

IMPACT OF E-INFRASTRUCTURE STIMULUS ON THE BIODIVERSITY  
SCIENCE DISCIPLINE: AN EMPIRICAL INVESTIGATION

Julio Eligio Ibarra

## Thesis Committee Members:

Chairman:	Prof. Dr. R. A. Wessel	Univ. Twente, MB
Secretary:	Prof. Dr. R. A. Wessel	
Promoter:	Prof. Dr. Kuldeep Kumar	Univ. Twente, MB
Other Committee Members:	Prof. Dr. Jos van Hillegersberg	Univ. Twente, MB
	Prof. Dr. Ir L. J. M. Nieuwenhuis	Univ. Twente, MB
	Prof. Dr. Roland M. Müller	Berlin School of Economics and Law
	Prof. Dr. Evelyn E. Gaiser	Florida International University
	Prof. Dr. Jacob de Vlieg	Radboud Univ. Nijmegen / Netherlands e-Science Centre

# CTIT

---

CTIT Ph.D. Thesis Series No. 12-227  
Centre for Telematics and Information Technology (CTIT)  
P.O. Box 217, 7500 AE  
Enschede, The Netherlands

ISBN: 978-0-615-66002-8  
ISSN: 1381-3617 (CTIT Ph.D. thesis series number 12-227)

Cover Design: Delaney-Designs. Rochester, NH -3867  
Printed by: Wohrman Print Service, Zutphen, The Netherlands

Copyright © 2012, Julio E. Ibarra, All rights reserved.

IMPACT OF E-INFRASTRUCTURE STIMULUS ON THE BIODIVERSITY  
SCIENCE DISCIPLINE: AN EMPIRICAL INVESTIGATION

DISSERTATION

to obtain

the degree of doctor at the University of Twente,

on the authority of the rector magnificus,

Prof.dr. H. Brinksma,

on account of the decision of the graduation committee,

to be publicly defended

on Wednesday the 4<sup>th</sup> of July 2012 at 16:45

by

Julio Eligio Ibarra

Born on the 15<sup>th</sup> of January 1959

in Havana, Cuba



For Therry,

*In appreciation for starting me on a life-long journey of learning.*

## Acknowledgments

Pursuing doctoral research alongside a full-time and demanding job has been a challenge. This dissertation would not have become a reality without the persistent and generous encouragement of many people.

Heidi Alvarez and Chip Cox, my dear colleagues and friends, thank you for your patience and unwavering support all these years. To our staff at CIARA, thank you for keeping the Center operating while I focused on writing this dissertation.

To my parents Juliano and Luisa Ibarra, I honor you both for the love and support you have provided me throughout this journey.

To Suzy Girard-Ruttenberg, thank you for the coaching that helped put plans into action. To Steve Vogel and Suzy, I thank you so much for your wonderful editing. To Juliana Morgan, thank you for your helpful research assistance.

I am very grateful to Prof. Dr. Jos van Hillegersberg and the University of Twente for providing a home for my dissertation. To the members of my doctoral committee, thank you for your thoughtful comments that have guided me in raising this research to a higher level of quality.

To the scientists with whom I've had the privilege to discuss my research and the Program Directors at the National Science Foundation, thank you for sharing your knowledge and giving me time to ask many questions, which ultimately helped me learn what I needed to write this dissertation.

To Prof. Kuldeep Kumar — research advisor, mentor and friend — there is no scale by which to measure your generosity and patience. Although we both were challenged by our busy schedules, we found time to meet in many cities, share good food and wine, and discuss this research. Prof. Kumar, you gave me the guidance and the time that enabled me to learn how to think about and how to do research. Your exemplary scholarship and service impressed upon me the importance to pass onto others the knowledge you shared with me. I also want to take this opportunity to thank Prof. Kumar's wife, Veronica Kumar, for her hospitality and kindness when I visited their home.

Lastly, I want to thank my dear friend, Jane Cameron. At the start of 2012, Prof. Kumar challenged me to finish this dissertation within six months. Jane, you agreed to help me finish by freeing me of day-to-day responsibilities, so that I could focus on my research. Your support created a tipping point that accelerated my progress dramatically which then allowed me to complete the milestones of this research.



## Table of Contents

<b>Acknowledgments</b> .....	<b>vi</b>
<b>Table of Contents</b> .....	<b>viii</b>
<b>Table of Figures</b> .....	<b>xii</b>
<b>Table of Tables</b> .....	<b>xiii</b>
<b>Chapter 1</b> .....	<b>1</b>
<b>1. Introduction</b> .....	<b>1</b>
<b>1.1 e-Infrastructure Development: Stimuli aimed at dramatic improvements in Scientific Progress</b> .....	<b>2</b>
<b>1.2 What’s the Problem?</b> .....	<b>6</b>
<b>1.3 Objective of the Research</b> .....	<b>8</b>
<b>1.4 Research Questions</b> .....	<b>8</b>
<b>1.5 Significance of Increasing Understanding of e-Infrastructure Development and Its Impact on Scientific Progress</b> .....	<b>10</b>
<b>1.6 Relevance and Potential Contribution</b> .....	<b>12</b>
<b>1.7 Roadmap and Organization of this Thesis</b> .....	<b>13</b>
<b>Chapter II Theory Construction: Literature Review</b> .....	<b>16</b>
<b>2. Literature Review</b> .....	<b>16</b>
<b>2.1 e-Infrastructure Development</b> .....	<b>19</b>
2.1.1 Infrastructure .....	24
2.1.1.1 Artifacts.....	25
2.1.1.2 Hard and Soft Infrastructures.....	26
2.1.1.3 Concepts and Patterns of Infrastructure Development.....	27
2.1.1.3.1 Large Technological Systems:.....	27
2.1.1.3.2 Substrate and Relational Properties of Infrastructure .....	32
2.1.1.3.3 Domains.....	33
2.1.2 Summary: e-Infrastructure Development.....	34
<b>2.2 Science Discipline: Properties and Concepts</b> .....	<b>36</b>
2.2.1 Community of Scientists .....	37
2.2.2 Methodology .....	38
2.2.2.1 Phenomena and Methodology.....	40
2.2.2.2 Paradigms.....	43
2.2.3 Problems and Puzzles.....	46
2.2.3.1 Normal (Evolutionary) Science.....	48
2.2.3.2 Extraordinary Science.....	50
2.2.3.3 Paradigm Shift.....	53
2.2.3.4 Scientific Revolution .....	54
2.2.4 Resources of a Science Discipline .....	56
2.2.4.1 Human Resources .....	57
2.2.4.2 Knowledge Resources.....	58
2.2.5 Summary: Science Discipline Properties and Patterns .....	59
<b>2.3 Integration of Concepts</b> .....	<b>60</b>
<b>CHAPTER III Theory Construction: Theoretical Underpinnings</b> .....	<b>62</b>
<b>3. Theoretical Underpinnings for This Study</b> .....	<b>62</b>



<b>3.1 Properties of Science Disciplines: A Philosophy of Science Perspective .....</b>	<b>63</b>
3.1.1 How Is Knowledge Created? .....	63
3.1.1.1 Knowledge as an Activity.....	64
3.1.1.2 Knowledge as Potential.....	65
3.1.2 Modes of Inquiry to Produce Knowledge .....	66
3.1.2.1 Theory Leads, Data Follows — The Leibnizian Mode of Inquiry .....	66
3.1.2.2 Data Leads, Theory Follows — The Lockean Mode of Inquiry.....	67
3.1.2.3 Theory and Data Shape Each Other: Singerian Form of Inquiry .....	68
<b>3.2 Co-Evolution Theory .....</b>	<b>68</b>
3.2.1 Biological Co-Evolution and Natural Selection .....	69
3.2.1.1 Biological Co-Evolution .....	70
3.2.1.2 Population Ecology Model (Natural Selection) .....	71
3.2.2 Relationship Between a Science Discipline and e-Infrastructure Discussion ....	75
3.2.3 Adaptive Structuration Theory.....	77
3.2.4 Resource Dependence Theory .....	79
<b>3.3 Summary of the Literature.....</b>	<b>82</b>
<b>CHAPTER IV: Conceptual Framework.....</b>	<b>87</b>
<b>4. Concept Map .....</b>	<b>87</b>
<b>4.1 Introduction to the Concept Map .....</b>	<b>88</b>
<b>4.2 Concept Map Components: Concepts and Operational Definitions.....</b>	<b>91</b>
4.2.1 Aspects of a Science Discipline .....	91
4.2.2 Environment of a Science Discipline .....	97
4.2.3 e-Infrastructure Development Process.....	100
4.2.4 Co-evolution Relationship of a Science Discipline and e-Infrastructure Development.....	107
4.2.5 Scientific Discovery Component.....	108
4.2.6 ICT Investments Stimulus Component.....	109
<b>4.3 Summary of the Concept Map.....</b>	<b>110</b>
<b>CHAPTER V: Research Design.....</b>	<b>116</b>
<b>5. Research Design.....</b>	<b>116</b>
<b>CHAPTER VI: Research Methodology.....</b>	<b>120</b>
<b>6. Research Methodology .....</b>	<b>120</b>
<b>6.1 Research Stance: Interpretivism.....</b>	<b>120</b>
<b>6.2 Connecting with a Research Paradigm: Interpretive Research.....</b>	<b>121</b>
<b>6.3 Case Study Research Methodology .....</b>	<b>121</b>
<b>6.4 Unit of Analysis .....</b>	<b>123</b>
<b>6.5 Data Gathering .....</b>	<b>124</b>
6.5.1 Methods for Data Collection.....	124
6.5.2 Sources for Data Collection .....	127
6.5.2.1 Informants.....	127
6.5.2.2 Documents as Sources of Data .....	127
<b>6.6 Data Analysis .....</b>	<b>130</b>
6.6.1 Recording Field Interviews and Other Data.....	131
6.6.2 Analyzing Field Interviews and Data .....	132
6.6.3 Developing the Iterative Process Between Field Data and Theory.....	134
<b>6.7 Summary .....</b>	<b>136</b>
<b>CHAPTER VII: EMPIRICAL RESEARCH.....</b>	<b>137</b>

<b>7. Case Studies: Introduction.....</b>	<b>137</b>
<b>8. Biodiversity Case: The U.S. Long Term Ecological Research Network.....</b>	<b>144</b>
<b>8.1 Setting the Case .....</b>	<b>144</b>
8.1.1 What is Biodiversity? .....	144
8.1.2 The U.S. Long Term Ecological Research Network .....	145
8.1.2.1 Characteristics of the US-LTER Network.....	147
<b>8.2 Sources of Information for Studying e-Infrastructure Development and Biodiversity and Ecological Research in the US-LTER Network.....</b>	<b>149</b>
8.2.1 Source Documents.....	149
8.2.2 Informants .....	151
<b>8.3 Data Analysis: Linking Data and the Conceptual Framework.....</b>	<b>152</b>
8.3.1 Stimulus: ICT Investments.....	153
8.3.2 Participation in the LTER Network.....	158
8.3.3 E-Infrastructure Development Process and the LTER Network .....	160
8.3.4 E-Infrastructure Development Process: Constructing the Data Infrastructure	172
8.3.5 E-Infrastructure Development Process: Constructing a Data Infrastructure...176	
8.3.6 Data Sharing and Governance to Support Data e-Infrastructure.....	181
8.3.6.1 Events That Gave Rise to Data Sharing and Its Acceptance in the LTER Network	182
8.3.6.2 Governance Structure of the LTER Network.....	183
<b>8.4 Findings .....</b>	<b>185</b>
8.4.1 Revision to conceptual framework .....	189
8.4.2 Conclusion: Mixed Results .....	192
<b>8.5 Literature Revisited .....</b>	<b>193</b>
8.5.1 IT Productivity Paradox.....	193
8.5.2 Impact of ICT on Scientists' Productivity .....	198
<b>9. Genomics Case .....</b>	<b>201</b>
<b>9.1 Introduction .....</b>	<b>201</b>
<b>9.2 Confirming Genomics as a Micro-level Discipline Connected to Biodiversity</b>	<b>203</b>
9.2.1 Species Classification and Discovery.....	203
9.2.2 Case Example: Connecting Genomics to Biodiversity via Data.....	205
<b>10. Consolidation of Findings and Conclusion .....</b>	<b>214</b>
<b>10.1 Discussion of the Findings .....</b>	<b>215</b>
10.1.1 ICT Investment Stimuli Impacts Technology Infrastructure Development ...215	
10.1.2 Technology Infrastructure Impacts the Growth Rate of Data .....	216
10.1.3 Technology and Data Infrastructures Influence How Science Is Practiced ...217	
10.1.4 Data Infrastructure influences Change in Socio-organizational Infrastructure	218
10.1.5 Impact to Scientific Progress: Mixed Results .....	219
10.1.6 Data Sharing Across Disciplines .....	221
<b>10.2 Contributions of the Research.....</b>	<b>222</b>
10.2.1 Contributions to Theory.....	222
10.2.2 Contributions to Practice.....	223
<b>10.3 Limitations of the Study.....</b>	<b>223</b>
<b>10.4 Future Research Directions .....</b>	<b>224</b>
<b>10.5 Overall Conclusions .....</b>	<b>225</b>
<b>References .....</b>	<b>226</b>

**About the Cover.....243**

## Table of Figures

<i>Figure 1 Organization of the Thesis.....</i>	<i>14</i>
<i>Figure 2 Conceptualization of e-Infrastructure Development.....</i>	<i>61</i>
<i>Figure 3 Literature Map of major literature streams.....</i>	<i>84</i>
<i>Figure 4 Conceptualization of co-evolution between e-Infrastructure development and Aspects of a Science Discipline.....</i>	<i>85</i>
<i>Figure 5 Concept Map of our Theoretical Framework.....</i>	<i>88</i>
<i>Figure 6 Major components of this study based upon the Interactive Research Design Model.....</i>	<i>119</i>
<i>Figure 7 U.S. federal investment in information and networking technology over 20 years.....</i>	<i>141</i>
<i>Figure 8 Intuition of ICT investment as a stimulus.....</i>	<i>142</i>
<i>Figure 9 Stimulus acting upon Site infrastructure to produce sharable data sets.....</i>	<i>159</i>
<i>Figure 10 E-Infrastructure Development Process: Enhancing LTER data, steps 1-4.....</i>	<i>174</i>
<i>Figure 11 E-Infrastructure Development Process: Enhancing LTER data, steps 5-10.....</i>	<i>175</i>
<i>Figure 12 Access to data sets and e-Infrastructure development process.....</i>	<i>177</i>
<i>Figure 13 Data Access and e-Infrastructure development process.....</i>	<i>179</i>
<i>Figure 14 Concept Map of findings in decade I.....</i>	<i>187</i>
<i>Figure 15 Complementarities of technology and socio-organizational sides of e-Infrastructure development.....</i>	<i>188</i>
<i>Figure 16 Concept Map with clustering of categories for e-Infrastructure Development Process.....</i>	<i>189</i>
<i>Figure 17 Revised Concept Map: Concept Map 2.....</i>	<i>191</i>
<i>Figure 18 Revised Concept Map with Productivity variables: Concept Map 3.....</i>	<i>200</i>
<i>Figure 20 Representation of connections between biodiversity and genomics using revised Concept Map.....</i>	<i>208</i>
<i>Figure 21 from S D Kahn, Science 2001; 331:728-729.....</i>	<i>209</i>
<i>Figure 22 Macro-Micro level relationship: Concept Map 4.....</i>	<i>212</i>

## Table of Tables

<i>Table 1 Research Questions.....</i>	<i>16</i>
<i>Table 2 Concepts and Patterns of Infrastructure Development.....</i>	<i>35</i>
<i>Table 3 Properties and Patterns of a Science Discipline.....</i>	<i>60</i>
<i>Table 4 Key Concepts, their Definitions and Observations.....</i>	<i>115</i>
<i>Table 5 IT Productivity Paradox variables and relationship to Concept Map.....</i>	<i>220</i>



## Chapter 1

### 1. Introduction

*— No matter what kind of challenge lies before you, if somebody believes in you, and you believe in your dream, it can happen. —*

*Tiffany Loren Rowe*

Nations around the world are increasingly concerned about their capabilities to innovate and compete in the changing global economy. Chief among those is the United States, whose status as the world leader in technology and the planet's dominant economic power is at risk. The National Science Foundation (NSF) raised this concern to the President's Council of Advisors on Science and Technology (PCAST, 2004):

*“Civilization is on the brink of a new industrial order. The big winners in the increasingly fierce global scramble for supremacy will not be those who simply make commodities faster and cheaper than the competition. They will be those who develop talent, techniques and tools so advanced that there is no competition.”*

Progress in science research and innovation has been recognized as central to achieving any nation's most critical goals, including raising living standards, creating good jobs, ensuring national security, strengthening education, improving public health, and protecting the environment (NAP, 1999; NAP, 2007).

Achieving dramatic advances in scientific progress will be critical to the U.S. and other leading nations, *if* they are going to prevail against rising competition and fierce economic rivalries. But what type of scientific progress has to be made in order to substantially impact the U.S. economy and support its global leadership position?

## 1.1 e-Infrastructure Development: Stimuli aimed at dramatic improvements in Scientific Progress

A 2007 report by the U.S. National Science Board (NSB, 2007) defined all scientific progress that enables economic growth as one of two types: *Evolutionary* or *Revolutionary*.

*Evolutionary progress* is evidenced by incremental advances in scientific understanding that builds upon the results of prior scientific knowledge. Using hypotheses and theories based upon a prevailing paradigm, evolutionary progress serves to *refine* the acceptance of existing hypotheses and theories, and therefore extends the lives of paradigms. The 2007 NSB report recognizes that the vast majority of research conducted in scientific laboratories around the world fosters evolutionary scientific progress.

*Revolutionary progress*, by contrast, takes place when scientific understanding advances dramatically, increasing the rate of discovery of new ideas, solutions and systems. The 2007 NSB report recognizes this phenomenon as "revolutionary" because it "transforms science by overthrowing entrenched paradigms and generating new ones." When this occurs, it is an opportunity for more rapid innovation and the most powerful economic development and growth.

Driving revolutionary progress is *transformative research*, a disruptive style of research. Transformative research is also widely viewed as key to the future of the U.S. continuing in its role as a leading global economic power.

The 2007 NSB report defines transformative research as "research driven by ideas that have the potential to radically change our understanding of an important existing scientific or engineering concept or leading to the creation of a new paradigm or field of science or engineering." Transformative research aims to increase revolutionary discoveries through the application of unconventional or radical approaches to actual problems and scientific puzzles (NSB, 2007). The desired effect of transformative research is to create the conditions that will achieve the kinds of discoveries that yield the greatest returns (NAP, 2007).



For example, when scientists and engineers discovered a solution to the limit of transistors<sup>1</sup> on an integrated circuit because of overheating, it enabled entrepreneurs to replace tape recorders with iPods, maps with global positioning systems, pay phones with cell phones, two-dimensional X-rays with three-dimensional CT scans, paperbacks with electronic books, slide rules with computers, and much more (NAP, 2010). Over time, this breakthrough innovation on an integrated circuit helped to create new industries and new infrastructure for the creation of new products.

Evolving in response to the requirements of transformative research was the phenomena of *Cyberinfrastructure* and *e-Science*. These government-funded initiatives — Cyberinfrastructure (NSF, 2007) in the U.S., and e-Science (Jankowski and Caldas, 2004) in the United Kingdom and European Union — share in common the notion of an advanced socio-technical substrate layer upon which transformative research can be enabled (Atkins et al, 2003). Moreover, they share a common vision of developing enabling infrastructure to support next-generation science, resulting in technological innovation and economic development.

The terms Cyberinfrastructure and e-Science emerged in the early 2000s to refer to a socio-technological infrastructure that integrated information and communications technologies (ICT) with human resources and organizations. This infrastructure was designed for the creation, dissemination and preservation of data, information and knowledge in the “digital age” (Atkins et al, 2003).

e-Infrastructure is yet another term that’s used in a similar manner as Cyberinfrastructure and e-Science, but with emphasis on the creation of national- or regional-scale infrastructures built upon existing ICT resources, such as national research and education networks, computing resources at supercomputing centers, data archives, etc.

---

<sup>1</sup> In 1971, the Intel 4004 Processor had 2300 transistors ([http://download.intel.com/pressroom/kits/events/moores\\_law\\_40th/MLTimeline.pdf](http://download.intel.com/pressroom/kits/events/moores_law_40th/MLTimeline.pdf)). In 2009, Intel released the Xeon® ‘Nehalem-EX’ Processor with 2.3 billion transistors (<http://www.intel.com/pressroom/archive/releases/2009/20090526comp.htm>).

e-Infrastructure is also referred to as “federated infrastructure” (e-IRGSP, 2005). Federated infrastructure normally refers to the sharing of resources owned and controlled by different organizations (including virtual organizations) that have agreed to federate. For example, in the U.S., Open Science Grid (OSG) users are able to share TeraGrid resources through an agreement implemented via a gateway system (Cummings et al, 2008). We found the term “e-Infrastructure” used mostly to describe Technological Infrastructure initiatives in Europe.

Separately and together, Cyberinfrastructure, e-Science and e-Infrastructure are viewed as investment worthy initiatives for those nations who wish to drive revolutionary scientific progress and stimulate national leadership in technological innovation, and economic development.

For example, in the U.S., investments in Cyberinfrastructure development initiatives approximate \$3.35 billion in a period of 9 years.

We refer to *e-Infrastructure development* as a process, consisting of stimuli of ICT investments towards creating national- or regional-scale federated infrastructure, aimed at increasing revolutionary scientific progress.

We will refer to *e-Infrastructure* as an object, or artifact, that embodies national- or regional-scale federated infrastructure that is the result of an e-Infrastructure development process. For example, we would use the term e-Infrastructure to characterize the outcome of an initiative in Europe to develop a new national-scale infrastructure towards enhancing multidisciplinary comparative research. Similarly, we may use the term “cyberinfrastructure” to describe a comparable initiative in the U.S., because the use of these terms tend to be tied to national initiatives.

For the remainder of this chapter, we will subsume Cyberinfrastructure and e-Science terms under e-Infrastructure. The terms “e-Infrastructure”, “Cyberinfrastructure” and “e\_Science” are given a more descriptive treatment in Chapter 2.

Since the Industrial Revolution, nations have gone through eras of development of various different infrastructures: railroads, telephone and telegraph networks, power and light networks, highway and public works systems, the Internet, among others. While these eras overlap, and the development of various infrastructures reinforce each other, they are often examined and described separately (Friedland, 1985). Infrastructure development of this type normally brings together public/private investments to stimulate growth and create demand that will, in turn, result in further accelerating growth.

Railroads, for example, profoundly affected the development of the U.S. as a nation during the 1850s, when the country was experiencing enormous geographic, demographic, social, and economic growth (Friedlander, 1985). Infrastructure development of a national railroad system leveraged public/private investments, and in so doing, became the dominant element of the national transportation system. It's a tried and true pattern: Growth attracts investment that fuels demand that spurs more growth.

e-Infrastructure development is for a nation's knowledge economy what infrastructure development was for an industrial economy (Atkins et al, 2003).

With investments in e-Infrastructure comes the expectation of high-risk, high-impact research, leading towards achievements of breakthrough discoveries (NSB, 2007; Atkins et al, 2003). The hope is that with these investments scientists will have access to new technologies and instruments that will lead to dramatic advances in scientific discovery (Bell et al, 2005; Anderson, 2003; NSF, 2007).

Capitalizing on such discoveries, nations would then be in a stronger position to compete and create opportunities for innovation. That's the hope. However, whether e-Infrastructure development leads to revolutionary progress — whether the hope and promise will match real returns — remains uncertain.

