/article/pii/S0048733305001721

Stake, R. E., & Savolainen, R. (n.d.). *The art of case study research* (Vol. 95004979). Sage publications Thousand Oaks, CA.

Star, S. L., & Ruhleder, K. (n.d.). Steps toward an ecology of infrastructure: Design and access for large information spaces. , *7*(1), 111–134. Retrieved 2015-05-01, from http://pubsonline.informs.org/doi/abs/10.1287/isre.7.1.111

Strandburg, K. J., Frischmann, B. M., & Cui, C. (n.d.). The rare diseases clinical research network and the urea cycle disorders consortium as nested knowledge commons. , 14–43. Retrieved 2015-05-01, from http://papers.ssrn.com/sol3/papers.cfm?abstract$_i d = 2538609$

Taylor, M. (n.d.). Cooperation and rationality: Notes on the collective action problem and its solutions. , 222–249.

Tilmes, C., Fox, P., Ma, X., McGuinness, D., Privette, A., Smith, A., ... Zheng, J. G. (n.d.). Provenance representation for the national climate assessment in the global change information system. , *51*(11), 5160–5168. doi: 10.1109/TGRS.2013.2262179

Tomlinson, B., Blevis, E., Nardi, B., Patterson, D. J., Silberman, M., & Pan, Y. (2013). Collapse informatics and practice: Theory, method, and design. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *20*(4), 24.

Trenberth, K. E., Stepaniak, D. P., Hurrell, J. W., & Fiorino, M. (n.d.). Quality of reanalyses in the tropics. , *14*(7), 1499–1510.

US Department of Commerce, N. (n.d.). *History of the national weather service.* Retrieved 2015-04-04, from http://www.weather.gov/timeline

Vaughan, D. (n.d.-a). NASA revisited: Theory, analogy, and public sociology1. , *112*(2), 353–393. Retrieved 2015-05-01, from http://www.jstor.org/stable/10.1086/506413

Vaughan, D. (n.d.-b). Theorizing disaster analogy, historical ethnography, and the challenger accident. , *5*(3), 315–347. Retrieved 2015-05-01, from http://eth.sagepub.com/content/5/3/315.short

Vertesi, J., & Dourish, P. (n.d.). The value of data: considering the context of production in data economies. In *Proceedings of the ACM 2011 conference on computer supported cooperative work* (pp. 533–542). ACM. Retrieved 2015-05-01, from http://dl.acm.org/citation.cfm?id=1958906

Walker, B., Gunderson, L., Kinzig, A., Folke, C., Carpenter, S., & Schultz, L. (n.d.). A handful of heuristics and some propositions for understanding resilience in social-ecological systems. , *11*(1), 13. Retrieved 2015-05-01, from http://dlc.dlib.indiana.edu/dlc/handle/10535/2570

Woodruff, S. D., Slutz, R. J., Jenne, R. L., & Steurer, P. M. (n.d.). A comprehensive ocean-atmosphere data set. , *68*(10), 1239–1250.

Yin, R. K. (n.d.). *Case study research: Design and methods.* Sage publications.