## 1.2 What's the Problem?

Based on literature and preliminary observation, there are two issues that call into question whether e-Infrastructure development will make, or is making, a transformative impact on scientific progress:

- (1) There is currently a lack of knowledge about how the development of e-Infrastructure is impacting scientific discovery;
- (2) We lack knowledge about how the problems and puzzles of a science discipline shape the development of e-Infrastructure, and conversely, how e-Infrastructure changes the problems and puzzles of a science discipline.

**The first issue** concerns return on investment. From a science policy perspective, nations are embarking in “transformative research” initiatives that supposedly introduce technology stimuli to science disciplines that hopefully may result in dramatic scientific progress. The U.S. National Science Board (NSB, 2007) made the following policy recommendation to the National Science Foundation:

*“That NSF develop a distinct, Foundation-wide Transformative Research Initiative (TRI) distinguishable by its potential impact on prevailing paradigms and by the potential to create new fields of science, to develop new technologies, and to open new frontiers.”*

From an infrastructure development perspective, nations are making investments in large-scale infrastructures with the hope of stimulating growth, achieving greater efficiencies and gaining a decent return in terms of scientific progress. In the case of e-Infrastructure development, the hope is to achieve scientific progress that results in breakthrough discoveries and innovations. However, it is not clear when or if these investments will result in moving scientific progress from mostly evolutionary to a revolutionary phase, where a nation could potentially achieve the greatest return.

Woolgar and Coopmans (2005) found that while much has been said about the likely effects of e-Infrastructures, not enough is known about their use and

effectiveness across science disciplines. Moreover, they emphasize that the nature and direction of change brought about by e-Infrastructures can be unpredictable. Woolgar's and Coopmans' argument is consistent with findings from information system (IS) researchers on the usability of advanced information technologies and user behavior (DeSanctis and Poole, 1994): "Actual behavior in the context of advanced technologies frequently differs from the intended impacts (Kiesler, 1986; Markus and Robey, 1988; Siegel, Dubrovsky, Kiesler and McGuire, 1986)."

**The second issue** looks at the phenomenon involving the interaction of two dynamic ecosystems: a science discipline and an e-Infrastructure. Imbalances could emerge as a result of the interactions between these two dynamic ecosystems. A science discipline is a dynamic ecosystem because it evolves as it works on its problems and puzzles (Graham et al, 2002). It also has a socio-technical infrastructure consisting of a community of scientists, knowledge and human resources (Kuhn, 1996; Graham et al, 2002).

We view e-Infrastructure as a dynamic environment because, on the one hand, it can emerge as part of a science discipline through the application of new instruments and technologies on problems and puzzles. For example, e-Infrastructure aims at providing scientists with a capability to resolve an anomaly between a hypothesis-driven experiment and empirical data. In this case, the e-Infrastructure we refer to comes from *within* science. On the other hand, e-Infrastructure can be introduced as a technology-led *intervention*, which potentially evolves into an imbalance in the discipline. Schroeder and Fry (2007) warn of potential imbalances occurring when social aspects of a science discipline are not taken into account in large-scale and complex technology-driven projects. Effects from interactions of a science discipline and e-Infrastructure — on both evolutionary and revolutionary progress — are not well understood.

**In summary**, we have raised two problematic issues concerning e-Infrastructures and their potential impact on science disciplines. While investments in e-Infrastructures continue to play a significant role as a stimulus towards increasing

transformative research, studies to understand the effectiveness of these investments are few or do not yet exist.

### 1.3 Objective of the Research

**The primary objective** of this study is to:

*Understand how the development of e-Infrastructure is impacting scientific discovery.*

**The secondary objective** of this study is to:

*Understand how the problems and puzzles of a science discipline shape the development of e-Infrastructure, and conversely, how e-Infrastructure influences the problems and puzzles of a science discipline.*

Both objectives are designed to provide insights and greater understanding into how the process of developing e-Infrastructure and the e-Infrastructure itself are impacting scientific progress. Moreover, we want to understand where e-Infrastructure development is paying off and providing gains in scientific discovery. For example, where have investments in e-Infrastructure development occurred that enabled scientists to fashion new problems and puzzles that provided gains in scientific discovery?

The impact of this study will be a contribution to an expansion of the body of knowledge from which stakeholders can draw, as they endeavor to make better-informed decisions about the requirements of e-Infrastructures and their potential for greater innovation and competitiveness than already experienced to date.

### 1.4 Research Questions

**The primary research question** is:

*How is the development of e-Infrastructure impacting scientific discovery?*

This is an exploratory research question that consists of three main components:

1. **The process of e-Infrastructure development:** What e-Infrastructure is, from its origins and concepts to properties, is an exploratory question. We will explain what e-Infrastructure is and its origin within a broader context, and then explain its role in the context of scientific progress over time.
2. **The e-Infrastructure itself:** The development of e-Infrastructure focuses our attention on investments and development involving e-Infrastructure in the context of stimulating scientific discovery.
3. **Its impact on scientific discovery:** Scientific discovery is a result that must occur within some context. The context we will explore is a particular science discipline, because a science discipline consists of knowledge and human resources, and embodies a creative ecosystem in which we can explore interactions between a science discipline and its components, and e-Infrastructure.

The primary research question leads to the following secondary research question.

**The secondary research question is:**

*How are the problems and puzzles of a science discipline shaping the development of e-Infrastructure, and conversely, how is e-Infrastructure changing the problems and puzzles of that science discipline?*

The second research question concerns itself with the interactions between two dynamic environments: a science discipline and e-Infrastructure development.

A science discipline, as previously described, is based on knowledge and human resources, and a community of scientists. Our objective is to observe the effects of e-Infrastructure development on a science discipline, and vice versa, so that we may explain how they potentially mutually shape each other, based on empirical results.

Our inquiry will seek historical information to identify patterns and to piece together how e-Infrastructure development can be fashioned to achieve the most dramatic scientific progress. Answering this question will provide us with concepts and a conceptual framework to investigate the existing relationship between e-

Infrastructure development and a science discipline, and how their interaction can potentially lead to effecting dramatic improvements in scientific progress.

### 1.5 Significance of Increasing Understanding of e-Infrastructure Development and Its Impact on Scientific Progress

e-Infrastructures are an important phenomenon to understand, because e-Infrastructure development potentially could be a pathway for scientific progress and transformative ideas, as well as an investment opportunity for nations seeking to innovate and compete in a global marketplace. Paraphrasing Popper (1959, 1994) and Wagner (2002), it is important to increase our understanding about the effects of e-Infrastructure on the progress of scientific research because the types of progress can result in transformative changes to a nation's economy.

In the U.S., the National Academies' 2007 report, *Rising Above the Gathering Storm*, assessed innovation and competitiveness capabilities along three primary categories: human capital, knowledge capital, and a healthy creative innovation ecosystem.

**Human capital** is a resource that consists of an educated, innovative, motivated workforce (NAP, 2007). In a global economy, an educated workforce must also be globally competent. Globally competent scientists and engineers are those with the ability to frame scientific questions or problems, and to seek solutions with people who have perspectives different than their own (Kirk, 2007). Science disciplines are institutions that offer established ways of developing human capital. **Knowledge capital** is a resource that fuels the growth of business and creates the potential to spawn new industries (NAP, 2007). These industries, in turn, can provide rewarding employment opportunities towards economic development. **An innovation ecosystem** is an interconnected web of "knowledge-creating institutions," conducting "basic research" or "applied research" to create knowledge. *Basic research* is aimed at original investigations for the advancement of scientific knowledge of the subject under study without specific commercial objectives (NSB, 2010). *Applied research* includes original research to increase knowledge, but it is



undertaken with the intent of commercial objectives (NSB, 2010). In an innovation ecosystem, knowledge-creating institutions form a web from interactions among inventors, technologists, entrepreneurs, world-class research universities, highly productive research and development (R&D) centers (both industrially and federally funded), a vibrant venture capital industry, and government funded basic research focused on areas of high potential (PCAST, 2004).

All of these factors — human and knowledge capital, the innovation ecosystem and knowledge-creating institutions — contribute to pushing scientific research forward. Yet there is another important argument why e-Infrastructure is so highly valued and seen as potentially transformative: Its ability to solve complex problems.

Complex problems, such as climate change, are beyond a single discipline's domain of understanding. These problems demand cross-disciplinary knowledge and resources to increase understanding of the phenomenon. Complicating matters, pressures for solutions come from multiple sources, from political to business to social.

Grand challenge problems and puzzles at this scale of complexity can create a demand that attracts investors, scientists and engineers, from both private industry and government. E-Infrastructure development plays a key role in providing an ecosystem of human brainpower, knowledge and technological resources that potentially leads to dramatic improvements in scientific progress.

Conversely, it is important to understand how transformative research — a disruptive style of research aiming to achieve revolutionary discovery — is shaping human, knowledge, and technological resources that collectively form an e-Infrastructure.

At present, not enough is known about how investments in e-Infrastructure development influence transformative research, the engine of revolutionary scientific discovery. Nor is enough known about how the requirements of science disciplines change when problems and puzzles create a demand, influencing

investments designed to both fund and exert pressure on the e-Infrastructure to develop more powerful technologies and instrumentation. For example, the exploding data crisis in science and society is creating a demand for investments in innovative data management solutions (NSF, 2010). These investments could result in the creation of a new e-Infrastructure for science disciplines, such as cloud platforms, as a solution to the data management problem.

## **1.6 Relevance and Potential Contribution**

The contribution of this proposed research is to increase understanding of the effects of ICT investment as stimuli towards e-Infrastructure development and how it potentially impacts scientific discovery. It will also contribute to an understanding of the requirements of a science discipline shaping the development of e-Infrastructure.

Scholarly research on infrastructure development draws on the works of Thomas Parke Hughes' *Networks of Power* (1983), authored about the evolution of electric power as large technological systems (Bijker and Law, 1992; Coutard et al, 2004). The phenomenon of e-Infrastructure development, and in particular its relationship to scientific discovery, is not well understood due to a lack of scholarly research. This void of scholarly research is a new and emerging phenomenon. A qualitative study on this phenomenon has been proposed to explore the interactions between an e-Infrastructure development process and its impact on scientific progress.

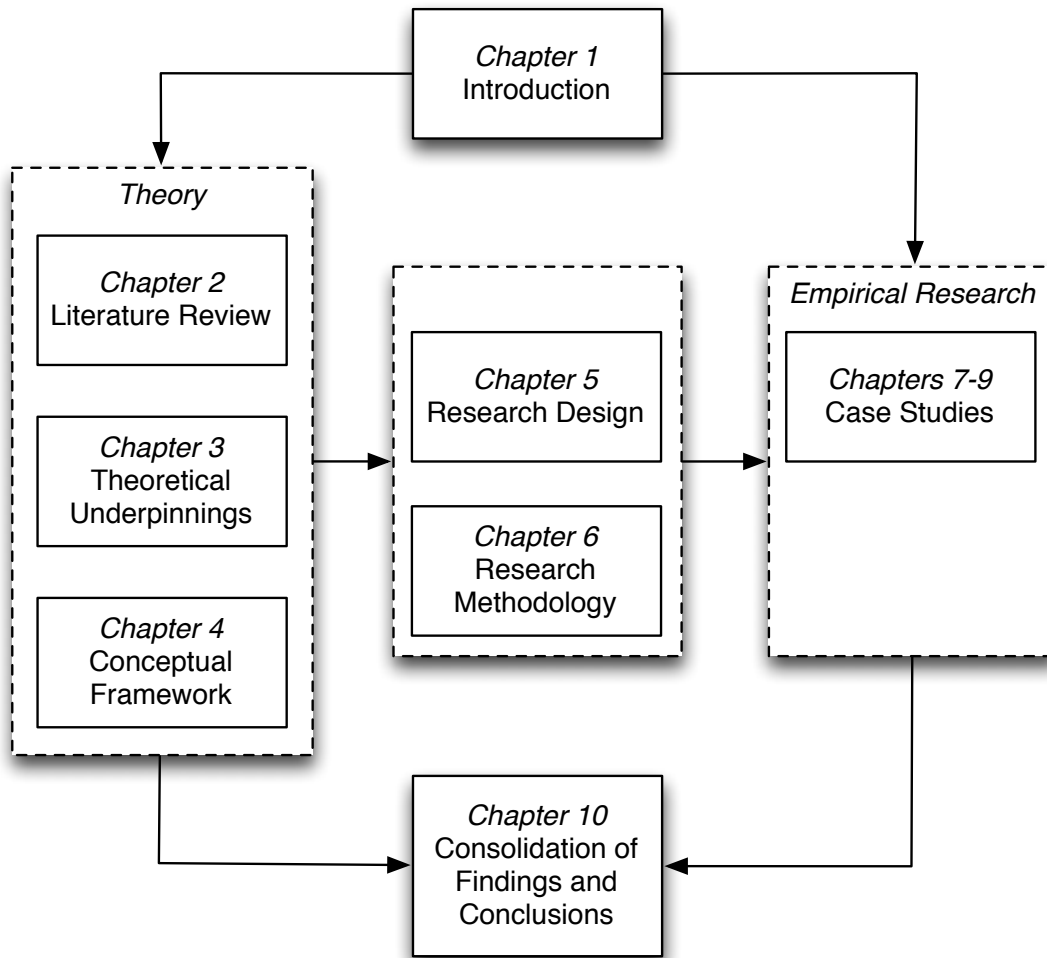
By establishing a reciprocal link between scientific progress and e-Infrastructure development, evidence supporting a powerful set of concepts and tools would be provided to stakeholders, such as government funding agencies as well as prospective investors from private industries, with a potential of increasing scientific progress. Theory will be developed to better explain the impact of e-Infrastructure development programs, such as cyberinfrastructure and e-Science, on scientific progress.

## 1.7 Roadmap and Organization of this Thesis

**Figure 1** below shows the organization of this dissertation.

**Chapter 2** provides a review of the relevant literature that serves as the foundation and scaffolding of our theoretical framework. Included in this literature are the properties of a science discipline and e-Infrastructure. We will also construct an explanation of the mutual shaping that results in scientific progress that is transformative.

Based on the insights derived and the gaps revealed from the literature review, **Chapter 3** explicates the theoretical underpinnings for the study, specifically, the concepts and theories upon which we construct an explanation of the properties of a science discipline, the properties of e-Infrastructure, and the relationship between them.



**Figure 1 Organization of the Thesis**

**Chapter 4** integrates the concepts from Chapters 2 and 3 to construct a conceptual framework for the empirical inquiry. The conceptual framework constructed in Chapter 4 will provide a conceptual lens upon which we can focus on the effects of ICT investment as stimuli on the process of e-Infrastructure development to reveal information that will guide us towards answers to the research questions.

**Chapter 5** elaborates upon the research design, providing a high-level description of the conceptual and empirical components, driven by the research questions.

**Chapter 6** presents the empirical research methodology in detail, explicating the research approach, the multiple case study design, and the methods and procedures

used for data collection and analysis. Issues of research quality and validity are also discussed in this chapter.

**Chapters 7 through 9** contain the case studies that provided the empirical basis for this research. The conceptual framework in Chapter 4 provides the structure and analytical framework for Chapters 7 through 9.

Finally, **Chapter 10** consolidates the findings of the research, provides answers to the research questions, discusses the contributions of the research, and provides directions for future research.

## Chapter II Theory Construction: Literature Review

### 2. Literature Review

Chapter 2 is a review of the literature on concepts — and relations between concepts — that will help us gain understanding about the phenomenon we’re studying. We will review the literature and identify its relevance to our research questions. To set the stage for the research of this study, Chapter 1 defined the research questions. Those questions provided the context for selecting the literature that best answers them.

<b>Research Questions:</b>	
<b>Primary Research Question:</b>	How is the development of e-Infrastructure impacting scientific discovery?
<b>Secondary Research Question:</b>	How are the problems and puzzles of a science discipline shaping the development of e-Infrastructure, and conversely, how is e-Infrastructure changing the problems and puzzles of science discipline?

**Table 1 Research Questions**

Chapter 2 helps us focus in on identifying literature to better understand the phenomenon of e-Infrastructure development, and its perceived impact of stimulating dramatic increases in scientific discovery. The literature examined in Chapter 3 will present theories supporting the answer.

It is important to clarify the difference between the literature presented and examined in Chapter 2 and Chapter 3. Steered by these research questions, the literature examined in Chapter 2 explicates the research problem, the lack of understanding about the phenomenon of e-Infrastructure development and its impact on scientific discovery. On the other hand, the literature examined in Chapter 3 presents theories that will support the proposed solution to answer the

research questions. Chapters 2 and 3 combined provide an interconnection between the research questions and which literature to review to illuminate the research problem, and also concepts from theories that will lead to answers of the research questions (Maxwell, 2005). At the end of Chapters 2 and 3, we will identify the major streams of the literature and summarize each of the topics of the literature review into a single integrated idea.

Our initial step towards explicating the research problem is to make sense of the components of our primary research question: the process of e-Infrastructure development, e-Infrastructure itself, and its impact on scientific discovery. The concept of e-Infrastructure and e-Infrastructure development are nascent, such that scholarly research examining the link between e-Infrastructure development and scientific progress is almost nonexistent. In Chapter 2, we will draw upon literature from the following scholars and researchers to help us illuminate concepts and patterns on e-Infrastructure and a science discipline:

### **Thomas Parks Hughes**

Thomas Parks Hughes, author of *Networks of Power (1983)*, proposed a model explaining the evolution of electric power. Hughes' theory is based on an evolutionary model of large complex technological systems. He characterized large technological systems as constructed in a social context, such that there is interaction with a social system (Hughes, 1987). Hughes' model of infrastructure development has been adapted and extended by historians and sociologists studying the development of infrastructure (Bijker and Law, 1992; Braun and Joerges, 1994; Coutard, 1999; Coutard et al., 2004; La Porte, 1991; Mayntz and Hughes, 1988; Bijker et al., 1987; Kaijser et al., 1995; Summerton, 1994).

### **Star and Ruhleder**

A second stream of foundational literature on infrastructure development draws on the work of Star and Ruhleder (1995). They conceptualized

infrastructure as “something that emerges for people in practice, connected to activities and structures.”

To inform this research and answer the primary research question, we will draw from Star’s and Ruhleder’s concepts of infrastructure as a substrate and as relational, in particular those concepts that transfer to infrastructure development.

The theoretical grounding to guide our answer will be completed in Chapter 3, where Star and Ruhleder’s theory of infrastructure as relational emerges as a result of activities and practices performed by people grounded in Adaptive Structuration Theory (Desanctis and Poole, 1994; Giddens, 1984). We shall borrow concepts of co-evolution theory (Erich and Raven, 1964; Thompson, 1994) and Adaptive Structuration Theory (DeSanctis and Poole, 1994) to guide us in creating a conceptual framework for examining the interactions between a science discipline and the process of e-Infrastructure development.

Concepts and properties of evolutionary and revolutionary scientific progress are present in the sociology and philosophy of science literature (Kuhn, 1996; Popper, 1961 and 1994).

### **Kuhn**

In the second half of Chapter 2, we will discuss Thomas Kuhn who, in his revolutionary work *The Structure of Scientific Revolutions* (1996), characterized the practice of science as undergoing “paradigm shifts,” instead of progressing in a linear accumulation of knowledge. Centered on the role of a paradigm in a scientific community, he introduced terminology for “normal (evolutionary) science,” referring to progress that is more precise or better articulated about an accepted paradigm; and “revolutionary science,” referring to progress that potentially transforms a science and its community with the reception of a new paradigm.



## **Popper**

In Chapter 2, we will review the work of Karl Popper (1961). Scientific progress supported by empirical results was greatly influenced by his work (1961, 1963). Popper revolutionized the classical scientific method based on observation and induction with empirical falsification, also known as “Critical Rationalism,” which advocated framing experiments with a falsifiable hypothesis. Hypothesis testing became a central part of experimental design and the scientific method, also referred to as hypothesis-driven science (O’Malley et al, 2009).

To explicate the concept of scientific discovery, and in particular, in the context of transformative research, we will review reports and contemporary literature on scientific progress emphasizing support for the use of radical and high-risk approaches (NSB; DoD; NAP, 2007). Scholarly literature on scientific practices describes the approaches of inquiry used by scientific communities to conduct empirical research. A review of the literature on scientific practices, and their evolution, has the potential to inform this research on changes in a science discipline. The literature also helps to determine if the progress achieved was evolutionary or revolutionary, and what factors were important in driving that progress.

### **2.1 e-Infrastructure Development**

In the exploration of our primary research question, we will first establish what e-Infrastructure is and its origin in a broader context.

e-Infrastructure is a new concept that has been appearing in reports and literature within the past eight years. e-Infrastructure has been referred to in different contexts as providing a substrate to develop e-Government, e-Business (Yu et al, 2005), and e-Science strategies (Fox, 2003; Hey and Trefethen, 2003). In the context of science, e-Infrastructure has been used as shorthand for “e-Science infrastructure” (Riedel et al, 2009; Voss et al, 2007b).

The term “e-Infrastructure” was adopted by the European Commission for its programs to develop next-generation ICT-based technologies in support of European science focused on innovation and economic development (e-IRG, 2007). Similarly, in the U.S., the National Science Foundation established the term “cyberinfrastructure” (Atkins et al, 2003), and in the U.K., the Office of Science and Technology established the term “e-Science” (Hey and Trefethen, 2002).

The U.S. National Science Foundation characterizes Cyberinfrastructure along the following four categories: (1) High-performance Computing (HPC); (2) Data, Data Analysis and Visualization; (3) Virtual Organizations for Distributed Communities; and (4) Learning and Workforce Development (NSF, 2007).

The National e-Science Centre<sup>2</sup> (NESC) in the U.K. refers to e-Science as “large-scale science that is increasingly carried out through distributed global collaborations.” Similar to the Cyberinfrastructure, e-science requires access to very large data collections, very large scale computing resources, dedicated databases and data storage, high-performance visualization back to the individual user scientists, and easy access to expensive remote facilities. Training outreach and education<sup>3</sup> is also a component of the U.K. e-science initiative.

The e-Infrastructures Roadmap document refers to e-Infrastructure as “the new generation of ICT-based Research Infrastructure in Europe (e-IRG, 2005).” This document describes e-Infrastructure along three main components: (1) Networking Infrastructures that deliver physical connections for the e-Infrastructure; (2) Middleware and Virtual Organizations that connect distributed resources and stimulate new processes altering the way organizations work; and (3) Resources of various types: physical facilities, technologies and instrumentation (e.g. telescopes, satellites, special physics equipment, weather balloons, lasers, spectrometers, and large sensor networks) that is of interest to scientists and how they practice science.

---

<sup>2</sup> <http://www.nesc.ac.uk/nesc/define.html>

<sup>3</sup> <http://www.nesc.ac.uk/training/>

Scholarly literature on these three terms, while limited, characterizes the development of e-Infrastructure as orchestrating a confluence of social and technological activities and artifacts into a “socio-technical arrangement” (Edwards et al, 2007; VOSS, 2009; Bani Mohammed et al, 2010). Some of the elements comprising this “socio-technical arrangement” include services and processes, technological resources, stakeholders of problem domains, organizational practices and social norms, and virtual communities.

Let’s explore each of these elements in more detail:

### **Services and Processes:**

A set of services and processes enables scientists to focus on doing science instead of concerning themselves with the heterogeneity and complexities of technological systems. For many years, scientists exchanged information using email, file transfer tools, and other tools made possible by distributed communication networks. Advancements in distributed high-performance computing, such as grid computing (Foster and Kesselman; Hey et al, 2005), Web services (Atkinson et al) and service-oriented architectures for science (Foster, 2005) have abstracted the heterogeneous characteristics of technological systems into a service layer, enabling researchers to collaborate around data, workflows, and resources across time and space in unprecedented ways (VOSS, 2009).

Foster (2005) classifies the service layer as domain-specific services and domain-independent services. Domain-specific services refer to content, such as data, software and processes specific to the practice of science towards knowledge creation. Domain-independent services refer to functions needed to operate and manage domain-specific services. Resources that are domain-independent can be shared across multiple domains (Foster, 2005). An example to clarify the difference between the two is in the context of data collection involving the use of scientific instruments, such as a telescope, a particle accelerator, a remote-sensing network, etc. The domain-specific service layer enables a scientist to develop a prescription,

in the form of a programming language that is executed as a workflow and produces results requested by that scientist.

### **Technological Resources:**

Technological resources refer to computing and information processing services built upon distributed high performance computation and ICT-based systems, such as the grid and service-oriented architectures. They create a substrate to support domain-specific and domain-independent services (Foster, 2005), operable through a collective of tools, facilities and digital resources. Technological resources include physical resources — networks, storage systems, computers, etc. — needed for remote scientific instrumentation (e.g., telescopes, microscopes or DNA sequencers), federated databases (content), and advanced simulation and visualization environments (computation) (Bani Mohammed et al, 2010, VOSS, 2009).

The concept of “separation of concern” is present in the interaction between Services and Processes, and Technological Resources (Stohr and Zhao, 2001). The presence of a Separation of Concern fosters specialization, and creates the conditions for a demarcation between domain-specific and domain-independent services.

### **Stakeholders of Problem Domains:**

Stakeholders of problem domains refer to the variety of people and/or organizations with specific expertise to solve problems or discontinuities in the evolution of an e-Infrastructure. Problems whose solution is required for the e-Infrastructure to continue evolving are often referred to as “reverse salients.” Reverse salients can be both technological and social. Social reverse salients can range from legal, political, cultural, governance issues, etc. (Edwards et al, 2007). A federal funding agency is a stakeholder by providing funding to conduct R&D to develop a prototype, for example. A financial firm is a stakeholder by providing funding to develop a prototype into a commercial product. Once the e-

Infrastructure stabilizes or evolves to its next stage, stakeholders of problem domains typically move on; therefore, their relationship with the e-Infrastructure development is normally temporary (Hughes, 1987).

### **Organizational Practices and Social Norms:**

Organizational practices and social norms refer to the social and cultural characteristics of a community of scientific practice, which collectively provides a basis for the conduct of scientific work conducted at a distance (Edwards et al, 2007). A scientific community of practice engages an e-Infrastructure through its practices and norms, shaping and growing the e-Infrastructure by drawing in new communities, each with its distinctive norms and practices (Edwards et al, 2007).

### **Virtual Communities:**

Virtual communities refer to a community of scientists whose members and resources may be dispersed geographically, yet are able to function as a coherent and coordinated unit. They are able to conduct advanced scientific collaborations, supported through the use of an e-Infrastructure (VOSS, 2009, Bany Mohammed, 2010).

At this point, we would like to refer to e-Infrastructure as a socio-technological arrangement, consisting of two dimensions: a technological dimension and a socio-organizational dimension.

The technological dimension includes physical resources, such as computers, networks, storage, and measuring instrumentation (e.g., telescopes, microscopes, environmental sensors, etc.). This dimension creates a substrate to support domain-specific and domain-independent services operable through a collective of tools and interfaces, and a set of processes and services that provide scientists the means to conduct their work around data, workflows, and resources across time and space.

The socio-organizational dimension consists of processes, stakeholders of problem domains, practices and norms from a variety of organizations and communities.

Collectively, these socio-organizational and technological dimensions support the use of an e-Infrastructure and the formation of virtual communities conducting advanced scientific collaborations.

### **2.1.1 Infrastructure**

The term “infrastructure” is used in many contexts and is often used as a noun, and preceded with a qualifier (e.g., highway infrastructure, telecommunications infrastructure, information infrastructure, etc.). Infrastructure consists of heterogeneous interacting artifacts (explained later), which must be adapted to work together or interoperate to exchange information (Law, 1987).

While scholarly literature on infrastructure development is well recognized in communities of social sciences, history, information sciences, engineering sciences, among others, the extent of scholarly literature on e-Infrastructure is limited. E-Infrastructure development can be interpreted as an enhancement of the larger phenomenon of infrastructure development (Bani Mohammed et al, 2010).

Our discussion will build upon the foundational concepts of infrastructure developed by Hughes (1983, 1987), Star and Ruhleder (1995) and Edwards et al (2007). To establish an understanding of the origins of the concepts of infrastructure, we shall start our discussion with Thomas Park Hughes and his explanation of infrastructure development in the context of Large Technological Systems (LTS). Hughes’ concepts have been adapted and extended to understand LTS, including telephone, railroads, air traffic control, and other major infrastructure (Bijker et al, 1987; others). We will follow Hughes with Star and Ruhleder (1995) and their theoretical perspective on infrastructure development. Lastly, we include the recent work by Edwards et al (2007) that establishes linkages between infrastructure and cyberinfrastructure.

We shall build upon these scholars' work to construct our conceptual framework for e-Infrastructure development and its relationship to a science discipline.

#### **2.1.1.1**      *Artifacts*

Artifacts are an integral part of infrastructure development. The social constructivism perspective, interpreted by Law (1994), states “artifacts are constructions of individuals or collectives that belong to social groups.” Similarly, Herbert Simon (2001), in his work on the Science of the Artificial, describes artifacts as “human constructions, adaptable to human goals and purposes.”

Artifacts can be fashioned as part of a socio-technical arrangement to develop infrastructure. Our description of artifacts comes from two sources: the theory of social constructivism (Pinch and Bijker, 1994; Law, 1994), and the theory of large technological systems (Hughes, 1994).

Artifacts become assimilated into a network as they interact with other artifacts and factors that shape them (Law, 1994). Adaptation occurs as artifacts get assimilated into a network.

Our use of the term “network” is both technological and social, meaning that a network works as a “constraint system” to stabilize artifacts in place. Through adaptation, artifacts are fashioned into the environment (becoming more fit). Adaptation is a result of the artifact becoming part of or being embedded into an infrastructure.

Hughes classified problem-solving components as either physical or non-physical artifacts, interacting with other artifacts, all of which directly or indirectly contribute to the goals of a Large Technical System (Hughes, 1984). He defined this classification of physical and non-physical artifacts, because his insight allowed him to perceive a continuum of potential artifacts that were socially constructed and could be adapted to function in a dynamic environment.

Law (1994) characterized artifacts as being shaped by social, economic, political, scientific, and other factors, which become interrelated over time through negotiation and controversy.

From this discussion, we adopt the classification of hard and soft artifacts. This classification is consistent with Hughes' classification of physical and non-physical artifacts, as well as the notion of a continuum of artifacts, both technologically and socially constructed. Our use of the terms hard and soft artifacts includes both the denotative (explicit) and connotative (implicit) properties of physical and non-physical artifacts. But it also includes an intuitive property by which to articulate a more precise explanation of the infrastructure development.

The intuitive property we refer to is the tendency of some artifacts to initially be characterized as soft artifacts when they are nascent, ill defined or fuzzy. As these artifacts evolve through improvement, better articulation, and elimination of sub-optimal traits, they harden, becoming more reliable, predictable and established within their environment. An example is the evolution of software running on a general-purpose system (in a test-bed environment) while its source code is unstable and changing frequently, as opposed to that software being casted onto a chip and the code sufficiently hardened where it can execute predictably.

#### **2.1.1.2**      *Hard and Soft Infrastructures*

In the context of Large Technological Systems (Hughes, 1987), **hard infrastructure** refers to technological structures, such as transportation systems, water supply systems, electrical power systems, telecommunications systems, etc., that support the development of a society. This type of infrastructure is also referred to as “hard infrastructure” because the elements of the infrastructure are physical, they consist of technological systems that are interconnected, owned and operated by public and private organizations, forming inter-networks and webs that support loosely-coupled coordination among the elements of the infrastructure (Edwards et al, 2007).



On the other hand, **soft infrastructure** refers to a body of rules or norms, such as the policies, laws, cultural and social standards, etc. of a nation, a community, or an enterprise (Niskanen, 1991). Soft infrastructure can include physical assets that are elements of an institutional infrastructure, such as a financial system, a government or law enforcement system, etc.

For the purpose of this study, the classifications of hard and soft infrastructures provide an array of categories that will help us order different types of infrastructure of science. This notion of hard and soft infrastructures, and their use will be detailed further in Chapter 4 — Conceptual Framework.

### **2.1.1.3**      *Concepts and Patterns of Infrastructure Development*

Having laid some groundwork about artifacts in the previous section, let us now study more closely the concepts and patterns that lead to the development of infrastructure. We will start with Thomas Hughes (1987, 1994), whose classic work identified patterns in the evolution of large technological systems. This will be followed by the work of Star and Ruhleder (1995), both of whom brought concepts of infrastructure as a substrate and possessed relational characteristics. Finally, we discuss the more recent work of Edwards et al (2007), where they build upon Hughes, Star and Ruhleder, and further introduce new concepts towards increasing understanding of Cyberinfrastructure.

#### **2.1.1.3.1 Large Technological Systems:**

Thomas Park Hughes (1994) developed the concept of Large Technological Systems (LTS) to explain the evolution of national electric power grids, the telephone system, and other complex technological structures. Hughes characterized LTS as “being constructed over time and growing organically, containing messy, complex, problem-solving components.” Hughes (1987, 1994) theorized that these large and complex technological systems exhibit patterns as they evolve. LTS can expand and contract as they evolve. Hughes explained that engineers, financial investors, government officials and other participating parties shape the development of LTS

during the invention and development phases by solving critical problems of a social-technical nature (Hughes, 1987).

### **Locally Constructed Technological System of Functioning Artifacts:**

Infrastructure development normally starts as a *locally constructed activity* (Hughes, 1987; Edwards et al, 2007), consisting of both hard and soft artifacts being locally constructed. As a local activity, infrastructure development is normally under the control of a single organization. For example, consider the evolution of grid computing (Foster and Kesselman, 1999). Grids emerged from local projects, typically funded through grants, to create local low-cost clusters of computational nodes that attempt solutions to complex problems (Bohannon, 2005). This class of complex problems would normally require the use of a super computer. However, the high cost of super computer cycles and lack of availability made it impractical.

### **Transfer of Technological System to Another Environment:**

Infrastructure formation becomes active when the artifacts of the local activity (both hard and soft infrastructures) are then *transferred* from one environment to another. This applies not only to technologies in the form of physical objects, but also to transferable organizational components such as parts of a company/institution, or the entire organization (Hughes, 1987). In the new environment, artifacts (hard and soft) experience pressures, placed upon them from technological, social, political, legal, economic, and other variations.

### **Adaptation to Variations in a New Environment:**

Artifacts of the infrastructure adjust to their new conditions and requirements. This is known as *adaptation* (Law, 1994; Edwards et al, 2007). Patterns of adaptation can be observed in our grid computing evolution example (Foster and Kesselman, 2004). Successful adaptation of cluster computing in a location eventually lead to the transfer of these clusters to other locations. These locations were interconnected across wide-area networks, eventually becoming part of the Internet. Resource sharing became one of the early challenges that exerted

pressures on cluster computing communities to join a larger grid computing community (Casanova, 2002).

Edwards et al (2007) extend Hughes' theory of evolution of LTS to explain infrastructure development. They define infrastructure as "networks or webs that enable locally controlled and maintained systems to interoperate more or less seamlessly." Consistent with Hughes (1983, 1987), they recognize that infrastructures can start as systems of hard and soft artifacts, local to a geographic area, centrally controlled, and often with homogeneous characteristics. As they evolve they adapt, becoming more fit, and interoperating with heterogeneous artifacts. They grow from local to regional networks, potentially becoming globally distributed inter-networks, shedding characteristics of centralized control, and replacing them with coordination and decentralized control (Edwards et al, 2007).

#### **Gateways:**

Interconnecting heterogeneous islands of infrastructures can be achieved with *gateways* — tailored artifacts, designed to link heterogeneous systems and infrastructure (Edwards et al, 2007).

Gateways are used to solve the heterogeneity problems with islands of infrastructure that may exist as incompatible clusters of computers or grids. While gateways are often understood as technological (hard artifact) solutions, they more accurately represent social choices (a set of compromises) combined with a technological solution (Edwards et al, 2007). The resulting standard is an artifact that enables the interoperation of two or more incompatible infrastructures.

#### **Reverse Salients:**

As infrastructures are interconnected through networks, parts of the infrastructure can fall out of phase with others, causing problems and anomalies to develop (Hughes, 1987; Edwards et al, 2007). Unresolved anomalies that attract a locus of intense research and engineering efforts from multiple groups of experts are called *reverse salients*.

The concept of reverse salients is defined in military contexts as points where an advancing front is held back (Edwards et al, 2007). Hughes (1987) referred to reverse salients during conditions in LTS where change among components vary at different rates. “Components falling behind or out of phase with others result in a constraint in growth.” For example, in electrical power systems, Hughes (1987) cites that local power systems experienced reverse salients in the form of failures when trying to satisfy growing electricity demands.

Solutions to reverse salients can involve subject matter experts from various disciplines depending on the changes or pressures made to the infrastructure or its components. This involvement can cause the developing infrastructure or component to go out of phase (Hughes, 1987). A **technological solution** would be appropriate to resolve a reverse salient if the characteristics of a component had been changed for the purpose of achieving a result, such as increasing efficiency of a technology resource, as in a cluster of processors. Other technology components may need to have their characteristics adjusted so that they too can operate at an efficient level with the cluster of processors (Hughes, 1987).

A **different organizational form** might, however, be a more appropriate solution for a reverse salient occurring in the development of a discipline or an industry (Hughes, 1987). For example, the U.S. high-energy physics community realized inefficiencies and inequalities in researchers having access to computational resources. As a result, a digital divide phenomenon occurred in the community. A solution to this reverse salient was to create a national organization<sup>4</sup> in order to facilitate access and sharing of resources among physicists and their institutions.

Reverse salients often emerge as a result of change and, therefore, attract problem solvers with specific expertise. Hughes (1987) defined these reverse salients as a set of critical problems. If the reverse salient is classified organizational or financial, a solution could come from stakeholders with specific expertise in financial industries (Hughes, 1987). For example, elite universities in the U.S. had the

---

<sup>4</sup> [International Committee for Future Accelerators \(ICFA\) - Standing Committee on Inter-Regional Connectivity \(SCIC\)](#)

privilege of owning a telescope to attract faculty and graduate students in order to conduct research under the auspices of a world-class astronomy program. Pressures for the science of astronomy to evolve created reverse salients that were both organizational and financial. As a result, the universities came together to create a national organization, with the support of the National Science Foundation, to fund and operate telescopes at a national and international level (Kennedy interview, 2007).

**Reverse Salients and Anomalies Comparison:** Hughes (1987) theorized that in each stage of growth of a LTS, reverse salients emerge and are detected as a set of problems with definable characteristics. These reverse salients would then attract appropriate types of problem solvers. Issues that could not be solved within the context of a socio-technological system, such as a LTS, were classified as anomalies by Hughes (1987) and Constant (1980). Anomalies emerge out of these reverse salients. Unresolved anomalies in reverse salients can lead to a solution that involves the introduction of a new and competing socio-technical system (Hughes, 1987) that replaces the established system. For example, DC-powered lighting systems in the 1880s, developed by Edison, failed to solve a reverse salient, which eventually led inventors and engineers to develop an AC-powered lighting system (Hughes, 1987). By the 1890s, a gateway device was invented to interconnect the heterogeneous DC- and AC-powered lighting systems (Hughes, 1987).

Solutions to reverse salients need not be technological (hard artifacts). Edwards et al (2007) point out that some important reverse salient solutions have been legal, political, social and cultural artifacts constructed as attempted solutions to problems. Edwards et al (2007) describes the following examples as reserve salient solutions relevant to Cyberinfrastructure: “(a) generating metadata; (b) techniques for federating databases held at multiple institutions using different equipment and data formats; (c) domain specific data sharing and publication cultures; etc.”

### **Path Dependence**

Another pattern of infrastructure development we recognize from Hughes (1987) is *path dependence*. Path dependence can be reached when infrastructure

momentum builds past a certain point and achieves sustainability (Hughes, 1987; Edwards et al, 2007). The term “path dependence” is also used to refer to the “lock-in” effects that result from choices between competing technologies (Edwards et al, 2007). For example, the choice of an inferior technology over a superior one — as in the case of VHS over Betamax video — occurred because of widespread adoption of the former.

Technological change builds upon the achievements of path dependent technology (Edwards et al, 2007). New technologies put pressures on existing path dependent technologies to increase widespread adoption. Older technologies are eliminated in favor of new technologies that can replace them and provide solutions or new capabilities. Inevitably, new technologies achieve path dependency when they reach a tipping point of widespread adoption, whereby choosing an alternative technology becomes too costly, not only financially, but also in time, effort to retrain and retool (Edwards et al, 2007).

In summary, Hughes recognized the following patterns in the development of infrastructure: (1) A locally constructed technological system of functioning artifacts; (2) The transfer of one technological system to another environment; (3) The adaptation to variations in a new environment; (4) Gateways; (5) Reverse salients; and (6) Path dependence.

#### **2.1.1.3.2 Substrate and Relational Properties of Infrastructure**

Star and Ruhleder (1995) theorized infrastructure as both substrate and as relational, conceptualizing the notion of layers in the relation between the substrate and the practices and artifacts it supports. Star’s and Ruhleder’s contribution on defining infrastructure was based on a study of a scientific collaboratory — a geographically distributed social-technological system, supporting digital communication and a scientific publishing (Star and Ruhleder, 1995).

Star and Ruhleder (1995) used metaphors to convey their images of infrastructure as a substrate; e.g., “something upon which something else ‘runs’ or ‘operates,’ such as a system of railroad tracks upon which rail cars run.” They interpret the

metaphor of infrastructure as a substrate as “something that is built and maintained, and which then sinks into an invisible background”; in other words, it becomes “transparent.”

Star and Ruhleder (1995) explain infrastructure as relational with the following metaphor: “A cook considers the water system a piece of working infrastructure integral to cooking dinner (organized practice); for the city planner, it becomes a variable in a complex equation.”

We argue that the notion of layers is important towards increasing our understanding of the relationship between a science discipline and e-Infrastructure development for the following reason:

- The presence of a substrate (support layers) is a result of conditions and structures for the creation of artifacts from what was created before.

Artifacts are created by harnessing the effects of phenomena, as well as by adapting extant artifacts to create new artifacts (DeSanctis and Poole, 1994; Simon, 1994; Arthur, 2009). Artifacts used to create other artifacts are called “supporting” or “component” artifacts (Arthur, 2009).

#### **2.1.1.3.3 Domains**

Artifacts that are constructed around a family of phenomena or effects share a common purpose or form subcomponents of useful combinations and tend to form clusters of artifacts (Arthur, 2009). For example, in genomics, methods used for DNA purification, DNA amplification, sequencing, etc. form a cluster of subcomponent artifacts, which can be used as building blocks to create new forms of artifacts (Arthur, 2009). Arthur (2009) refers to these clusters of artifacts as “domains.” A domain is any cluster of component artifacts selected in order to form an instrument or method (Arthur, 2009). In addition to the component artifacts, a domain embodies practices and knowledge. Hughes (1987) made a similar observation involving patents, explaining that patents (component artifacts) tend to

cluster around problem areas. Selected along with its component artifacts, domains and component artifacts can collectively give rise to the formation of a substrate.

Why is domaining important? Domaining, either by design or through an evolutionary process, is important because when it occurs, there is the potential for transformational change. In the case of domaining by design, component artifacts and methods are selected to serve a particular purpose. For example, before the arrival of digital technologies, astronomers observed the sky using a methodology called "classical observing," where they worked with a physical telescope to make observations of the sky. When data mining methods became widely available as a technology, the astronomy discipline created a new domain for detecting celestial objects, based on the mining of historical data.

In scientific practice, domaining occurs when anomalous data is not being resolved by the prevailing paradigm. As previously discussed (Kuhn, 1996; Graham et al, 2002), scientists develop new methods, enhancing their instrumentation or creating new ones, to eventually resolve an anomaly. This process can result in establishing a domain, consisting of new artifacts, principles and methods, all of which collectively solve the anomalous data problem.

The concept of domaining could be important when we are looking for patterns of infrastructure development. Domaining can guide us in recognizing which domains' artifacts were drawn to form an infrastructure.

### **2.1.2 Summary: e-Infrastructure Development**

In summary, we presented concepts by Hughes (1987), Star and Ruhleder (1994), and Edwards et al (2007), in order to lay the groundwork to construct an understanding of concepts and patterns of infrastructure development.

We discussed Hughes' concepts in the context of the construction of large technological systems. Star and Ruhleder (1994) broadened our understanding of infrastructure in two ways: first, infrastructure as a substrate, working as a support structure for something else; and second, infrastructure as relational, with respect to its use or purpose. Edwards et al (2007) broadened Hughes', Star's and



Ruhleder’s ideas in the context of the cyberinfrastructure phenomena by further articulating that infrastructures grow as local and regional networks, and potentially into larger more globally distributed inter-networks, shedding characteristics of centralized control, replaced by coordination and decentralized control.

Table 2 below lists these concepts and patterns, and provides a summarized description of their properties. We shall refer to them as the Concepts and Patterns of Infrastructure Development.

Infrastructure Development Patterns and Concepts	Description
Locally Constructed	Infrastructure development normally starts as a locally constructed activity
Transfer and Growth of Socio-technical System	Successful locally-constructed socio-technical system (infrastructure) is transferred to a new environment
Adaptation	Artifacts of an infrastructure adapt to their new conditions and requirements
Gateways	Tailored artifacts designed to link heterogeneous systems and infrastructure
Reverse Salient	Condition that arises when parts of an infrastructure fall out of phase with other parts or components
Path Dependence	A state of infrastructure development where sustainability and widespread adoption is achieved to an extent where an alternative is too costly
Domaining	A clustering of component artifacts selected to tailor a solution to a problem domain
Substrate	A layer providing support for artifacts or another layer
Relational	Property of infrastructure based upon use or purpose.

**Table 2 Concepts and Patterns of Infrastructure Development**

Coming back to e-Infrastructure as a socio-technical arrangement, our intention at this point is to use the concepts and patterns of infrastructure development (Table

2) along with our conceptual framework (Chapter 4) to identify and describe events, conditions, or changes in categories represented in our conceptual framework that are the result of e-Infrastructure development.

In the next section, we will examine literature on the concepts and properties of a science discipline, so we can then develop a conceptual framework by which to study the interactions between e-Infrastructure development and a science discipline.

## 2.2 Science Discipline: Properties and Concepts

To answer our research questions, we need to understand what e-Infrastructure development is interacting with, before we can understand its affect on scientific progress. To achieve this understanding, our focus will be on a science discipline. We argue that a science discipline consists of knowledge and human resources, and is a knowledge-creating ecosystem. While this might appear to be a plausible argument, we have not yet grounded it in accepted literature.

To explain what constitutes a scientific discipline, we will start with concepts from Kuhn (1996) and Popper (1959, 1994). It is our belief that a science discipline evolves. Both Kuhn (1996) and Popper (1994) believed this as well and explained how science can evolve, using aspects of Darwin's evolutionary theory (1859). Kuhn's (1996) explanation of the evolution of a science discipline as overlapping periods of "normal science," "extraordinary science," "paradigm shift," "revolution," and then back to "normal science," will serve as a guide by which we can understand how e-Infrastructure development is changing as a science discipline. Popper's (1994) falsifiability theory and hypotheses complements Kuhn's theories. Our belief is that they will guide us and provide insights into seeing how e-Infrastructure development could be a factor in changing a science discipline.

As we will explain in more detail later, our approach to [enhancing](#) understanding of how e-Infrastructure development is affecting scientific progress is to construct a theory about the co-evolutionary relationship between a science discipline and e-

Infrastructure. This theory will be grounded on the work of Kuhn, Popper and others in the field of philosophy and sociology of science.

Based on Kuhn (1996), our preliminary definition of a science discipline consists of the following properties: (a) is a particular community of scientists; (b) has problems and puzzles that its members work on; (c) has a methodology; (d) has resources that consists of knowledge and human resources. The sections that follow elaborate on these properties and how they collectively constitute a science discipline.

### 2.2.1 Community of Scientists

Kuhn (1996) refers to a “particular scientific community” that practices research as being defined by its problems and puzzles, which are grounded upon one or more scientific achievements. A community of science builds upon its scientific achievements, which over time accumulate as the scientific community’s body of knowledge accumulates. (Kuhn, 1996). Popper (1959, 1963) further characterized accumulated scientific knowledge as being built upon direct observation, and then synthesized and shaped over time with questions and hypotheses. A scientific community practices its science from an agreed upon set of paradigms (Kuhn, 1996). Briefly, a *paradigm* provides a framework that “defines the legitimate problems and methods of a research field (Kuhn, 1996).” Paradigms serve to unify a community of scientists. Kuhn (1996) explains that during a pre-paradigm period there is much debate<sup>5</sup> that occurs over “legitimate methods, problems, and standards of solution.” Scientific communities debate to build consensus in the community towards adoption of a paradigm (Kuhn, 1996). Paradigms will be discussed more in depth in the next section titled “Methodology.”

A community of scientists is not only defined by the scientists who are its members, but by what they are doing — their practice (Wenger, 2002). Practice in a

---

<sup>5</sup> Even after a paradigm is adopted, debates might subside, but they do not completely stop. Debates intensify among community members when paradigms are under pressure from experimental results not resolving with the community’s hypotheses (Kuhn, 1996).

community of scientists exists because scientists are engaged in actions whose meanings they negotiate with one another (Wenger, 2002). Actions refer to the work, negotiations and debates among community members about problems and puzzles they mutually recognize as important. Wenger (2002) classifies this property of a community practice as the mutual engagement of its participants.

A community of scientists is composed of people who have been educated to work on the community's problems and puzzles. A community aims to attract members through recognition from other communities of its achievements in solving problems and puzzles. Popper (1994) explains that for scientists, the motivation for doing science in a particular community comes from identification of problems, and challenges of puzzles. This motivation stems from their "amazement about something that may be quite ordinary in itself, but becomes a problem or a source of amazement for scientific thinkers." Kuhn (1996) explained that scientists will join another community because the achievements of the other community are recognized as sufficiently unprecedented, and because it offers scientists the opportunity to solve a puzzle that no one has solved before or solved as well.

Similar to a community of practice, a community of scientists is engaged in the generative process of producing its own future (Lave et al, 1991). Kuhn (1996), in his essay on the practice of normal science, explains that a community of science increases its membership when students (or newcomers) elect to study its body of knowledge in preparation for membership.

### **2.2.2 Methodology**

Paradigms provide scientists with a set of basic beliefs and methodology by which to conduct their research. The set of basic beliefs are assumptions (what we think, what we believe, etc. about the world, but cannot prove). Basic beliefs are also called "metaphysical beliefs," because it is not known how to test these beliefs against some external norm (Lincoln and Guba, 1985). For example, what are the origins of mass in the universe? Acceptance of a set of beliefs or metaphysical

principles gives the community an agreed upon methodology to conduct research (Lincoln and Guba, 1985), even if the assumptions are not provable.

Graham et al (2002) explained paradigms as providing a methodology for ecologists, specifically: “Paradigms represent the belief systems that dictate how ecological data are collected and analyzed, and the standards by which they are compared.” Maxwell (2005) refers to Kuhn’s work on paradigms as “a set of very general philosophical assumptions about the nature of the world (ontology) and how we can understand it (epistemology).” Researchers (scientists) of a field or tradition (community) generally accept these assumptions (Maxwell, 2005). Paradigms are discussed in depth later in this section.

Kuhn (1996) described that the methodology includes, first, a criterion for choosing problems that have solutions guided by a paradigm. The result is a set of problems that the community of scientists accepts as scientific and, hence, members are encouraged to further articulate the paradigm from which they were derived (Kuhn, 1996). Also, within the criterion is a methodology by which the problems and puzzles are consistent with the theories and hypotheses previously accepted by the scientific community (Kuhn, 1996). Kuhn explains that under normal science — a period of time when problems, puzzles and solutions are consistent with the established paradigm — solutions are achieved because the scientific community is focused on a clearly defined set of puzzles.

This process of normal science explains evolutionary scientific progress. As defined, evolutionary scientific progress is incremental, building upon the results of previous achievements or further articulating long-standing hypotheses and theories (NSB, 2007).

A second component of the methodology described by Kuhn (1996) is a set of agreed upon rules to determine both ‘what are acceptable solutions’ and ‘how to obtain them.’ What did Kuhn mean by rules to determine acceptable solutions? Kuhn (1962) stated, “Rules are derived from paradigms.” Rules are expressed as

explicit statements of scientific concepts and theories (Kuhn, 1996). Rules were categorized by function — for example, Newton’s laws of motion. These rules were then applied to fashion puzzles out of problems, and to “limit acceptable solutions” (Kuhn, 1996). As an example, a dominant area of research in physics was to understand the forces that acted between bits of matter. Using Newton’s laws, quantity of matter was used as a “fundamental ontological category” to search for solutions to its puzzles.

The rules about how to obtain acceptable solutions to a particular puzzle were influenced by the endorsement of particular instrumentation. The selection of an instrument for an experiment was influenced by the commitments made by that community to preferred types of instrumentation and the norms about the legitimacy of certain instrumentation for experiments (Kuhn, 1996). For example, rules required that selected instrumentation produce results in the same units of measure as the ones that were entered into a theoretical model (Kuhn, 1996). Kuhn (1996) explained that a scientist, as a member of a scientific community, has commitments to “understand the world and to extend the precision and scope with which it has been ordered.”

#### **2.2.2.1**      *Phenomena and Methodology*

In this section, we will further describe the concept of methodology in the context of scientists searching for phenomena. The activity of searching for new phenomena is directly coupled with methodology. In this section, we will explain the interaction between methodology and discovery of phenomena. Based upon this explanation, we will scaffold the methodology of a science discipline with e-Infrastructure development, and their connection to increasing scientific discovery

In general, scientists search for phenomena to uncover their effects, which are observable and can be measured (Arthur, 2009). Using a phenomenon, scientists can measure an effect that varies alongside the thing they are trying to measure. A phenomenon of nature, which exists in the absence of human influence, such as gravitation, generates effects that can be measured (NRC, 2001; Arthur, 2009). The

behavior of objects in nature produces certain effects, which in turn can be exploited by scientists to achieve a purpose. For example, the steady oscillating frequency of quartz crystals or pendulums is an effect that constitutes a phenomenon in nature. Because scientists develop concepts to make use of phenomena for a purpose, they used the effect of the oscillating crystals to conceptualize time keeping. This leads to the invention of the clock (Arthur, 2009).

Phenomena also occur from the interactions of natural and human systems. For example, ecologists are able to sense biological, physical and chemical phenomena below, on and above the Earth's surface using methods based upon advances in computational capabilities and microbiology. For example, data from wireless sensor networks, deployed to measure environmental variables, are enabling scientists to sense new phenomena at different temporal and spatial scales (Porter et al, 2005; Carey et al, 2012; Hanson, 2008).

Another example, but this time in the environmental sciences discipline, concerns hydrologic forecasting. Hydrologic forecasting is about the challenges of making predictions such as “changes in freshwater resources and the environment caused by floods, droughts, sedimentation, and contamination ... (NRC, 2001).” The NRC (2001) report recognized enhancements in the technological capabilities of remote sensors that are driving a revolution in hydrologic science. The enhanced capabilities of these remote sensing technologies made it possible for scientists to measure hydrologic phenomena never before seen (NRC, 2001).

Discovery of phenomena can lead to revolutionary scientific progress (NSB, 2007). In general, scientists seek phenomena to uncover their effects. Once detected, scientists work intensively to harness those effects towards the aim of discovery. They then build upon the effects harnessed previously to construct new instruments and methods in order to increase understanding of the new effects, and ultimately, harness them to further advance the discipline's knowledge. Faraday, for example, uncovered the effects of electromagnetic induction using instruments, technologies

and understandings from effects<sup>6</sup> harnessed earlier into existing methods and understandings (Arthur, 2009; Kuhn, 1996).

Effects uncovered earlier enable scientists to create methods, allowing them to see deeper into a phenomenon, which could result in increased scientific understanding that could lead to uncovering later effects (Arthur, 2009). Methods employing instrumentation and technologies result from harnessing the effects of phenomena (Arthur, 2009).

*Is e-Infrastructure development, then, giving rise to the discovery of new phenomena?*

Arthur (2009) explains that the effects of electrical phenomena were uncovered between 1750 and 1875. “Capturing and harnessing these effects for use resulted in numerous methods and technologies: the electric battery, capacitors and inductors, transformers, telegraphy, the electric generator and motor, the telephone, wireless telegraphy, the cathode ray tube, the vacuum tube (Arthur, 2009).” The uptake and use of these technologies enhance scientists’ capabilities and enable them to take action, such as the development of new instruments with greater precision to see deeper into phenomena. DeSanctis and Poole (1994) developed Adaptive Structuration Theory (AST) to explain this “technology-action” relationship as “continually intertwined.” They further explain, “There is a recursive relationship between technology and action, each iteratively shaping each other (DeSanctis and Poole, 1994).”

In summary, scientists harness the effects from phenomena by developing new methods or improving known ones. These methods embody new instrumentation

---

<sup>6</sup> Faraday uncovered electromagnetic induction using knowledge and instruments that had been developed from effects harnessed earlier (Arthur, 2009). First, the battery is a technology that used electrochemical effects harnessed earlier. Second, the principle that a coil when wound around a magnetic material, such as iron, produces a stronger magnetic field (effect) had been a recent discovery (Arthur, 2009). There were other methods and technologies he used in his experiment to discover electromagnetic induction.



developed from technologies based on the effects of discovered phenomena. The community of scientists is shaped from the use (action) of this technology-driven instrumentation. The process begins again when scientists harness the effects of another phenomenon.

To answer our research questions, we will examine how methodologies have changed, supported by e-Infrastructure developments, and how these changes in methodologies have given rise to discovery of new phenomena.

#### **2.2.2.2**      *Paradigms*

We are now ready to give a thorough treatment to the concept of a paradigm. An explanation of paradigms is important in answering the primary research question (*How is the development of e-Infrastructure impacting scientific discovery?*), since paradigms are inextricably linked to the concepts of evolutionary and revolutionary science.

Thomas S. Kuhn (1996) introduced the concept of a paradigm in his controversial book *The Structure of Scientific Revolutions* to provide an explanation of how science evolves from cycles of articulation of problems and puzzles, followed by attempted solutions (Weinberg, 2001). Guba and Lincoln (1989) refer to paradigms as “basic belief systems.” Paradigms steer scientists of a discipline to adopt a consensus view, “to agree on what phenomena are relevant and what constitutes an explanation of these phenomena, about what problems are worth solving and what constitutes a solution of a problem (Weinberg, 2001).” Once a consensus is achieved, paradigms guide scientists to establish a commensurable system by which to conduct empirical work (Mahajan, 1992).

Paradigms guide scientists in the design and construction of measuring instrumentation for an experiment to solve a puzzle. Through paradigms, scientists develop instrumentation for empirical research with greater confidence. This confidence is reinforced by the established consensus in their scientific community. A paradigm, then, can result from a theory that appears to be better than its

competitors. Theories compete — and sometimes fail — to ascend as paradigms; Franklin’s fluid theory of electricity, for example, (Kuhn, 1996) failed to ascend to a paradigm.

### **Paradigms and Progress:**

While controversial, Kuhn’s theory of paradigms is considered useful in understanding progress and the evolution of a science discipline. For example, Graham et al (2002) use Kuhn’s idea of paradigms for studying the evolution of ecology. Graham et al (2002) explain that Kuhn (1970) selected the term “paradigm” to depict how “humans acquire knowledge, which inevitably leads to a suite of methodological, philosophical, and even social constructs that guide scientists and their investigations.” In their study of present-day ecologists, Graham et al (2002) found that “paradigms represent the belief systems that dictate how ecological data are collected and analyzed, and the standards by which data are compared.” They explain that constructs derived from theories, such as “theories of island biogeography, continental drift, or the biological species concept,” are used to construct a framework that contains a broad and coherent set of rules, standards and hypotheses that help guide future research. Kuhn (1996) described repeated use of such frameworks as “further articulation and specification under new or more stringent conditions.” Progress of this type is “evolutionary” because the prevailing paradigm is further articulated.

For the science discipline of ecology, Graham et al (2002) view paradigms as possessing a temporal property, meaning that they represent the current state of scientific knowledge held by the members of a science discipline. They also view paradigms as possessing a relational property. For example, for one ecologist, a paradigm could represent the single model that characterizes a specialized domain of study, whereas to another ecologist, it is a set of methods that defines the way in which an evolving research domain collects and analyzes its data.

Paradigms can gain and lose popularity within a discipline (such as ecology), normally contingent upon how more or less successful they are than their

competitors in solving problems, particularly those recognized as acute (Kuhn, 1996). Graham et al (2002) explained that a method ecologists use to measure their progress in understanding ecology is to compare how well all of the paradigms of the discipline collectively represent the “status quo”; meaning, how well these paradigms represent the current state of scientific understanding.

We will refer to paradigms in our exploration of how e-Infrastructure development has affected them. Changes to paradigms could provide insights on the types of scientific progress being made.

### **Paradigms and Practice:**

Paradigms provide patterns that can guide scientists in their investigations to create knowledge. Weinberg (2001), although critical of Kuhn’s essays in *Structure of Scientific Revolutions*, recognized Kuhn for explaining progress in science as being similar to the biological evolution described by Darwin (1859). Weinberg’s critique of Kuhn also underscores the consistency of both Kuhn’s and Popper’s explanations for the evolution of scientific theories. “For Kuhn, the natural selection of scientific theories is driven by problem solving. When during a period of normal science it turns out that some problems cannot be solved by using existing theories, then new ideas proliferate, and the ideas that survive are those that do best at solving these problems (Weinberg, 2001).” Similarly, Popper (1994), with his theory of falsifiability, explained that theories that are more fit survive the pressures brought upon them.

Kuhn’s and Popper’s theories reveal patterns in scientific practice, which Kuhn described as a cycle of different modes of scientific practice: “normal science,” “crisis,” “extraordinary science,” “paradigm shift,” “revolution,” and then back to “normal science.” These patterns of change are described below. Kuhn, Popper and Weinberg each explained that the cycle starts with a problem or problem situation. As per Popper (1994), “We are always learning a whole host of things through

falsification. Above all, we gain a new and more sharply focused problem; and a new problem is the real starting point for a new development in science.”

### 2.2.3 Problems and Puzzles

In this section, we explore problems and puzzles of a science discipline so that we can continue to build our conceptual framework. We shall borrow concepts and patterns of scientific practice developed by Kuhn and others in the philosophy and sociology of science to guide us in establishing relationships and answers to our research questions.

Within the framework of an adopted paradigm, scientists define problems and puzzles that are recognized by their community. Kuhn (1996) categorizes puzzles as a special type of problem, used to further articulate the paradigm from which they were derived. He further explains that the types of puzzles useful to scientists are ones for which a solution is expected to already exist. Although a particular outcome is anticipated, how scientists go about constructing an experiment, with instrumentation that increases accuracy and scope, is considered a contribution. Scientists are recognized for the ingenuity and skill they employ in making this contribution (Kuhn, 1996).

What are problems and puzzles of a scientific community? Popper (1994) in his essay on “The Logic and Evolution of Scientific Theory” explains that each new development in science has as its starting point: a problem or problem situation. Problems appear as inconsistencies or anomalies in relation to a discipline’s accumulated knowledge. Anomalies are detected when data are inconsistent with the generality of accepted theories of a discipline’s body of knowledge (Graham et al, 2002).

Problems and puzzles are designed to support the paradigms of a discipline. A paradigm provides a methodology to guide scientists to choose problems. “One of the things a scientific community acquires with a paradigm is a criterion for choosing problems that, while the paradigm is taken for granted, can be assumed to have solutions (Kuhn, 1996).” Problems within the scope of a paradigm are

normally the only ones recognized by its scientific community as legitimate scientific problems (Kuhn, 1996). Kuhn further explains that problems outside the scope of the paradigm are normally classified as “metaphysical,” or “as the concern of another discipline,” or as too problematic to be worth the effort (Kuhn, 1996). Legitimate scientific problems are reducible to the puzzle form and accepted by a scientific community, because they can be stated in terms of the conceptual and instrumental tools that the paradigm supplies (Kuhn, 1996).

Kuhn (1996) defined four modes of scientific practice to explain the ways in which scientists work on problems and puzzles defined by their paradigms. Kuhn's modes of scientific practice are:

1. Normal (evolutionary) Science
2. Extraordinary Science
3. Paradigm Shift
4. Scientific Revolution

The design of problems and puzzles to achieve revolutionary progress is not explicitly explained by Kuhn and other authors of the philosophy and sociology of science. Literature explains that revolutionary progress occurs from the effort of resolving an anomalous condition, and not from the point of view of design of problems and puzzles to achieve revolutionary progress.

Kuhn's four modes of scientific practice are discussed in the following sections. Where he discusses revolutionary changes to paradigms, we compare his explanations with contemporary literature.

An important part of our effort to answer the research questions is to inquire about changes to a science discipline and the conditions that alter the design of problems and puzzles in order to achieve revolutionary progress.

### 2.2.3.1 *Normal (Evolutionary) Science*

Previously, we referred to evolutionary science and normal science in the same way. Kuhn (1996) uses the term normal science. Contemporary literature (NSB, 2007) refers to this mode of scientific practice as evolutionary science.

Kuhn (1996) described “normal science” as periods of time when scientists work comfortably using constructs, theories and hypotheses within the framework of their paradigms. During periods of normal science, scientists work on problems and puzzles that strive to bring theory and experimental results into closer agreement (Arneson, 2006; Loehle, 1987; Mentis, 1998). They use theories and constructs of the paradigm to work on recognized problems and puzzles to provide confirmation or falsification of the paradigms (Arneson, 2006).

Further articulation of a paradigm’s theories and hypotheses is a significant part of normal science. “Normal science research is directed to the articulation of those phenomena and theories that the paradigm already supplies (Kuhn, 1996).” Kuhn described this activity as the “clean up” or “mop up” work of a paradigm. Kuhn (1996) characterized mop up work as part of bringing a normal science research problem to a conclusion. It provides an incentive to attract scientists to become members of a discipline because it furthers their careers while contributing to the knowledge of a scientific community. Although the solution of a normal research problem is anticipated, the contribution is derived from achieving the anticipated in a new way.

Kuhn’s concept of normal science is consistent with Popper’s three-stage model. Using Darwin’s theory of evolution as an underpinning, Popper (1994) formulated a three-stage model to explain how scientific theories evolve. The three stages of Popper’s model are: (1) the Problem; (2) the Attempted Solutions; and (3) the Elimination.

The first stage — the Problem — arises when a disturbance causes a change in the environment to occur. Popper argued that the starting point of the process of normal science is “a problem” or “a problem situation.” The second stage – the

Attempted Solution — attempts to solve the problem. The third stage — the Elimination — eliminates non-solutions. Popper explains the use of his three-stage model as a closed-loop system between stages 2 and 3. If an attempted solution fails, it is eliminated, and then another solution is formulated. This loop continues until a solution is arrived at.

During normal science, a science discipline will identify problems (or problem situations). Attempted solutions will consist of theories (Popper, 1994). Popper explains that theories are often wrong because they are conceived through trial and error. Theories that are non-solutions are subjected to elimination. Normal science continues as long as a theory, within the context of the current paradigm, is found to be a solution. If no such solution is found, a period of crisis commences, leading into extraordinary science.

### **Discovery and Normal Science:**

It's important to note that Kuhn (1996) makes a distinction between a “solution” to a normal research problem, and a “discovery.” Kuhn (1996) explains that a solution in a normal research problem is considered significant to a community of scientists, because it “adds to the scope and precision with which the paradigm can be applied.” A solution does not result in a change to the paradigm (Kuhn, 1996).

Discovery, on the other hand, “commences with the awareness of anomaly (Kuhn, 1996).” *Awareness of an anomaly is the event that triggers a potential for progress that is revolutionary.*

Kuhn (1996) refers to a “new scientific fact” as the object of a discovery. Under normal science, these scientific facts are anticipated and, therefore, are not new. A new scientific fact is the resolution of an anomaly that results from a change to the paradigm being used (Kuhn, 1996). This anomaly is the gateway to extraordinary science.

### 2.2.3.2 *Extraordinary Science*

An indicator that a paradigm is under pressure occurs when anomalies start to appear in the conduct of normal science. “When an anomaly comes to seem more than just another puzzle of normal science, the transition to crisis and to extraordinary science has begun (Arneson, 2006).” What is an anomaly in this context? An anomaly is something observed or perceived as an irregularity or abnormality within the theories or hypotheses of a paradigm, or normal practice. Scientists use technology artifacts as a means by which to detect anomalies. Anomalies can be detected in data that are inconsistent with the generality of accepted paradigms (Graham et al, 2002). Anomalies emerge when both the measuring apparatus (hard artifact) and the knowledge of the science community have sufficiently developed to make an anomaly recognizable as an inconsistency that is reproducible. Higher-precision measuring apparatus for example has provided scientists with enhanced capabilities to see deeper into phenomena (Robertson, 2003).

What happens during periods of extraordinary science? During these phases, anomalies are recognized as more than another problem or puzzle of normal science (Arneson, 2006). When anomalies are not being resolved, a discipline enters a period of crisis. Kuhn (1996) reasoned that “anomaly leads to crisis, crisis leads to extraordinary research, and then extraordinary research leads to revolution.”

An example of a period of extraordinary science occurred in the late 19<sup>th</sup> century in the discipline of physics using Newton’s paradigm for the study of motion and gravitation. The crisis occurred when Newton’s theories for motion could not resolve with experimental data for the motion of light (Weinberg, 2001; Kuhn, 1996). “This problem was solved through a paradigm shift, a revolutionary revision in the understanding of space and time carried out by Einstein in the decade between 1905 and 1915, and going far beyond the crisis that had inspired it (Weinberg, 2001).”



During crisis, more attention is devoted to the anomalous problem by more of the field's most eminent scientists (Arneson, 2006). It is during these periods of extraordinary science, when current theories and paradigms are put under stress, that achievements are extraordinary, accumulation of knowledge is greater, and the discipline evolves (Kuhn, 1996, Graham et al, 2002). Resolution of the anomaly can evolve into opportunities to further articulate and refine the paradigm in order to better explain observed patterns (Graham et al, 2002).

### **Paradigms Under Pressure:**

Pressures exerted on paradigms increase the potential of revolutionary progress. Popper (1994) explains theories under pressure in Stage 2 of his model of "attempted solutions." If attempted solutions fail, theories of paradigms are subjected to the elimination process (Stage 3 of Popper's model). Popper's attempted solution stage is similar to Kuhn's explanation of the pressures exerted on paradigms in periods of crisis and extraordinary science. If successful in resolving anomalous data, such modifications make paradigms more fit to continue evolving while under pressure in their environment (Popper, 1994). A reason they become more fit is because learning occurs as an outcome of a solution. Popper (1994), in the context of evolutionary theory, defined learning to mean "that unsuccessful or discarded solutions drop more and more to the level of passing references, so that eventually the successful attempt at a solution appears to be almost the only one left."

On the other hand, if resolution is not achieved because data continues to be anomalous, then it can spur the beginning of extraordinary research and revolution (Graham et al, 2002). Popper (1994) explained this extraordinary research effort as the "conscious application of the critical method" that is part of the elimination process (Stage 3) within his model of the evolution of scientific theory. In the context of the scientific method, Stage 3 of Popper's model postulated a conscientious critical method towards the elimination of non-solutions. Popper (1994) argued that it is during this period of extraordinary science that the critical method would involve the development, uptake and use of technological artifacts

for more precise and further articulated attempted solutions, theories, or hypotheses, so that they could become “objects of conscious critical investigation.” Robertson (2003) refers to these objects as “conceptual or technological innovations (artifacts),” which scientists develop to enable them to “see” more about a phenomenon.

### **Hardening Paradigms:**

Under a period of extraordinary science, scientists attempt to harden their discipline’s paradigms by refining theories and hypotheses to be more objective, less ambiguous, and operational about the sorts of phenomena to which they apply (Weinberg, 2001). If successful, this process results in making the paradigms more fit to survive new pressures. In his critical essay on Kuhn’s theory of scientific revolutions, Weinberg (2001) argues that there are “hard” and “soft” parts to scientific theories, where “hard” means “durable,” “mature,” and less “ad hoc.” As theories become more fit and mature, the hardened theories and hypotheses of a paradigm represent permanent and durable accomplishments (Weinberg, 2001). Using a metaphor involving T-shirts with Maxwell’s equations of electromagnetism on the front, Weinberg (2001) states, “If you have bought one of those T-shirts, you may have to worry about it going out of style, but not about it becoming false.”

Theories that are not durable, Weinberg (2001) classifies as “soft.” The soft parts of scientific theories do change; they are ad hoc, and more ambiguous (Weinberg, 2001). Examples of soft theory are Maxwell’s theory of an ether to explain electromagnetism or Newton’s theories of particles and forces not being sufficient to explain Nature. Weinberg (2001) relates soft parts of scientific theories to the commitment of the scientific community to harden them through consensus. “It is only when scientists share a consensus that they can focus on the experiments and the calculations that can tell them whether their theories are right or wrong, and, if wrong, how they can show the way to a new consensus (Weinberg, 2001).” Weinberg (2001) referred to the standard model of elementary particle physics, which is considered very successful in accounting for measuring the properties of

known particles, as an example of a field theory with soft parts due to physicists not firmly committed to the theory.

### **2.2.3.3**      *Paradigm Shift*

“Paradigm shift” is the next mode of scientific practice defined by Kuhn (1996). A paradigm shift is a transition from a paradigm in crisis to a new one from which a new tradition of "normal science" can emerge (Arneson, 2006). A paradigm shift period commences when a paradigm remains in crisis from unresolved anomalies. Paradigm shifts can lead to a reconstruction of the field from new fundamentals, changing some of the field's most elementary theoretical generalizations, as well as many of its paradigm methods and applications (Arneson, 2006). An example of a paradigm shift is the replacement of Newtonian mechanics with Einstein's theories of mechanics within special relativity (Robertson, 2003).

In adopting a paradigm, scientists of a discipline acquire theories, methods, and standards that have become path dependent over time as they have become more fit from surviving against other pre-paradigm theories. As a result, change to a paradigm impacts the accepted problems and puzzles of a discipline, and their solutions. “When paradigms change, there are usually significant shifts in the criteria determining the legitimacy both of problems and of proposed solutions (Kuhn, 1996).”

Scientists will exhaust all possible solutions using the rules of normal science, applied in the area of difficulty with the hopes that they can be made to work (Arneson, 2006; Popper, 1994). Arneson (2006) explains that it is during this time that scientists seek ways of magnifying the breakdown and seeing deeper into the area of difficulty. Robertson (2003) explains the act of being able to see deeper as an ability that is created through a technological or conceptual invention. The ability to see deeper enables scientists to detect anomalies with greater clarity and precision. Also, the ability to see more connects scientists with the ability to do more (Robertson, 2003).

In our exploration of a science discipline, our objective will be to uncover facts about contemporary paradigm-shift events and what role e-Infrastructure played in influencing discovery.

Paradigm shifts are not scientific revolutions (Kuhn, 1996). Paradigm shifts can help improve paradigms and increase their path dependencies when their theories and hypotheses become more fit (Popper, 1994). Likewise, as theories and hypotheses become more fit, so do paradigms become more fit (Popper, 1994). As a result, the discipline matures and becomes more path-dependent on its paradigm (Kuhn, 1996).

It is during these time periods of transition from a paradigm shift that a science discipline is (a) receptive to the uptake and use of hard and soft artifacts, and (b) active in the creation of hard and soft artifacts to arrive at solutions to reestablish normal science. In astronomy, important paradigm shifts resulted from the invention of the telescope (Robertson, 2003). For example, it was the telescope that allowed Galileo to see deeper into space and discover the moons of Jupiter and observe the rotation of the Sun. From his observations, he then empirically validated the theory of the Sun-centered Copernican model, which eventually led to paradigm shifts (Robertson, 2003).

#### **2.2.3.4**      *Scientific Revolution*

Kuhn (1962) used the term “revolution” as the next phase of an expanding paradigm shift. It depicts a growing sense that an existing paradigm is no longer useful in explaining a phenomenon (Arneson, 2006). Kuhn (1996) argued that a revolution results when an “existing paradigm has ceased to function adequately in the exploration of an aspect of nature to which the paradigm itself had previously led the way.” Graham et al (2002), in their application of Kuhn’s theories to ecology, characterized the period of revolution as “accepted paradigms suffering from intense pressures and tensions.” This occurs because they can no longer resolve experimental results, “forcing scientists to shed the constraints of the paradigms in search of new understanding.” The breakdown in the paradigm, resulting in

revolution, could impact a narrow subdivision of the scientific community (Kuhn, 1996).

What do scientists experience during a time of paradigm revolution? They experience new theories emerging to resolve anomalies that refuse assimilation into existing paradigms (Kuhn, 1996). They experience the redefinition of their problems and puzzles with the reception of a new paradigm (Kuhn, 1996). Known problems may be perceived to be less significant. Other problems, perhaps not perceptible or insignificant with the old paradigm, could, with a new paradigm, become the standard bearer for significant scientific achievement (Kuhn, 1996).

The discipline of chemistry experienced a redefinition of its science when its paradigm, based on the forces of mutual affinity and led by French chemists, was eventually replaced by Dalton's chemical atomic theory (Dalton et al, 1893; Nash, 1956; Kuhn, 1996). John Dalton, an English meteorologist investigating physical problems of the absorption of gases by water and of water by the atmosphere, revolutionized the practice of chemistry with his paradigm (Kuhn, 1996). Dalton's paradigm also changed the way chemists' recorded scientific data (Kuhn, 1996).

Kuhn (1996) also describes the realization of working with a new paradigm as being able to see familiar objects, but from a new and broader perspective, and being able to see these new objects that were not visible within the previous paradigm. He called it a "transformation of vision." "During revolutions scientists see new and different things when looking with familiar instruments in places they have looked before (Kuhn, 1996)." Kuhn (1996) provides examples from the history of electricity, the history of chemistry, and the history of astronomy that describe the effects of a new paradigm, transforming scientists' vision, and enabling them to see different things when looking at familiar objects.

When scientific revolution leads to the next normal science tradition, Kuhn (1996) tells us it can be both incompatible and incommensurable with the old paradigm. The new paradigm provides the means for new theories and new hypotheses, a new methodology, new problems and puzzles, and even a new way of representing data.

### **Summary:**

We have discussed problems and puzzles of a science discipline. A science discipline defines its problems and puzzles within the framework of an adopted paradigm. Modes of scientific practice — developed by Kuhn (1996) and Popper (1994) that explain how scientists work on problems and puzzles within the framework of a paradigm — were presented.

In our exploration to answer our research questions — i.e., *How is the development of e-Infrastructure impacting scientific discovery? How are the problems and puzzles of a science discipline shaping the development of e-Infrastructure, and conversely, how is e-Infrastructure changing the problems and puzzles of that science discipline?* — we will use these modes of scientific practice as a guide for recognizing events and patterns as we examine the interactions of e-Infrastructure development with a science discipline.

#### **2.2.4 Resources of a Science Discipline**

Up to this point in our review of the literature, we have conceptualized a science discipline to consist of (1) a community of scientists, (2) a set of problems and puzzles defined by its paradigm, and (3) a set of methodologies by which to solve problems and puzzles. We also know from our review of the literature that science evolves by constructing new solutions to problems and puzzles using previous scientific achievements, embodied in its methods and technologies (Arthur, 2009; Kuhn, 1970; Robertson, 2003).

It would be reasonable to say that elements 1, 2 and 3 of a science discipline listed above are both human and knowledge resources of a science discipline. A community of scientists is a resource of information and experience (knowledge and human) that contributes to the evolution of a discipline. Problems and puzzles are knowledge to a science discipline, because they create inconsistencies and unresolved tensions within a discipline's accumulated knowledge. Successes and failures to problems and puzzles provide a knowledge resource that builds community through sustainability and prosperity. Methodologies can serve both as

knowledge and technology resources. As a knowledge resource, methodologies provide a set of rules (derived from paradigms) that scientists of a discipline agree to follow (Kuhn, 1962). Rules can be grouped by function, as in Newton's Laws of Motion (Kuhn, 1962). Arthur (2009) describes them as groupings of technologies that embody the rules of a paradigm, because they can be grouped based on the phenomena they harness, such as chemical, electrical and optical phenomena. In this fashion, they also serve as technology resources to create the instrumentation that is necessary to see deeper in order to observe and understand the next phenomenon.

In the next two sections, we will examine the concepts of "knowledge" and "human resources" based upon contemporary literature. This literature is focused on increasing revolutionary scientific practice to secure the prosperity of society within an information and knowledge-driven economy (NSB, 2010, S&T Indicators; RAGS, 2010). Collectively, we refer to these objects as the "resources of a science discipline." Finally, once the grounding of these concepts is established, we will summarize our understanding towards the development of our conceptual framework.

#### 2.2.4.1 *Human Resources*

Human resources, along with knowledge resources, is one of three categories defined by the National Academies of Science as an essential ingredient for achieving scientific progress that is revolutionary (NAP, 2010). Science disciplines must have a supply of skilled and creative individuals who come with new ideas and unconventional approaches to stimulate revolutionary progress. Likewise, the demand for scientists and engineers should be driven by the goals for the prosperity and well being of society.

"Human resources of a science discipline" is a population of people who are practicing basic and applied research to achieve progress that is either evolutionary or revolutionary. In the U.S., this population includes the Science and Engineering (S&E) workforce, and the population of graduate students and post docs conducting

research that potentially contributes to the knowledge of a science discipline. Our use of the term “workforce” is consistent with the National Science Foundation and the National Science Board (NAP, 2010), where it refers to “employed scientists.” In contrast, the term “labor force” includes the scientist population that is employed and the population that is unemployed (NAP, 2010). We do not include “the labor force” in our definition of human resources of a science discipline, because we consider the unemployed population to not be involved with research practices that will lead to evolutionary or revolutionary progress.

How are science disciplines developing their human resources? In the U.S., development of a science workforce has been characterized as being based on three activities: (1) increasing the number of graduates with science degrees, (2) increasing immigration of foreign scientists, and (3) delaying retirements of senior scientists because of the relative youth and inexperience of the science workforce compared to the total U.S. workforce (NSB, 2010).

In summary, we have characterized human resources of a science discipline as a population of people, consisting of a global S&E workforce, and a population of graduate students and post-docs, who are practicing basic and applied research to achieve progress that is either evolutionary or revolutionary.

#### **2.2.4.2**      *Knowledge Resources*

Knowledge resources refer to the methodology (practices), problems and puzzles, and technology domains scientists use to create knowledge. Methodology and problems and puzzles were discussed previously based on Kuhn (1996) and Popper (1994).

A technology domain refers to a family of components, and to the rules for combining those components in order to form technological devices (Arthur, 2009). A technology domain forms a cluster of technological practices; whereas, a technology is a device (artifact) that does a specific job (Arthur, 2009). For example, the radar technology domain consists of a set of rules and practices for combining individual radar technologies (artifacts) to create a radar device or system.



Referring back to Kuhn (1996), during periods of extraordinary science and paradigm shift, the knowledge resources of a science discipline can experience dramatic changes as scientists look for new ways to magnify the anomaly.

In summary, we have explained the meaning of resources of a science discipline as two broad categories: Human Resources and Knowledge Resources. These categories are consistent with contemporary literature (NSB, 2007; NAP, 2010, RAGS) recommending changes to the practice of science in order to increase scientific progress classified as “revolutionary.” Driven by our research questions to understand how e-Infrastructure development is affecting scientific discovery, we will examine changes to these categories of resources and their linkages to other categories of a science discipline.

**2.2.5 Summary: Science Discipline Properties and Patterns**

In summary, we have described properties of a science discipline and elaborated on the meaning of evolutionary and revolutionary science based on concepts and patterns from Kuhn (1996) and Popper (1959, 1963). Table 3 below provides a list of the properties and patterns that we will use to observe the effects of e-Infrastructure development on a science discipline.

Properties of a Science Discipline	Descriptions
Community of Scientists:	<ul style="list-style-type: none"> <li>• Defined by its problems and puzzles, and how it practices its science;</li> <li>• Leverages achievements to builds its body of knowledge and grow its memberships;</li> <li>• Practices science from an agreed upon set of paradigms.</li> </ul>
Methodology:	<ul style="list-style-type: none"> <li>• Derived from a paradigm;</li> <li>• An accepted process for data collection, analysis, comparison, synthesis;</li> <li>• Rules to determine acceptable solutions;</li> <li>• Constructed to uncover phenomena.</li> </ul>
Problems and Puzzles:	<ul style="list-style-type: none"> <li>• Designed to support the paradigms of a discipline;</li> </ul>

	<ul style="list-style-type: none"> <li>• Articulate the paradigm from which they were derived by the modes of Scientific Practice: <ul style="list-style-type: none"> <li>○ Normal (evolutionary) Science</li> <li>○ Extraordinary Science</li> <li>○ Paradigm Shift</li> <li>○ Scientific Revolution</li> </ul> </li> </ul>
Resources:	<ul style="list-style-type: none"> <li>• Human</li> <li>• Knowledge</li> </ul>

**Table 3 Properties and Patterns of a Science Discipline**

Similarly, Graham et al (2002), using Kuhn’s modes of scientific practice, found patterns that represented the different trajectories taken by the discipline of ecology as it achieved evolutionary or revolutionary progress. Trajectories they observed were interpreted in the following way: Specialization occurred during normal science; Refinement or Conceptual Evolution occurred during paradigm shifts; and in the final trajectory, paradigms were either abandoned or usurped during revolutions.

### **2.3 Integration of Concepts**

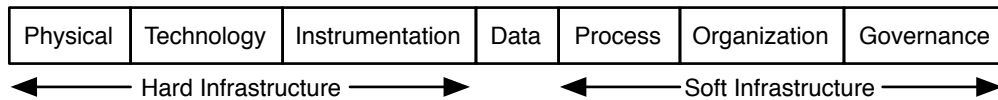
Previously, we characterized the development of e-Infrastructure as a socio-technical arrangement. Concepts and patterns of infrastructure development were identified, which we argued could be adapted to detect patterns of e-Infrastructure development. Concepts and properties of infrastructure were categorized as hard and soft infrastructures.

Hard infrastructures referred to physical infrastructure, such as transportation systems, water supply systems, telecommunications systems, etc. In relation to a science discipline, we further categorize hard infrastructures into the following three categories: Physical, Technological, and Instrumentation. Soft infrastructures referred to the rules and norms, policies, cultural and social standards of a community. We further categorize soft infrastructures into the following three categories: Process, Organization and Governance. These six categories will

represent hard and soft infrastructures that collectively form an e-Infrastructure as a socio-technical arrangement.

We will add a seventh category, which we label “Data.” By design, data is positioned in the center of our model in relation to the other artifacts, because it plays a central role in contemporary e-Infrastructure development in science disciplines. Figure 2 shows our conceptualization of e-Infrastructure development as a socio-technological arrangement of hard and soft infrastructures.

The demands for digital data are changing how science is conducted. For example, the National Science Foundation sees data as more than a product of research. “Digital data are not only the products of research, but provide input to new hypotheses, enabling new scientific insights and driving innovation (NSF, solicitation 07-601).”



**Figure 2 Conceptualization of e-Infrastructure Development**

While data may logically seem more soft than hard (since data acquires its meaning through symbolic representation), our preliminary observations suggest that data plays a mediating role between hard and soft artifacts, and therefore in scientific progress and in evolving disciplines. In our preliminary investigations we have found data to also have a mediating role between disciplines, such as between biodiversity and crystallography (Beach interview, 2008). Crystallographers combined biodiversity data with their data to create new knowledge that was revolutionary for the crystallography discipline.

We will use this model to construct a theoretical lens (Chapter 4) that expresses the relationship between a science discipline and e-Infrastructure. This, in turn, will guide us in answering our research questions concerning the relationship between e-Infrastructure development and scientific discovery.

## CHAPTER III Theory Construction: Theoretical Underpinnings

### 3. Theoretical Underpinnings for This Study

Chapter 2 applied existing concepts to pose new research questions about how e-Infrastructure development has influenced scientific discovery.

Chapter 3 will:

- Describe how changes to a paradigm reflect advancement in knowledge.
- Examine the changes that paradigms experience in the context of scientific practice (Kuhn, 1996 Popper, 1959; 1963).
- Review how these changes were explained in the concepts of Evolutionary Science, Extraordinary Science, Paradigm Shift and Scientific Revolution.

In order to answer our research questions, Chapter 3 offers a literature review supporting the conceptual framework developed in the preceding chapter. In Chapter 3, we:

- Discuss experimental design and the practices that scientists strive to use to create knowledge when working within their paradigm. We base this discussion on Churchman's "modes of inquiry," scholarly research that explains different ways of creating knowledge.
- Refer to Darwin's theory and other theories of co-evolution to support our ideas about the interactions between a science discipline, e-Infrastructure development, and mutual reciprocal shaping.
- Refer to the concepts of Advanced Structuration Theory (AST) to explain the structure of advanced technologies in the context of e-Infrastructure. Our objective here is to establish a link between AST theories and scientific practice, and the shaping that occurs through this interaction.

From these concepts and their vocabulary, we will shape and apply our observations of the effects of interactions between e-Infrastructure and a science discipline.

### 3.1 Properties of Science Disciplines: A Philosophy of Science Perspective

Science disciplines are under great pressure to produce the kinds of discoveries that yield huge economic gains nationally and globally. National science policy is promoting the use of new scientific approaches to foster these revolutionary discoveries (NSB, 2007). The purpose of this section is to explain how classical science and the way people think about science is shifting in response to this pressure. We will do this by:

- Illustrating these shifts through the example of data mining and the ways in which it is changing how science is practiced.
- Building upon the modes of inquiry defined by Churchman in order to make sense of the changes we are observing in classic science.
- Referring to Churchman's modes of inquiry to support our ideas of co-evolution between a science discipline and its infrastructure.
- Utilizing the theories behind co-evolution, and the work of Churchman and other science philosophers to understand how infrastructure development is affecting science, its processes, and the way people think about science.

#### 3.1.1 How Is Knowledge Created?

Churchman (1971) characterized knowledge as a product of inquiry. He defined inquiry as an activity that produces knowledge.

#### **Data, Information, and Knowledge**

The terms data, information and knowledge are interrelated and frequently overlap in their usage. Aamodt and Nygaard (1995) discuss the three concepts from two complementary and interdependent perspectives: The roles they take in a decision-making step, and their frame of reference. From the perspective of inquiry,

Churchman (1971) refers to a decision-making step as “an activity” or “potential action”.

### **Data**

Data are often the result of measurement (or observation), which are codified using a syntactic convention. Data on its own carries no meaning. Aamodt and Nygård (1995) refer to data as an initial set to decision making. To become information, data must be interpreted and take on meaning.

### **Information**

Information is data with meaning (Aamodt and Nygård, 1995). When data are interpreted and take on meaningful form, then the object for that form is referred to as “information” (Churchman, 1971). An example comparing the relationship between data and information could be the following: The stream of bits of the Ethernet protocol could be classified as “data,” whereas a book that describes the concepts and the properties of the Ethernet protocol could be classified as “information.”

Churchman (1971) considered knowledge to be “a collection of information, or as an activity, or as a potential.” Let’s consider each of these characterizations of knowledge:

#### **3.1.1.1      *Knowledge as an Activity***

An activity can result from the reaction that a user has to a collection of information (Churchman, 1971). Churchman tells us “it is how the user reacts to a collection of information that matters.” It is this reaction that enables the user to establish a link between known information and new information to create knowledge. Aamodt and Nygård (1995) describe knowledge as having an active part in the processes of transforming data into information (referred to as data interpretation), deriving other information (referred to as elaboration), and acquiring new knowledge (referred to as learning).

Churchman (1971) classified the “action conception of knowledge” as “pragmatic”. In other words, “knowledge is an ability of some person to do something correctly,” and therefore, to achieve a desired outcome.

Knowledge, as a pragmatic activity resulting from a reaction, can be illustrated with an example obtained from a personal communication with James Beach (2007) involving two communities in chemistry: crystallography and bio-chemical pharmaceutical.

- Crystallographers study protein structure by crystallizing them, then doing X-ray diffraction. This is how they determine the structure of proteins. They did this for a long time, because they were interested in the structure of proteins.
- Biochemical pharmaceutical scientists recognized that the crystallography data contained a wealth of information, which they could use for their drug discovery and development programs to produce pharmaceutical products.

Knowledge resulted when the biochemical pharmaceutical scientists reacted to the crystallography data as relevant and meaningful enough to link with their own data in order to solve some of their problems and puzzles.

#### **3.1.1.2**      *Knowledge as Potential*

Churchman (1971) argued that it is the user that possesses knowledge, and not the collection of information. A collection of information by itself would not be considered knowledge, because it does not provide a meaningful result without the potential action from its users.

For example, a library — as a collection of meaningful information — and its users possess a potential for knowledge, if the actions of its users (their inputting of queries, for example) result in meaningful responses (Churchman, 1971).

### 3.1.2 Modes of Inquiry to Produce Knowledge

What are modes of inquiry to produce knowledge? Modes of inquiry refer to different approaches towards creating knowledge. Earlier, we referred to “inquiry” as “an activity, which produces knowledge (Churchman, 1971).” In this section, we explore how different scientists approach inquiry and the creation of knowledge.

Churchman (1971), in his classic work *Design of Inquiring Systems*, presented four inquiring systems, each using different approaches whose purpose are to produce knowledge for its human user. We refer to them as modes of inquiry, because each approach operates under a different set of conditions. This will be further explained in the next sections where we will discuss each mode of inquiry.

Karl Popper’s *Critical Rationalism*, also known as “hypothesis-driven science” (O’Malley et al, 2009), shall be included in our discussion on modes of inquiry. Hypothesis testing is considered to be at the core of classical scientific practice (O’Malley et al, 2009). Evolutionary science practice, recognized as the dominant scientific practice (NSB, 2007), is based upon the hypothesis mode of inquiry.

Let us take a closer look at the differences between these modes of inquiry.

#### 3.1.2.1 *Theory Leads, Data Follows — The Leibnizian Mode of Inquiry*

The Leibnizian mode of inquiry is based upon the following set of assumptions. First, knowledge is built from “innate ideas.” The inquirer starts off with a knowledge base, and from that knowledge base can construct new knowledge (Churchman, 1971).

This mode of inquiry is consistent with classic, hypothesis-driven science. In hypothesis-driven science, a scientist starts by developing a hypothesis based upon prior knowledge. The hypothesis is written as a statement of fact, which is then subjected to experimentation to confirm the hypothesis (Glass, 2007). Popper (1961) believed that it was much more difficult to confirm a hypothesis than to disconfirm it. He developed the principle called “critical rationalism,” which advocates framing experiments with a falsifiable hypothesis (Glass, 2007). A single



experiment, showing an inconsistency based upon prior knowledge, was sufficient to disconfirm the hypothesis.

A confirmed hypothesis becomes a true statement, which is then added to the knowledge base of “innate ideas.” Churchman refers to this knowledge base as “fact nets.” Fact nets grow by linking truths to them, creating new fact nets. The hypothesis-driven approach to scientific practice builds on the achievements of a science discipline (Kuhn, 1996) in order to expand its knowledge base. Churchman (1971) refers to this knowledge base as a storehouse of fact nets interlinked through appropriate relationships.

### **3.1.2.2**      *Data Leads, Theory Follows — The Lockean Mode of Inquiry*

In the Lockean mode of inquiry, fact nets (knowledge) are formed from “observation” sensed in the environment outside the inquiring system. This is a deviation from the Leibnizian mode of inquiry that is based upon truth within its knowledge base. Knowledge in a Lockean inquirer is created from sensory inputs.

The initial state of a Lockean inquirer has no knowledge, “no a priori information” about the outside world in its memory (Churchman, 1971). It starts with a clean slate — “tabula rasa.” The inquirer has the ability to receive an input, and code it as data. It can perform logical operations on data, such as combination, to form knowledge.

The Lockean mode of inquiry creates knowledge through consensus. Churchman (1971) explained that the goal of a Lockean inquirer is to create knowledge based on empirical data accepted by its community. Lockean inquirers attempt to achieve consensus by means of induction from community agreements about observations (Churchman, 1971).

Induction is reasoning (assertions) from a limited set of personal experience (Glass, 2007). In hypothesis-driven science, many believed using inductive reasoning to confirm theories from personal experience would lead to errors, because data was not sufficiently trusted upon which to base hypotheses. Datasets based largely on

inductive reasoning were not a good fit for hypothesis-driven science, as criticized by Popper, Kant, Hume and other science philosophers.

In modern science, as datasets have become much larger and data management tools have become more powerful (e.g., data mining), scientific communities, with the use of these tools, are influencing the development of theory from data. An example of this change occurred in the biology discipline when sequencing the human genome. Glass (2007) explained that it was not until enough data had been interpreted that it was possible to formulate a hypothesis on the basis of the human genome.

### 3.1.1.2.3 *Theory and Data Shape Each Other: Singerian Form of Inquiry*

In the Singerian mode of inquiry (after E. A. Singer), theory is confirmed by using data. This is consistent with the Leibnizian mode of inquiry, because Leibniz confirms theory from prior knowledge. The point of departure between the Singer and Leibniz approaches is where data is used to create theory, and also to confirm theory.

The Singerian mode of inquiry is consistent with modern sciences' approach to working with data and theory. Modern science practices embrace methods that start by finding patterns in data to confirm theory, and conversely start with theory, then generate data to validate a hypothesis consistent with theory.

The role of data in the Singerian mode of inquiry is consistent with the role of data in modern science, where it has gained more prominence. This observation links back into our notion about e-Infrastructure and our 7-artifact model where data is becoming central.

## 3.2 Co-Evolution Theory

In Chapter 2, we formulated our ideas on the properties of a science discipline and of infrastructure. Properties of a science discipline were based upon a review of the literature of the philosophy and history of science (Kuhn, Popper, others).

Properties of infrastructure were based upon literature on the history of large technological systems (Hughes), and the literature of the construction of technological systems (Bijker, Hughes and Pinch, 1994; Star and Ruhleder, 1995; Edwards et al, 2007).

Our secondary research question inquires about the interaction between a science discipline and e-Infrastructure:

*How are the problems and puzzles of a science discipline shaping the development of e-Infrastructure, and conversely, how is e-Infrastructure changing the problems and puzzles of science discipline?*

To answer this question, we will borrow concepts from two sources:

- Biological co-evolution in natural selection. We adapt the work of Campbell (1960, 1969) and Aldrich (1979), in their use of co-evolution and natural selection in the domain of organizational change.
- Sociology and the work of DeSanctis and Poole (1994) on Adaptive Structuration Theory (AST).

We shall use concepts from these two sources to develop a theoretical lens for examining the relationship between a science discipline and e-Infrastructure.

### **3.2.1 Biological Co-Evolution and Natural Selection**

Campbell (1969) applied concepts of variation, selection and retention of biological natural selection for the study of social organization. Karl Weick (1979) drew from the work of Campbell to develop a social psychological theory of how individuals coordinate their actions, based upon the variation, selection and retention logic of evolutionary models (Murmann, 2003). Aldrich (1979) extended the work of Weick from the individual to the level of entire organizations to develop what he called the Population Ecology Model.

Principles from Thompson (1994) on **reciprocal evolutionary change** between interacting species driven by natural selection, could provide insights to help us

abstract the interactions between the components of a science discipline and the components of e-Infrastructure. We could then determine if these interactions then resulted in reciprocal change.

### 3.2.1.1 *Biological Co-Evolution*

The concept of co-evolution can be traced back to Darwin, in 1862, when he observed the phenomenon of reciprocal mutual influence between orchids and pollinators. Erlich and Raven (1964) later formulated the process of co-evolution as they recognized patterns of interaction between two or more biological species, which resulted in the creation of a close and evident ecological relationship.

Thompson (1994) formally defined biological co-evolution as “reciprocal evolutionary change between interacting species driven by natural selection.”

Pazos et al (2008), working on protein co-evolution, expanded on this definition by adding that each interacting species’ evolution is linked to the other species through a process called *reciprocal mutual selective pressure*. Interacting species include host and parasite, predator and pray, plant and herbivore, etc. (Blerkom, 2003). For example, herbivores, dependent on plants for food, evolve traits to obtain this food despite the evolution of a diverse arsenal of plant defenses. The reciprocal relationship between herbivores and their host plants results in the evolution of traits by each in a defensive response to selection pressures.

Based on Thompson (1994) and Pazos et al (2008), Biological Co-evolution refers to two or more biological species that exert selective pressures (change) on the other, where each affects the others’ evolution (reciprocal evolutionary change).

Selective pressures are conditions that bring about (trigger) a biological selection process (Thompson, 1994). Conditions known to trigger biological selection include limits on resources (e.g., nourishment, habitat space, mates) and the existence of threats (predators, disease, adverse weather). The selection process results in preventing some species from surviving, while allowing others to do so. The surviving and propagating species pass on traits to the succeeding generation.

The natural selection process can increase the prevalence of traits. Traits that increase reproductive success of a species are selected, whereas those that reduce success are not selected.

In summary, we have described Biological Co-evolution as resulting from a natural selection process. Conditions in the environment trigger a biological selection process, which brings about interactions between two or more species to exert pressures on the other. If the interaction between the species is reciprocal, then it's co-evolutionary. Going forward, we will introduce theories based on natural selection and biological co-evolution to describe the interactions between Aspects of a Science Discipline and e-Infrastructure development.

### **3.2.1.2 *Population Ecology Model (Natural Selection)***

Aldrich (1979) defines the Population Ecology Model as being based on the natural selection model of biological ecology. This model examines the nature and distribution of resources in organizations' environments to explain organizational change (Aldrich, 1979).

From an organizational perspective, change can be either internal or external to an organization. Organizations exist within environments. Environments exert pressures that result in competition between organizations for resources and survival. This pressure is a driving force in organizational activities. Resource dependence puts pressures on organizations (disciplines) to focus on tactics and strategies to manage their resources within the changing environment.

Resource Dependency Theory and Population Ecology Model provide us with concepts that underpin our ideas about co-evolutionary change between the Aspects of a Science Discipline and e-Infrastructure development. Resource Dependency Theory is described in section 3.2.4. Population Ecology Model is described in the remainder of this section. The discussion will focus primarily on the level of the science discipline.

The Population Ecology Model provides a methodology we can adapt to guide us in observing change in organizational forms affecting the practice of science.

According to Aldrich (1979), “change may occur either through new organizational forms eliminating old ones or through the modification of existing forms.” Aldrich (1979) characterizes organizational forms as consisting of elements, which refer to goals, boundaries and activities. These elements are subject to the selection process of the environmental selection criteria. What emerges, based on the model, is change that characterizes new or modified organizational forms.

### **Impact on Science Disciplines**

The process of natural selection means organizations are moving toward a better fit with their environment. Natural selection is a property of the environment resulting from change and pressures on the organizations (disciplines or groups of a particular discipline) in the environment. The property of natural selection is complementary to Popper’s and Kuhn’s concepts regarding change to paradigms and their impact on disciplines. Changes in the environment can result in change being induced on the organizations (disciplines, communities) in the environment. Environmental constraints can come from both internal and external pressures to the environment.

- Internal environmental constraints come from pressures generated within an environment, such as reduction in funding.
- External environmental constraints come from pressures generated outside an environment; for example, the introduction of e-Infrastructure development to create an outcome of dramatic increases in scientific progress.

Aldrich (1979) explains that the “natural selection model is general and can be applied to any situation where the three stages are present.” This three-stage model of variation, selection and retention can be used to describe trial and error learning, organic evolution, and socio-cultural and organizational evolution (Aldrich, 1979)

**Stage 1: Variation Process:** In biological systems, a change in either environmental conditions or the inner structure of an organism triggers a variation process. Aldrich (1979) defines the variation process as “the first stage in the natural selection process.” In a variation process, change occurs through processes of recombination or mutation. Recombination refers to new combinations of genes. Mutation refers to change in genes from one form to another (Levine, 2000). Occurrence of variation could be random or predicted. Recombination and mutation generate random variability within a biological environment (Levine, 2000). Genetic variability results in variation in individual traits and gives rise to differential fitness of individuals. Furthermore, traits can be inherited. Variation leads to the selection of organisms that are most suitable and most fit, based on the selection criterion (Aldrich, 1979). Recombination in science can be enhanced through increasing exchange of ideas between disciplines, improving communication and collaboration technology, sharing data, technology and methods, and enhancing student and research exchange programs (Aldrich, 1979).

Variation can result from scientific breakthroughs leading to increases in technological change, which can then set off periods of change where a discipline becomes more fit. For example, genomics became more fit from technology breakthroughs after the human genome project. A variation process occurred, which then led to the formation of the sub-discipline of “comparative genomics” (Robbins, 1996).

**Stage 2: Selection Process:** Stage 2 of the natural selection process takes place when the operation of selection criteria are performed that selectively eliminates certain types of variations (Aldrich, 1979). From the perspective of scientific practice, a science discipline sifts through different theories or paradigms until finally selecting one that is able to adapt (or fit) to the pressures exerted upon it.

We shall borrow from Aldrich's (1979) concepts for our conceptual framework to help us observe how a discipline would apply to a selection process. Aldrich (1979) explains that in organic selection, certain mutant forms are more fit to exploit the resources in their environment and, as a result, are able to survive the resource-based selection criteria. Organisms unable to survive the selection criteria are eliminated. Aldrich (1979) explains that selection criteria are based upon constraints in the environment. Organisms (organizations or communities in science disciplines) that fit environmental criteria are positively selected and survive, while others either fail or change to match environmental requirements (Aldrich, 1979).

External constraints/stimulus change the selection rate by impacting the extent of resource abundance in the environment. Variation arises through organizations' active attempt to generate alternatives and seek solutions to problems (Aldrich, 1979) within the organization or in its external environment. When resource abundance is high, the degree of organization formation has the potential to increase (Specht, 1993) and likewise, the rate of selection is likely to decrease. Organization formation is a type of between-organization variation (Aldrich, 1979). From the level of a science discipline, if resource abundance is high (e.g., increase in federal budget for research), the rate of variation could potentially increase between disciplines or between communities in the same discipline (Aldrich, 1979). A possible example is the rate of formation of sub-disciplines or virtual organizations for sharing of ideas or collaborations; the rate of investments forming projects or organizations to deploy new technologies and form e-infrastructure to support these projects; the rate at which policies are changed that ease intellectual property rights in order to stimulate the formation of projects or virtual organizations involving foreign counterparts. On the other hand, if resource abundance is low, the variation rate could remain high, reflecting a high degree of competition for resources; e.g., an open call for proposals by a federal agency resulting in a high



degree of proposals submitted, but a low number of awards, due to lack of funding. The culling of proposals indicates an increase in selection rate.

**Stage 3: Retention Process:** Stage 3 involves the selective retention of the variations that survived the selection process (Aldrich, 1979). “Retention occurs when selected variations are preserved, duplicated, or otherwise reproduced so that the selected behavior is repeated on future occasions or the selected structure appears again in future generations (Aldrich, 1979).” In organic evolution, selective retention transfers the traits of selected variations to succeeding generations of plants and animals.

At the level of a science discipline, retention results as time passes. For example, selected variations of problems and puzzles, or knowledge resources of the human genome project in the discipline of microbiology were transferred as traits to the succeeding sub-discipline of comparative genomics. Retention is consistent with the concept of path dependence in the context of infrastructure.

According to Campbell (1969), “When the three conditions of the 3-stage model of variation, selection and retention are met, an evolution in the direction of better fit to the selective system becomes inevitable.”

### **3.2.2 Relationship Between a Science Discipline and e-Infrastructure Discussion**

Our approach is to use Variation, Selection and Retention processes as a pattern of co-evolutionary change. Aldrich (1979) compared the processes of the population ecology model with processes of social organization. His claim was “if the processes are isomorphic — similar in form — then many of the insights of evolutionary theory can be used to understand organizational change (Aldrich, 1979).”

Our approach is to include Variation, Selection and Retention in our Concept Map, as functional components, to guide us in observing change resulting from the interactions between a science discipline and e-Infrastructure development activities.

Let us consider the components of the model applied to a science discipline:

**Environment:** We consider the environment as encompassing a science discipline. Earlier, we abstracted a science discipline to consist of a community, problems and puzzles, methodologies, and resources. These elements are subject to selection criteria where change could be measured based on pressures from the environment.

**Variation:** Aldrich (1979) explains that for organizational change to occur, variation must be present within and between organizations. Variation is triggered through organizations' active attempt to generate alternatives and seek solutions to problems within the organization or in its external environment (Aldrich, 1979). E-Infrastructure is a potential for increasing resources in a science discipline or its environment. E-infrastructure offers resources (physical, technology, instrumentation, data, processes) to a science discipline in order to increase variation towards solving problems and puzzles, in particular complex problems and puzzles, with the potential to lead to breakthrough discovery. In science, the effect of high variation could have a positive impact in increasing the number of funding opportunities (niches) and competition for the resources in those niches.

For example, a community of scientists and a resource provider collaborate to derive a solution that augments the capability of an instrument through the use of e-Infrastructure technology. This instrument consisted of environmental sensors. The e-Infrastructure technology combined wireless technology with the environmental sensors, enhancing their capabilities to collect and transmit data in real-time to scientists. This technology created an e-Infrastructure that enabled biodiversity and ecology scientists to see deeper and answer questions they were unable to before (Porter et al, 2005).

**Selection:** Using the population ecology model, Aldrich (1979) observed the selection of new or changed organizational forms that resulted from environmental constraints. e-Infrastructure can enable an organization to become more fit to survive the selection process. During variation, organization formation or

restructuring occurs to become more fit to compete for resources and survive the selection process. Linkages to e-infrastructure providers are one approach to acquiring e-infrastructure resources. For example, if the selection criteria are to increase data sharing, organizations more fit to provide approaches to increase data sharing are more likely to acquire resources.

**Retention:** During retention, e-Infrastructure potentially accelerates the transfer and replication of the traits of the organizations that survived the selection process. Continuing the example using data sharing, once XML was adopted as a standard, the infrastructure of the Internet accelerated its adoption. XML became the standard for describing metadata, facilitating the exchange (sharing) of data sets.

**Summary:** We described the concepts of co-evolution developed by Thompson (1994) and Aldrich (1979), and proposed to adopt these concepts to guide us in observing change resulting from the interactions between a science discipline and e-Infrastructure development in its environment.

These concepts provide us with a framework to answer our secondary research question — *How are the problems and puzzles of a science discipline shaping the development of e-Infrastructure, and conversely, how is e-Infrastructure changing the problems and puzzles of science discipline?* — involving the co-evolutionary relationship between a science discipline and e-Infrastructure development as they co-evolve from pressures that are both internal and external to the environment.

### 3.2.3 Adaptive Structuration Theory

We borrow concepts from Adaptive Structuration Theory (AST) to provide a lens by which we can study three effects of the interactions between a science discipline and e-Infrastructure development, and the outcome of scientific progress.

1. The first effect we consider is: **Variations in the change of science disciplines that occur, as e-Infrastructure development (advanced technologies) is used (DeSanctis and Poole, 1994).**

AST focuses on the concept of advanced information technologies (e-Infrastructures). For this discussion, let us use Advanced Information Technology (AIT) interchangeably with e-Infrastructure. It distinguishes them from other technologies as having properties capable of creating changes to traditional structures (transformative properties). E-Infrastructures hold the “potential to change traditional organizational design (DeSanctis and Poole, 1994).”

On the one hand, e-Infrastructures can be appropriated (fashioned) into a system (structure) of orchestrated (coordinated) technologies (Arthur, 2009) to suit the organization’s (discipline’s) structures, processes, norms, rules and resources. New structures also can emerge in an organization (discipline), based upon human action, as people interact with e-Infrastructure (DeSanctis and Poole, 1994).

2. The second effect we consider comes from AST providing a lens by which to focus on the **issues of uptake and use of e-Infrastructure**. DeSanctis and Poole (1994) explain that the effect of AITs is more a function of how they are used by people if people are able to adapt AITs (e-infrastructure) to their particular work needs, or fail to use them. Hine (2006) and Woolgar et al (2006) made similar arguments.
3. The third effect we consider results from the **range (continuum) of potential interactions between groups, communities**, etc. from the use of AITs. In addition to supporting traditional discrete technologies, AITs focus on coordination among people and provide procedures for accomplishing complex interpersonal exchange (DeSanctis and Poole, 1994). Variation from the uptake and use of e-Infrastructure by science groups and communities is likely to occur, in particular between different disciplines.

AST and its concepts of social structures — and the model it provides describing the interplay between AITs, social structures and human interaction — will provide a

potential lens by which we can better understand how e-Infrastructure development is changing science disciplines.

AST could be helpful in guiding us to see different forms of interactions among communities of scientists, and the structures they generate. DeSanctis and Poole (1994) refer to these forms as “structural potential.”

#### **3.2.4 Resource Dependence Theory**

The study of resources and their relationships to the behavior of organizations and their environment led to the development of the Resource Dependence Theory (RDT). This study explores resources from two perspectives: (a) resources of a science discipline; and (b) resources resulting from e-Infrastructure development. Our objective is to borrow concepts from Resource Dependence Theory to create a conceptual lens by which to identify patterns and relationships between the Environment of a Science Discipline, Aspects of a Science Discipline and e-Infrastructure Development. Our conceptual lens is defined in Chapter 4: Conceptual Framework.

Resource Dependence Theory (RDT) informs us that environmental pressures, such as the lack of funding or absence of new discoveries, will cause organizations (science disciplines) to seek ways to acquire access or control over resources (Boyd, 1990). Population Ecology Theory (PET), introduced earlier, also focuses on resources, but from the perspective that ecological constraints on resources lead to variation, selection and retention processes in the environment. In chapter 2, we characterized resources of a science discipline as consisting of human and knowledge resources. E-Infrastructure development, which we characterized external to the environment of a science discipline, has Physical, Technology and Instrumentation resources. In this section, we will study concepts to help us identify and understand how science disciplines acquire resources as a means towards discovery, based upon RDT and PET.

RDT defines relationships between organizations (disciplines), resources, and their environment. The environment exerts external forces (constraints) on organizations (disciplines), causing change and adaptation (Pfeffer and Salancik, 2003).

A *niche* refers to “resource space” in an environment (Specht, 1993; Hawley, 1988). A niche forms from distinct combinations of resource utilization and constraints on resources. A niche attracts organizations to exploit the niche’s resources (Specht, 1993; Brittain and Freeman, 1980). As organizations exert control over resources from the environment, they establish boundaries to contain those resources (Katz and Gartner, 1988). *Boundaries* serve to distinguish one organizational population from another. Organizations, as they form, create boundaries to establish identity, such as physical and/or legal identity (Katz and Gartner, 1988).

Our approach is to use these concepts by Aldrich (1979), Dess and Beard (1984) and Specht (1993) to identify conditions in science disciplines that results in the formation of niches. We will want to observe if e-Infrastructure development is creating such conditions that result in increasing niches and resources.

*Linkages to External Resources:* RDT states that the need for external resources and information drives the degree of organizational dependence on the environment (Boyd, 1990). Linkages to external resources and information are a function of organization dependence on the environment (Boyd, 1990). RDT states that an organization must gain control over resources in its environment in order for it to survive (Pfeffer and Salancik, 1978; Pfeffer, 1987). External pressures — such as competition, lack of funding, and social forces — will cause organizations to seek out linkages to external resources in the environment (Boyd, 1990). Our approach is to transfer this concept of “linkages to external resources” into our conceptual framework as a feature by which to observe linkages between a science discipline and e-Infrastructure development.

How would we employ these concepts? Boyd (1990) informs us that linkages to resources external to a discipline are high when munificence is low. In other words,

if environmental munificence<sup>7</sup> is high, meaning there is an abundance of resources within a discipline, then the number of linkages to resources external to the environment should be low. In our conceptual framework, we established that a science discipline has two types of resources: human and knowledge. Munificence within a discipline is low when its paradigm is under pressure and the discipline lacks resources to ease the pressure. A discipline will seek to establish linkages to external resources to ease the pressure on its paradigm. These linkages to external resources could take on the form of: (1) a linkage to expertise of another domain to import knowledge and human resources; e.g., the linkage can take on the form of a legal contract involving employment or intellectual property, etc. (2) an external linkage to technology that will provide a capability to the discipline; e.g., the technology could be from e-Infrastructure development. Reichman et al (2011) found that scientific workflows and data provenance technologies were introduced into disciplines to facilitate reproducibility.

*Performance:* RDT recognizes that as a result of external linkages, the performance of an organization should improve (Boyd, 1990). In our conceptual framework, we will measure “performance” in terms of scientific progress that is either evolutionary or revolutionary.

**Summary:** We have introduced the Resource Dependence Theory (RDT) and the Population Ecology Theory (PET) — two important theoretical frameworks — in order to recognize patterns and relationships between resources and scientific progress. In our conceptual framework, concepts from RDT and PET will be employed to observe impacts on resources from the effects of environmental pressures on a science discipline. The properties of Resource Dependence and Population Ecology Theories provide theoretical underpinnings for our conceptual framework, explaining the behaviors of science disciplines seeking resources in an

---

<sup>7</sup> Environmental munificence refers to the degree of resource abundance in the environment. It can be measured as a function of growth in a science discipline or domain (Boyd, 1990).

environment. Our Concept map uses these underpinnings to recognize relationships between resources and scientific progress.

### 3.3 Summary of the Literature

We finalize the review of the literature in a map of discussions in Chapters 2 and 3. Figure 3 shows the primary literature supporting the major components of our conceptual framework, that will be discussed in Chapter 4. The major components represented in the literature map are: e-Infrastructure Development, Aspects of a Science Discipline and Co-evolution.

The major threads of literature supporting our conceptualization of e-Infrastructure and its development are Large Technological Systems, Social Construction of Technology and Infrastructure, Substrate and Relational Properties of Infrastructure and Domains. The characterization of e-Infrastructure as a “socio-technical arrangement” is supported in connected streams of literature by Hughes, Pinch and Bijker, Law, and Star and Ruhleder. Domains and the pattern of domaining are represented as a separate stream connected to the process of e-Infrastructure development.

Four streams of literature are shown in the literature map that supports the concepts and properties in the Aspects of a Science Discipline: Properties of a Science Discipline, Modes of Scientific Practice, Resources of a Science Discipline, and the Philosophy of Science literature. Streams of literature for the Philosophy of Science are Knowledge Creation and Modes of Inquiry literature. It is from these streams of literature that we will construct a lens to observe patterns of discovery.

Finally, four streams of literature support the concepts and theories we use for the process of Co-evolution. Two streams (shown on the left) come from the biology literature: Biological Co-evolution and Natural Selection, which is a stream that brings in the Population Ecology model theory. Two other streams (shown on the right) come from the Organization and Technology literature, and Organization and the Environment literature. These two streams of literature are shown connected through the Population Ecology Model and the Resource Dependence Theory.



We will now connect the concepts and properties we have derived from these streams of literature into a cohesive idea.

In Chapter 1, we argued that nations have been making significant investments in ICT infrastructures with the intention of stimulating dramatic increases in scientific progress, with the hope of achieving discovery. By achieving discovery, the hope is that these investments will result in technological innovation and economic prosperity for these nations.

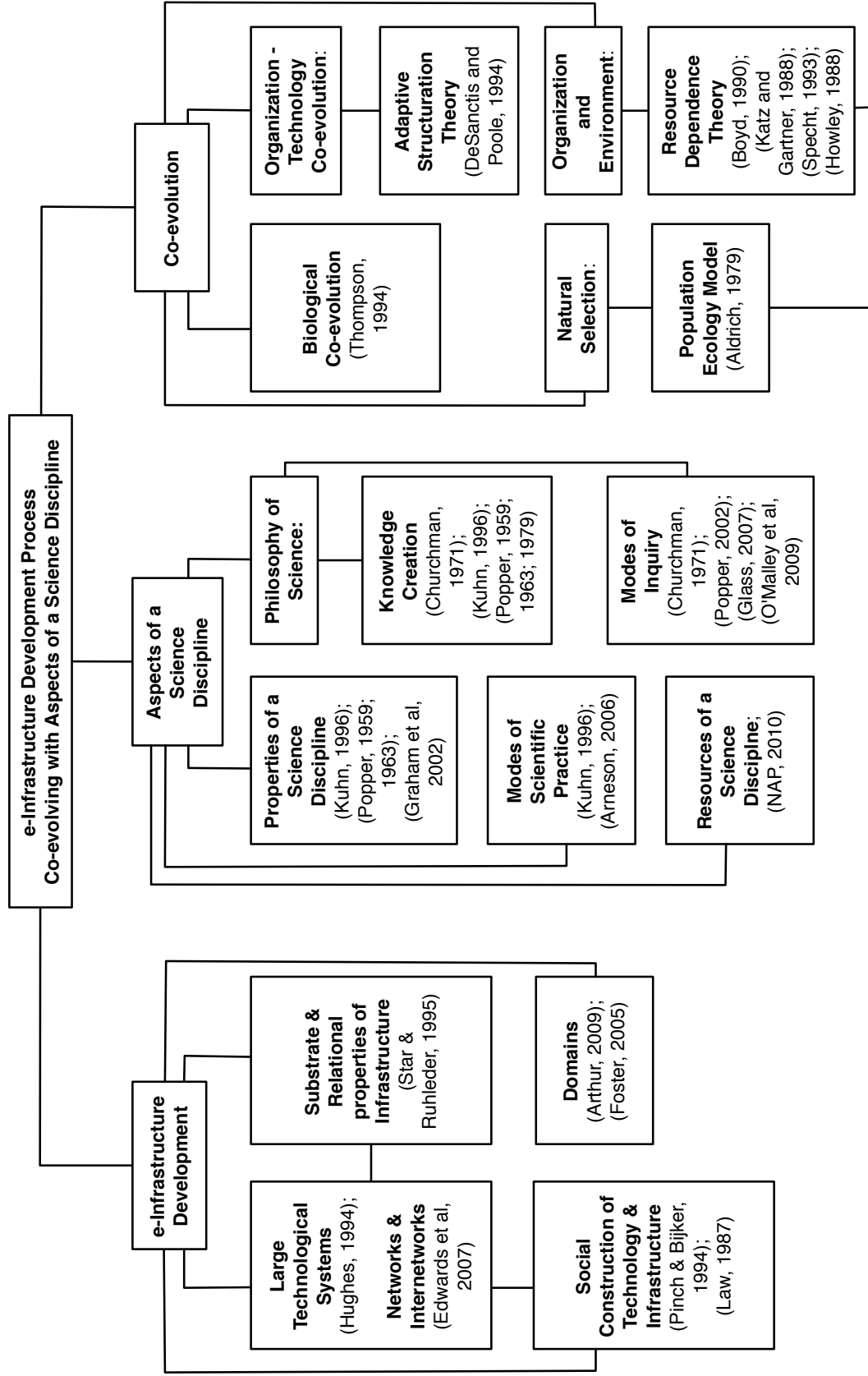
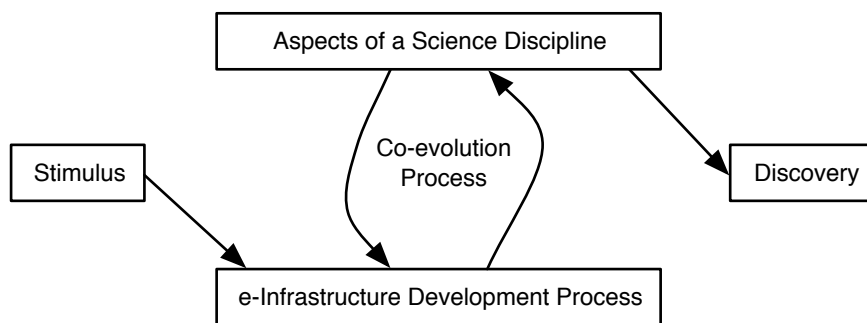


Figure 3 Literature Map of major literature streams

In Chapter 2, we introduced concepts and properties on e-Infrastructure development and Aspects of a Science Discipline. We conceptualized e-Infrastructure development as a social-technical arrangement. The concepts and properties of e-Infrastructure development were categorized as hard and soft infrastructures, and organized as a set of seven interconnected categories. Figure 2 conceptualized e-Infrastructure development and its seven categories.

Aspects of a Science Discipline were conceptualized as possessing the following properties: a Community of Scientists; Problems and Puzzles; a Methodology; and Resources. Modes of scientific practice described how these properties are inter-related. Knowledge creation and Modes of Inquiry provided concepts for how science is thought about and understood.

Concepts and patterns of co-evolution and the environment of a science discipline were described in Chapter 3 to inquire about the interaction between a science discipline and e-Infrastructure. Our conceptualization of these components and their interconnections are represented in the following figure.



**Figure 4 Conceptualization of co-evolution between e-Infrastructure development and Aspects of a Science Discipline**

Figure 4 shows the components of our preliminary theory of e-infrastructure development and connections to other components that could result in discovery.

The next chapter on the Conceptual Framework will describe each concept and provide operational definitions, describing how we intend to use each concept and how we will observe it for changes.

## CHAPTER IV: Conceptual Framework

### 4. Concept Map

Chapters 2 and 3 surveyed the literature on the history and philosophy of science for insights that could ground a theoretical explanation of e-Infrastructure development's impact on increasing scientific discovery. In this chapter, we construct a Conceptual Framework to guide the process of making sense of the effects of significant national ICT investments. These investments are intended to stimulate transformative research which, when supported by developing e-Infrastructure, lead to revolutionary discovery.

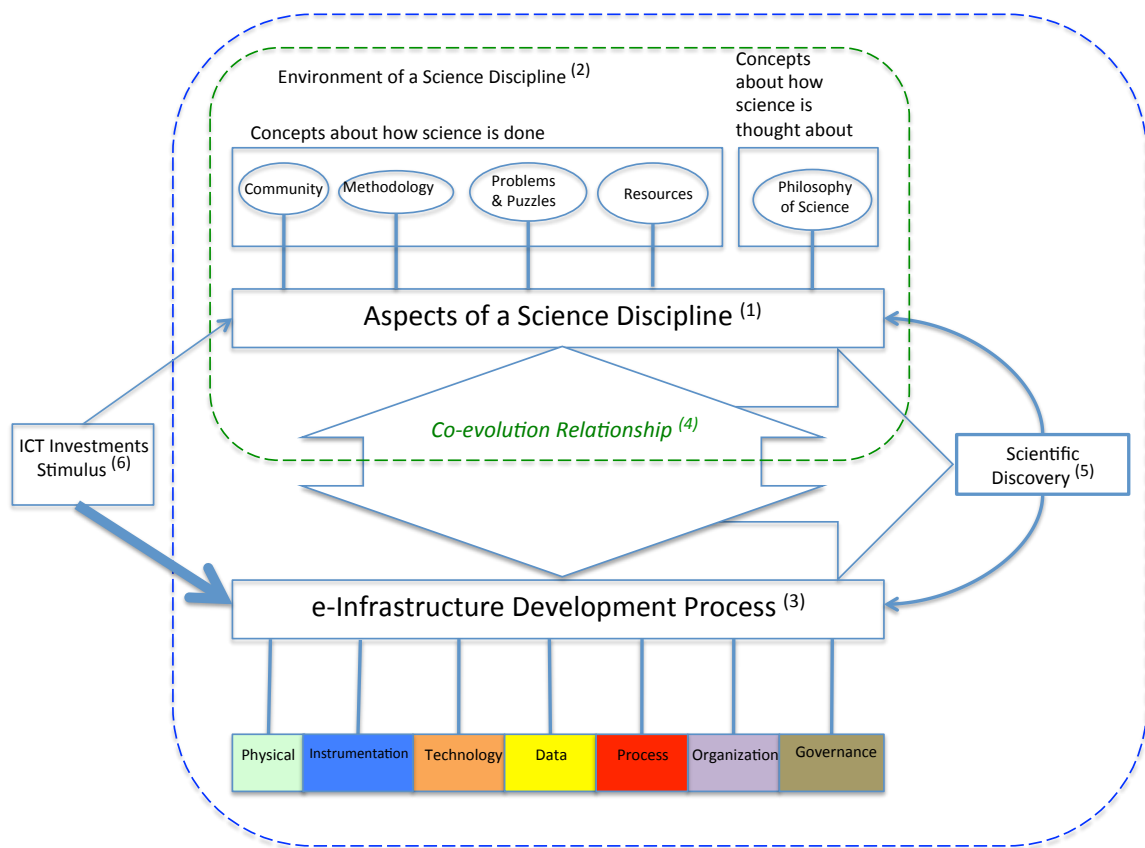
A Conceptual Framework is a key component of the design of a research study. Conceptual Frameworks serve to interconnect concepts, assumptions, expectations, beliefs and theories that support and inform the research of a study (Maxwell, 2005). A Conceptual Framework also represents a tentative theory for the phenomena that is being investigated (Maxwell, 2005).

A tool for the construction of a Conceptual Framework is called a "Concept Map." A Concept Map develops and clarifies a theory (Maxwell, 2005), while also providing a visual representation of that theory. It is also referred to as a "conceptual lens," because it enables its user to zoom in and out to make the proposed theory more visible. A Concept Map usually consists of two types of symbols: (1) labeled ellipses or rectangles representing concepts, and (2) arcs or lines, with or without arrows representing relationships between concepts.

Coupled with explanation, a Concept Map illustrates what a theory says is happening with the phenomenon being studied (Maxwell, 2005, Miles and Huberman, 1994).

#### 4.1 Introduction to the Concept Map

The Concept Map of our study is shown below in Figure 5. It describes six main components of our conceptual lens: (1) Aspects of a Science Discipline; (2) the Environment of a Science Discipline; (3) the e-Infrastructure Development Process; (4) the Co-evolutionary Relationship between a science discipline and e-Infrastructure development; (5) Scientific Discovery as a result; and (6) ICT Investment as a stimulus. Each of the six components is numbered in the Concept Map for easier identification.



**Figure 5 Concept Map of our Theoretical Framework**

In the remainder of this section, we describe the six components of the Concept Map, and their function within the overall framework. In the next section, we describe each of the concepts in the Concept Map, provide an operational definition of each concept, and explain how we plan to observe or measure each of the concepts. We conclude with a display of all of the components and each of their

concepts, a brief description of their operational definitions, and the plan for observing them.

### 1. Aspects of a Science Discipline

The Aspects of Science Disciplines are organized as two activities: (a) concepts about how science is done (the modes of practice in a science discipline), and (b) concepts about how science is thought about and discussed (Philosophy of Science). Based on Kuhn (1996), and upon the Literature Review (Chapter 2) and theoretical underpinnings (Chapter 3) of this study, we organized the concepts of how science is done into four dimensions: (a) **membership** that collectively defines a particular community of scientists, (b) **methodology** used by that community to conduct research; (c) **Problems and Puzzles that** community members work on, and (d) **human and knowledge resources** stemming from that community's achievements. The concepts determining how science is thought about are based upon the philosophy of science. This philosophy is included in our Concept Map, because it's important to understand how science inquiry is changing.

### 2. Environment of a Science Discipline

When an environment exerts pressures on a science discipline, it can result in competition for resources between sub-disciplines or other organizations. For example, governments can exert pressures on science disciplines through policies that control funding. We conjecture that e-Infrastructure development increases resources for a science discipline. Likewise, the requirements of a science discipline increase demand for resources in response to its goals. In turn, those goals exert pressure on e-Infrastructure. We will refer to the Population Ecology Model (Aldrich, 1979) and Resource Dependence Theory (Boyd, 1990) for concepts and relationships to guide our exploration of the environment of a science discipline.

### 3. E-Infrastructure Development Process

E-Infrastructure Development Process represents a set of categories derived from the literature streams on infrastructure and its development. These categories were

described back in Chapter 2. Hughes (1987) represented infrastructure development as a continuum of technologies or domains of technologies (Arthur, 2009). Moreover, these categories are socially constructed and can be adapted to function in a dynamic environment. Within our conceptual lens, the component of e-Infrastructure development shall guide us through linking events from stimulus to discovery.

#### 4. Co-evolutionary Relationship Between a Science Discipline and e-Infrastructure

The fourth component of our Concept Map concerns itself with the interactions between a science discipline and e-Infrastructure development. It is our conjecture that the Aspects of a Science Discipline and e-Infrastructure co-evolve as they mutually influence each other. Their relationship is represented with a bidirectional arrow to represent their mutual influence (shaping). The dashed line intersecting the arrow demarcates other influences either coming from within the environment of a science discipline to the e-Infrastructure, or vice versa. The arrow pointing to the right is a part of the fifth component, which we will describe next. When applied to different conditions, we theorized that the co-evolutionary relationship impacts whether the outcome follows a normal (evolutionary) process or a revolutionary process. What determines evolutionary versus revolutionary paths to discovery is also a significant part of the empirical work contained in this study.

#### 5. Scientific Discovery

The fifth component of our Concept Map guides us in our observation of the path from stimulus to discovery. This component of the Concept Map will help answer our primary research question of how the development of e-Infrastructure is impacting discovery. We theorized that a path starts from a stimulus — in particular, ICT investments. (We discuss characteristics of a stimulus in the sixth component). A path develops as it crosses between e-Infrastructure development and the Aspects of a Science Discipline. Directed arrows from discovery back to Aspects of a Science Discipline and e-Infrastructure Development Process represent



the outcome of the discovery emerging from the science discipline, the e-Infrastructure, or both.

## 6. E-Infrastructure Development as a Stimulus

The sixth and final component of our Concept Map focuses on how ICT investments have stimulated the development of e-Infrastructure, leading to dramatic improvements in scientific discovery. Here, we observe whether the intention is to develop e-Infrastructure through an independent initiative, such as Cyberinfrastructure or e-Science, or through a science discipline, as it develops instrumentation or adapts technologies as part of its practice. Arrows indicate that stimuli could be external to a science discipline (represented by the directed arrow towards the e-Infrastructure development process), or internal to a science discipline (represented by the directed arrow towards environment of a science discipline). We use a thicker arrow to represent our theory that ICT investments in e-Infrastructure development are greater than investments in the environment of a science discipline.

### 4.2 Concept Map Components: Concepts and Operational Definitions

We have described the six major components of the Concept Map. In this section we describe the concepts and properties within each component. We develop an operational definition for each concept that describes how we will use it and observe it for changes. Concepts will be shown in bold italic. Since we're interested in tracing a path from stimulus to discovery, our concepts must be able to tune in to (or understand) the effects caused by environmental pressures. Operational definitions given here are not exhaustive and they're subject to change as we use the Concept Map in our empirical work.

#### 4.2.1 Aspects of a Science Discipline

As described in Chapter 2, a Science Discipline contains the following concepts and properties: (a) *is a particular community of scientists*, (b) *has Problems and*

*Puzzles its members work on, (c) has a methodology, and (d) has resources that consist of knowledge and human resources.*

1. *“Is a particular community of scientists”* lets us know that we will be observing a subset of a broader science discipline. Scientists in a particular community engage in actions whose meanings they negotiate with one another (Wenger, 2002). We label a community of scientists as “particular” because of defining properties, such as a “community” sharing a particular goal, or working on a particular set of problems or puzzles. Each “community” uses a particular practice of normal science defined by its paradigm, etc.

**Observing a particular community of scientists:** The concepts we will use to observe a particular community of scientists are: (a) **Size**, (b) **Membership characteristics**, and (c) **Structure**.

The **Size** of a particular community of scientists will be measured over time. Measures we will use to characterize size are number of scientists, number of students, and number of organizations participating in the community while conducting research.

**Membership** measures the different forms of participation in a community of scientists. For example, “funded membership” means a community of scientists of an organization funded to participate and conduct research within a particular community. “Affiliated membership” means participation in a particular community through another organization that has an affiliation agreement with a particular community. “User membership” means joining a particular community as a user of its resources, e.g. data product. Membership characterizes the different types of groups that have joined a community.

**Structure** categorizes the different forms of a particular community of scientists. We will use a nominal scale for structure with values of low, medium and high. **Low** represents a community of scientists that is conducting local science. For example, a biological field station where scientists go to (physically) conduct

their own research has low structure. **Medium** means a community of scientists that is geographically dispersed, with some sharing of resources. An example of this would be a particular community of two or more field stations with scientists collaborating across stations and some degree of sharing resources. **High** has the properties of medium, but with a high degree of sharing of resources, such as sharing data.

- 2. Methodology:** A methodology within the Aspects of a Science Discipline refers to a procedure or prescription to perform an experimental design. Both experimental design and methodology are guided by paradigm (Glass, 2007; Kuhn, 1996). Graham et al (2002) argued that in the Ecology Discipline, methodology concerns itself with how data are collected and analyzed, and the standards by which they are compared. A community of scientists applies a methodology consistent with the paradigm that community uses. Case in point, if a community of scientists uses the Critical Rationalism paradigm, then the methodology normally involves making a hypothesis and then subjecting it to falsification (Popper, 2002; Glaser, 2007).

Our Concept Map does not concern itself with the specifics of the methodologies that a particular community of scientists uses in its practice. Instead, it guides us to observe a change in methodology and explore the cause of that change within a particular community of scientists. For example, what if 20 years ago a community of ecologists started to study the biodiversity of the Everglades and used a hypothesis-driven paradigm, such as Critical Rationalism, to design experiments? Second, what if data were collected and used primarily to test experimental results against hypotheses? Third, what if scientists who collected the data were also the primary users? Finally, what if 10 years into the program, science requirements changed and they demanded the use of data from other sources, such as remote sensing? Then, changes in science requirements or changes in policy are potential impacts on the methodology a community of scientists uses to achieve its stakeholders' goals.

**Observing change in methodology:** Observing change in methodology involves searching for influences affecting methodology from either within the environment or external to the environment of a science discipline. During data analysis, we will look for cause-effect relationships between methodology and other categories of the Concept Map. Cause-effect relationships we will search for include: (a) Change in policy or science requirements — for example, the adoption of a standard format for biodiversity, such as ecological data for use in GIS systems; and (b) the adoption of technology (or technologies) that influences change in methodology — for example, the adoption of DNA barcoding for identification and classification of organisms that has dramatically changed the methodology of taxonomy.

- 3. Problems and Puzzles:*** As we described in Chapter 2, Problems and Puzzles are designed to support paradigms of a discipline. A community of scientists defines its Problems and Puzzles in a manner consistent with its paradigm (Kuhn, 1996).

The scientific goals of a community of scientists influence its Problems and Puzzles. Goals provide a stimulus for scientists to frame Problems and Puzzles in novel or bold ways to make discoveries. For example, governments reward scientists for proposing Problems and Puzzles that challenge theories and prevailing paradigms towards achieving discovery (NSF, 2012). How Problems and Puzzles are framed in relation to a paradigm, and in the context of the goals of a community of scientists, could result in discovery.

Change in the scientific goals of a community of scientists is potential stimuli entering the environment of a science discipline. Problems and Puzzles are subject to change as the goals of a community of science change (Glass, 2007).

**Observing change in Problems and Puzzles:** We plan to observe change in Problems and Puzzles by comparing the propositions and questions of a particular community of science over multiple periods of time. We will observe

for changes in the goals and in the scope articulated in the propositions and questions that frame the Problems and Puzzles.

Change in policy affecting a science discipline is an event we will observe as a potential cause of change in Problems and Puzzles. Change in policy affecting the Problems and Puzzles of a science discipline would be represented as a path of events linking the Governance category into a Problems and Puzzles category in the Aspects of a Science Discipline component.

4. **Resources:** As previously discussed, e-Infrastructure development receives stimuli of ICT investments, intended to develop the technological infrastructure. This infrastructure is represented by Physical, Technology and Instrumentation categories. As e-Infrastructure development interacts with a science discipline, we will refine our conceptual framework and use it to answer our primary research question.

In Chapter 2, we argued that a science discipline has resources, and we partitioned these resources into two classes: Human and Knowledge resources. We generalized that ***“Resources of a Science Discipline”*** is a broad concept, and that its use in our Concept Map can coincide with the other concepts; therefore, we must be precise in how we intend to use the concept, “Resources of a Science Discipline.”

***“Human resources of a science discipline”*** is a population of people who are practicing basic and applied research to achieve scientific progress and discovery. We found three activities that are used in the U.S. to develop human resources of a science discipline (NSB, 2010):

- Increasing the number of graduates with science degrees;
- Increasing immigration of foreign scientists; and
- Delaying retirement of senior scientists because of the relative youth of the science workforce compared to the total U.S. workforce.

**Observing change in human resources of a science discipline:** We will observe change in these activities, and remain open to discovery of other activities, as we study a particular community of science. The following is an initial list of variables we will use to observe change:

- Change in the number of graduate students;
- Change in the participation of foreign scientists; and
- Change in the employment status of senior scientists.

***“Knowledge resources of a science discipline”*** is a broad concept too, and it could be included in the other categories in the Aspects of a Science Discipline. For example, the experience that scientists within a community of scientists possess about how science is conducted is a knowledge resource (NAP, 2010). Another example of a “knowledge resource,” as characterized by Popper (1999), is an objective representation of Problems and Puzzles of a community of scientists, codified in a language as objective propositions or questions (Popper, 1999), where meaning can be transferred and reused to create knowledge.

We chose to treat knowledge resources of a science discipline as its own category in the Concept Map. In this category, we will track events from stimuli to discovery resulting from interactions between e-Infrastructure development and a science discipline.

**Observing change in knowledge resources of a science discipline:** Knowledge resources we plan to observe are: (a) experience of scientists, (b) change in Problems and Puzzles, and (c) multidisciplinary use of data sets. Note: This list is not exhaustive.

***Experience of scientists:*** Observe change in total number of peer-reviewed papers in a community of scientists. We intend to measure the experience of scientists in a particular community through the increase in peer-reviewed publications over a period of time.

***Change in Problems and Puzzles:*** Observe change in Problems and Puzzles. For example, change in scientific goals is an indication of potential change to types of Problems and Puzzles. For example, the scientific goal to address problems characterized in greater spatial and temporal scales is likely to change the research questions of a community of scientists. We argued that an objective form of the research questions is a knowledge resource.

***Data sharing across multiple communities of science:*** We would measure and compare papers that report findings by using data from across multiple communities and over multiple time periods. A particular community of science can be the creator of the data set(s). Reused data sets are potential knowledge resources to multiple disciplines. We would also observe if data sets were enhanced or combined with data from other disciplines. We refer to these types of data sets as heterogeneous data sets. It is our conjecture that e-Infrastructure provides a technological substrate to facilitate the creation of heterogeneous data sets. Moreover, we believe that an increase in heterogeneous data sets is an indicator of potential increase in discovery. The empirical component of this study will look for data that supports this proposition.

#### 4.2.2 Environment of a Science Discipline

An environment exerts pressures on a science discipline that can result in competition for resources between sub-disciplines or other organizations. For example, governments constrain science disciplines through policy and funding. In Chapter 3, we introduced concepts and properties from Resource Dependence Theory (RDT) and Population Ecology Theory (PET) to explain the relationships between organizations, resources and their environment.

An environment consists of organizations and resources. ***Organizations*** depend on resources for survival. Organizations control resources, and share them with other organizations. There are many types of environmental resources, including financial, physical, informational, and technological (Pfeffer and Salancik, 2003).

The environment of a science discipline consists of particular communities of science (domains) and resources, which we earlier classified as knowledge and human resources. Our conceptual framework treats physical, technological, instrumentation, and data as external resources in the component of e-Infrastructure development. A particular community of science we will study is the U.S. Long Term Ecological Research Network (LTER). LTER is a network of organizations, referred to as Sites. Each Site has resources that they control and share with other Sites.

**Observing organizations and resources:** We will use our Conceptual Framework and Concept Map to classify organizations and resources, and their interactions, as we explore events from stimulus to discovery. Organizations in the environment of a science discipline include *universities* (providers of human and knowledge resources, as well as physical and technology resources), *museums* (physical and instrumentation resources), and *government agencies* (e.g., funding agencies that provide funding resources and policies that exert controls on resources and goals). This is not an exhaustive list of organizations and resources in the environment of a science discipline. Classification of others is a component of the empirical work of this study.

Other organizations in the environment include *resource providers*. We described two types: domain-specific providers and domain-independent providers. *Domain-specific resource providers* are organizations with expertise and resources in domain specific knowledge (Foster, 2005). Expertise comes from human resources in the community of science workforce, and graduate student population (NAP, 2010). Resources in domain specific knowledge include domain-specific content in data, domain specific software, domain specific scientific processes, etc. For example, a domain-specific resource provider, using domain-specific knowledge and technology, processes raw remote sensing data into a ready-for-science data product. *Domain-independent resource providers* are organizations that possess expertise and resources that can be shared and reused across multiple disciplines;



for example, a scientific workflow system that can be adapted for reuse in different science disciplines.

***Niches:*** A niche was described earlier in Chapter 3 as a “resource space” in an environment that forms from combinations of actions on resources. Niches are important to detect, because they represent potential for discovery. For example, a niche could form from discovery of anomalous data, causing a paradigm and its discipline to transition from normal science to extraordinary science (Kuhn, 1996).

**Observing for niches:** When a niche forms, we should be able to observe its effects, such as the presence of a cluster of resources in the environment, or the potential for resources to come in to the environment. For example, a niche could form around a technology and a policy that facilitates the sharing of data. The niche grows in resources in order to increase adoption of this technology and policy when a funding stimulus is introduced into the environment.

***Linkages to external resources:*** As we described in Chapter 3, linkages to external resources are a function of organization dependence on the environment (Boyd, 1990). We adopt this idea to recognize events or effects that cause a science discipline to create linkages to external resources. It is important to categorize and describe newly discovered external linkages and to recognize patterns of e-Infrastructure development, as these will help answer our primary research question.

**Observing linkages to external resources:** Linkages to external resources will be observed through change in number of agreements or contracts a community of scientists has with other organizations. A contract or agreement between two organizations represents a linkage to access or control over one another’s resources. For example, an award number from a grants funding agency represents an agreement between the grantor (funding agency) and the grantee (organization who performs the work). This can also be observed in peer-reviewed publications; for example, an agreement between a provider of a data set and a user to access a

data set for an experiment. We can observe the presence of a linkage through a publication that cites the source of the data.

In summary, we have described the environment of a science discipline as consisting of organizations, resources, and niches. We have described how we plan to observe their interactions and the events that create conditions to establish linkages to external resources.

### 4.2.3 e-Infrastructure Development Process

Figure 2 in Chapter 2 represents our conceptualization of e-Infrastructure. The categories organized in a side-by-side fashion represent the development of e-Infrastructure as a continuum of hard infrastructure and soft infrastructure, socially constructed and adaptable enough to function in a dynamic environment. Hard infrastructure includes categories for Physical, Technology and Instrumentation. Soft Infrastructure includes categories for Process, Organization and Governance. Data, we hypothesized, plays a central role between Aspects of a Science Discipline and the process of e-Infrastructure development. We now describe each concept, and how we plan to observe it.

#### **1. *Physical:***

Physical is a category for physical objects or resources. An object in the physical category is not necessarily part of an e-Infrastructure, but it has the potential to be. For example, a museum that houses a collection of specimens is a physical object. It might be part of an e-Infrastructure if, for example, its collection is available online in a digital format that is accessible to a science discipline. We use the Physical category to categorize physical objects external to a science discipline's environment. These objects are being changed or created to be part of an e-Infrastructure.

An important reason to include this category in the Concept Map is to guide us in observing the level of past and present investment in changing physical resources to become a component of an e-Infrastructure. What it means for a physical resource,

such as a museum or a biological field station, to become a component of an e-Infrastructure is not well understood. Physical objects found to be components of an e-Infrastructure are observed in an effort to learn their role in an e-Infrastructure. In the empirical process, we will use our Concept Map to explore what investments have been made towards changing physical objects into components of an e-Infrastructure or to create and ready physical resources to be on an e-Infrastructure. It could be important to determine a correlation between changing a physical object into a component of an e-Infrastructure and a resulting discovery.

**Observing change in Physical objects into components of an e-Infrastructure:**

Physical objects changing into components of an e-Infrastructure will be observed by analyzing investments aimed at adapting physical facilities or resources into components of an e-Infrastructure. Our search will also include information on investments directed towards creating physical resources, which from the start are components of an e-Infrastructure. Investment information of this kind will be observed in public record reports or from informant interviews.

Also, we will observe for linkages between Aspects of a Science Discipline and the physical resources found within an e-Infrastructure. For example, a resource provider and a community of scientists have developed a workflow system to automatically harvest data across multiple field sites. Linkages (e.g., one or more agreements) to access physical resources between the sites and a resource provider are an indicator of a potential development of an e-Infrastructure.

***2. Technology:***

Technology as a category of e-Infrastructure development refers to the technologies that are fashioned into components of an e-Infrastructure. For example, a computer is an artifact fashioned from the technology of computation. We argued in Chapter 1 that nations are making huge investments in technology infrastructure development. This technology category is in our Concept Map to guide us in

observing artifacts in an e-Infrastructure or technologies in the process of being changed into components of an e-Infrastructure.

Our focus will be on technology investments by governments on basic and applied research. We are prepared to present investment information on these technology sub-categories: Computation, Storage, Networking, Software, and Visualization. While this list of categories is not exhaustive, our preliminary investigations indicate they cover a significant range of investments.

**Observing change in Technology Infrastructure:** We will search for information about change in technologies aimed at stimulating e-Infrastructure development. Of particular interest are investments in technology aimed at linking islands of physical and instrumentation resources into national shared e-Infrastructure aimed at enhancing scientific discovery.

### ***3. Instrumentation:***

Instrumentation refers to a measuring apparatus or system, such as a remote sensing system, microscope, telescope, DNA sequencer, etc., that collects or generates data for scientific use. Instruments offer precise measurements as data, which can then be applied to a measurement scale.

The Instrumentation category will be used to observe the level of investment in changing or creating instrumentation as a component of an e-Infrastructure. We recognize this as an outcome of a stimulus to develop e-Infrastructure.

In our Concept Map, we expect to find and represent events that are linkages from the environment of a science discipline to instrumentation. Case in point, consider an instrument at a field station measuring a range of variables characterizing the surrounding environment. The field station and the instrument are accessible through a workflow system within a national e-Infrastructure for biodiversity-ecology. Access to this instrument is only through the workflow system. Researchers must submit proposals requesting time to use the instrument. Receiving a grant to access the instrument is an event designed to establish a

linkage between a scientist and an instrument. Finding these linkages is important to guide us in characterizing the meaning of e-Infrastructure and how it develops.

**Observing change in Instrumentation:** We will observe change in Instrumentation by examining investments to adapt instrumentation to work in an e-Infrastructure. Reusing the previous example of a workflow system, an instrument, such as a data logger, may be adapted to transmit its data set whenever data volume reaches a threshold. We will observe for information on Investments to change existing instrumentation, as well as investments in new instrumentation to enable the instrument to work as a component of an e-Infrastructure, will be included in our inquiry.

#### ***4. Data***

Chapter 2 describes the Data category as having a key role in the development of e-Infrastructure. This category will be used for observation of investments for enhancing or creating data resources towards the development of an e-Infrastructure. A data resource refers to a digital representation of an object that is a source of data. As an example, an object that is a source of data might be a preserved butterfly that is in a collection at a museum. A taxonomist can study that butterfly at the museum, or request to borrow it from the collection. (This might not be possible depending on the relationship the taxonomist has with the museum). On the other hand, a data resource could also be a digital representation of that butterfly that also includes scientific information describing its properties. An example of a data resource similar to this is the Encyclopedia of Life<sup>8</sup>.

We would argue that a data resource is not a replacement for a real object that provides data through observation. On the other hand, data resources accessible through an e-Infrastructure create a potential to widen access to a significantly broader and more diverse community of scientists and other communities of interest, this widened access potentially increases the democratization of science.

---

<sup>8</sup> Encyclopedia of Life, <http://eol.org/>

Investments to stimulate development of data resources are the type of information the Data category would describe.

**Observing change in Data Infrastructure:** We will observe sources of information on investments to create data resources. An effect we will look for from these investments is how they are changing work practices in a science discipline. Changes in work practices of a community of scientists could be a finding in a developing e-Infrastructure. Moreover, we will observe from published studies or from subject matter experts what impact access to data resources has had on democratization. Publications showing co-authorships between scientists in the U.S. or other developed countries with authors in developing countries are an indicator of investments that we will use to observe investments to create data resources and data infrastructure.

## ***5. Process***

The Process category guides us in observing changes to processes as the e-Infrastructure develops. The Process category will guide our observation of the effects of linkages to resources between the categories in hard infrastructure, data infrastructure and Aspects of a Science Discipline. An effect we expect to find is the presence of conditions creating a demand to change a process or adopt a new one. An example of this is, investments in remote sensing technology, transforming data into geo-referenced data. Geo-referenced data increased demand for sharing data between communities of scientists, because it provided a common format that could be used with GIS technology. Scientists were reluctant to share data outside of their community. Eventually, the data providers adopted a guideline, which led to a change in the process of sharing their data.

**Observing change in Process Infrastructure:** Our approach to observing change in Process infrastructure is to observe effects of infrastructure development, including transfer, adaptation and path dependencies (see Table 2). The eventual result is a change in processes, or particular scientific practices. We will witness

changes in processes in publications, such as academic journals, internal documents of the community of science, or through informant interviews.

## **6. Organization**

The Organization category in our Concept Map refers to two classes of organizations interacting with Aspects of a Science Discipline. The first class is organizations that are changing; the second class is organizations that are emerging. Organizations that are changing refer to communities of a science discipline that are undergoing change or adaption resulting from environmental pressures. Organizations that are emerging refer to communities of a science discipline that are forming (coming into existence), as a result of conditions in the environment, such as availability of resources. Organization does not imply a legal entity. Organization more closely resembles a community or a “virtual organization.” An organization can also represent a project that consists of human and technology resources.

Let us first look at an example of an organization that is changing. Consider a synthesis center that produces a synthesized data product for its communities of science. What if this synthesis center is under severe pressure, because the government and its primary resource provider (funding), chose to terminate its contract (sever the linkage)? The synthesis center is changing as a result of the loss of the contract with the government. The synthesis center must establish new linkages to other communities of science (organizations) that will provide resources in exchange for its data synthesis product. Otherwise, it faces dramatic changes, or eventually ceases to exist due to the loss of its resources.

A second example is an organization that is emerging. Consider a community of scientists that have made a discovery. As Graham et al (2002) found when discoveries occurred in the discipline of ecology, communities of scientists would splinter from their previous paradigm to develop the emerging paradigm. Complementary to the new paradigm, a new organization emerges, with linkages to internal and external environmental resources designed to support the emerging paradigm.

**Observing change in Organization Infrastructure:** An organization can be observed through its physical, technological and human resources. Variation in the goals of an organization is an approach that we will take to observe change. Organizations that will be observed have research goals, and are normally funded by a government. Changes in organizations' funding levels, composition of their human resources (i.e., repeated declines in the number of scientists employed in the organization), and linkages to resources in other organizations are the avenues we will use to observe changes in organizations.

## **7. Governance**

The Governance category guides us in observing changes in the ways in which organizations are governed. Governance in the Concept Map is where we will categorize properties of governance, such as structure of an organization, and rules by which it operates and works towards achieving its goals. Governance of an organization is not static, meaning it can change as the organization evolves. For example, governance of the LTER Network evolved to represent 26 sites, enabling the organization to work as a coordinated network (Gosz et al, 2010).

**Observing change in Governance:** Our approach to observing change in governance of an organization is to compare its governance structure over time. For example, structural and functional changes in an organization over time indicate changes in its governance structure made mostly likely to address environmental pressures.

In summary, we have described the seven categories in the e-Infrastructure development process included in our Concept Map. For each of the seven categories, the following was provided: (1) a description of each category's function within the e-Infrastructure development process and its interactions with other components, and (2) a description of how it will be used to observe changes over time in the properties of that category. Examples of realistic scenarios were given to illustrate the use of each category. We now continue with the next component of the Concept Map.



#### 4.2.4 Co-evolution Relationship of a Science Discipline and e-Infrastructure Development

The Co-evolution Relationship (fourth) component of our Concept Map provides a lens by which we can observe change between Aspects of a Science Discipline and e-Infrastructure development. This component is equipped with concepts of co-evolution, natural selection (Thompson, 1994; Pazos et al, 2008), and organization behavior in relation to its environment (Aldrich, 1979; Pfeffer and Salancik, 1978). Communities of scientists require resources as they work on Problems and Puzzles. When resources are needed external to a community, one or more linkages (Boyd, 1990) are created between organizations to provide access to them.

The Co-evolution Relationship component guides us in organizing information on linkages and resources between a science discipline and e-Infrastructure development. This component observes paths between e-Infrastructure development and Aspects of a Science Discipline. A path refers to a linkage between a community in a science discipline and an e-Infrastructure resource provider (organization). A path supports access to and exchange of one or more resources and is, therefore, a substrate between a community of scientists and a resource provider.

Referring to the Resource Dependence Theory, a linkage in the context of this conceptual framework can be conceptualized as a process that creates a mechanism by which one or more paths can be established between an organization (in a science discipline) and a resource (or resource provider). An example of a linkage is the process of establishing an agreement. The agreement itself is the mechanism by which one or more paths can be established.

A path supports access to and exchange of one or more resources, and therefore, works as a substrate between a community of scientists and a resource provider.

Case in point; consider a network of field sites for biodiversity and ecology research that produce data sets for long-term ecological studies, such as the Long Term Ecological Research (LTER) Network. Prior to data sharing occurring between sites

or other research networks, linkages were established that defined a policy for data sharing. The linkages took on the form of agreements, normally between a data provider and a data user. Once the linkage (agreement) was established, mechanisms were created, such as access credentials, necessary to access systems or databases, containing desired data sets.

**Observing change in the Co-evolution Relationship component:** We will track the number of paths created over time between Aspects of a Science Discipline component and the e-Infrastructure development process component. Comparisons can be performed at two different time intervals to observe change in the number of paths representing linkages supporting the exchange of resources. For example, in the first decade (1980 – 1989) of the LTER Network, paths between sites that supported cross-site comparative research were fewer than five (there were none in the first four years). Sharing of resources, such as data and information, was very low, as reflected in the number of publications with co-authorships (Johnson et al, 2010). During the second and third decades, as stimulus came to increase cross-site comparative research, the number of paths between sites increased dramatically.

#### **4.2.5 Scientific Discovery Component**

Previously, in component (4) of the Concept Map, we described the co-evolution relationship between a science discipline and e-Infrastructure development. We also introduced the concept of a path as a linkage between a community in a science discipline and an e-Infrastructure resource provider.

The Scientific Discovery component (component 5 of the Concept Map) guides us in observing for discovery events. A “discovery event” marks the occurrence of a discovery. As described in Chapter 2, a “discovery” commences with the awareness of an anomaly (Kuhn, 1996). When a discovery event occurs, the Scientific Discovery component marks the path between the Co-evolution Relationship component (5) and Scientific Discovery component (6). The path from stimulus to discovery can be traced. The path can also be studied to retrospectively observe the e-Infrastructure development process. From this, we can empirically determine the

impact that stimulus and the e-Infrastructure development had on discovery, when combined with interactions from a science discipline..

**Observing change in the Scientific Discovery component:** The Scientific Discovery component changes when a discovery event occurs. Information on discovery events is observable in peer-reviewed publications. Discoveries are published and obtained by interviewing scientists who were members of a particular science community and who possess historical information about discovery events.

#### 4.2.6 ICT Investments Stimulus Component

The ICT Investments Stimulus component (6) in the Concept Map guides us in observing stimuli of ICT investments intended for either the development of e-Infrastructure or the environment of a science discipline.

We hypothesized that the majority of the stimuli in ICT investments would be for the development of technological infrastructure, which eventually would dramatically increase scientific discovery. The conceptual framework also supports an investment pull from science communities into Aspects-of-a-Science-Discipline components. For example, an external stimulus of funding for science research, obtained through a proposal process, instead of through a federal technology e-Infrastructure initiative. This component will guide us in testing this hypothesis as part of our empirical work. When a discovery event is observed, this component will provide us with information to retrospectively trace back to the stimulus that led to a discovery. This will help us answer our primary research question.

**Observing change in the ICT Investments Stimulus component:** The ICT Investments Stimulus component changes when a stimulus of ICT investments is made. The ICT investments we observe will be made primarily by the U.S. government to federal agencies investing in ICT investments. These investments are intended for the creation and support of a shared national e-Infrastructure, supporting basic Research and Development (R&D), and Applied Research. Changes will be observed by comparing investments on an annual basis across different

investment categories. We will refer to online publications with federal ICT investment information for this data.

### **4.3 Summary of the Concept Map**

We have described each of the concepts in the Concept Map. For each concept, we provided an operational description as well as explained how we plan to observe the concept. Examples of realistic scenarios were given to illustrate how we would use the Concept Map for exploration and testing our conceptual framework.

We now display the components and categories of the Concept Map in Table 4. Included is a brief description of each component and concept, its operational function, and how it will be observed.

Concept	Definition/Description	How the Concept is Observed
A Particular Community of Scientists	<p><b>Aspects of a Science Discipline</b></p> <p>A particular community of scientists is a body of scientists that practices research defined by its Problems and Puzzles, and is grounded upon its achievements (knowledge).</p>	<ul style="list-style-type: none"> <li>• <i>Size</i>: number of scientists; number of students; number of organizations participating in the community.</li> <li>• <i>Membership</i>: measure of different forms of participation; e.g., funded, affiliated, user, etc.</li> <li>• <i>Structure</i>: measures geographical dispersion and the degree of sharing of resources.</li> </ul>
Methodology	How science is practiced in a particular community of scientists, guided by its prevailing paradigm.	Changes in how the science is practiced as a result of adoption of a standard or a policy. For example, the adoption of a data-driven approach, such as data mining versus hypothesis driven approach.
Problems and Puzzles	Problems and Puzzles define the propositions and questions used to further articulate the paradigm from which they were derived. An anomaly is a particular kind of problem that appears as an inconsistency in relation to a discipline's accumulated knowledge.	<ul style="list-style-type: none"> <li>• Compare propositions and questions of a particular community of science over multiple periods of time.</li> <li>• Identify any change in the goals and scope that frame the Problems and Puzzles.</li> </ul>
Resources of a Science Discipline:	A science discipline has human and knowledge resources. These resources are within the internal environment of a science discipline.	
<ul style="list-style-type: none"> <li>• Human Resources of a Science Discipline</li> </ul>	A population of people who are practicing basic and applied research to	Changes in size and composition of the scientific workforce:

	achieve scientific progress and discovery	<ul style="list-style-type: none"> <li>• Change in the number of graduate students.</li> <li>• Change in the participation of foreign scientists.</li> <li>• Change in the employment status of senior scientists</li> </ul>
<ul style="list-style-type: none"> <li>• Knowledge Resources of a Science Discipline</li> </ul>	An objective representation of Problems and Puzzles of a community of scientists, codified in a language as objective propositions or questions (Popper, 1999), where meaning can be transferred and reused to create knowledge.	<ul style="list-style-type: none"> <li>• <i>Experience of scientists</i>: Change in number of peer-reviewed papers.</li> <li>• <i>Change in Problems and Puzzles</i>: Observe change in Problems and Puzzles.</li> <li>• <i>Data sharing from across multiple communities of science</i>: Compare papers that report findings using data from across multiple communities, over multiple time periods.</li> </ul>
<b><i>Environment of a Science Discipline:</i></b>		
Environment	An environment exerts pressures on a science discipline that can result in competition for resources between sub-disciplines or other organizations.	
Organizations (Communities of Scientists):	Organizations control resources, and share them with other organizations.	<ul style="list-style-type: none"> <li>• Change in types of organizations in the environment; e.g., academic, government, non-governmental.</li> <li>• Change in broader participation of organization; e.g., from other nations.</li> </ul>
Resources in Environment of Science Discipline	Organizations with resources include <b><i>universities</i></b> (providers of human and	Change in number and types of resources in the environment of a

	knowledge resources, as well as physical and technology resources), <i>museums</i> (physical and instrumentation resources), and <i>government agencies</i> (e.g., funding agencies that provide funding resources and policies that exert controls on resources and goals).	science discipline. Compare resources over time. Observe information on resources in reports and information obtained from informant interviews.
Linkages to external resources	A linkage represents an organization's dependence on external resources	Change in number of agreements or contracts over time.
<b><i>e-Infrastructure Development Process:</i></b>		
Physical	Physical is a category for physical objects or resources. Observes physical objects that are being changed or created as part of an e-Infrastructure development process.	Change in the investments to modify or create physical resources
Technology	Refers to the technologies that are fashioned into components of an e-Infrastructure to enhance basic and applied research.	Change in technologies aimed at stimulating e-Infrastructure development; e.g., using technology to link islands of resources into an e-Infrastructure
Instrumentation	A measuring apparatus or system, such as a remote sensing system, a microscope, a telescope, DNA sequencer, etc., that collects or generates data for scientific use.	Change to instrumentation for it to work as a component of an e-Infrastructure; e.g., an instrument at a field station changed to work on an e-Infrastructure.
Data	Data refers to the socio-technological resources that are orchestrated towards the generation, collection, archiving, and transmission of data.	<ul style="list-style-type: none"> <li>• Stimulus for change to number and type of data resources.</li> <li>• Effects of data infrastructure; e.g., increases in sharing and democratization.</li> </ul>

Process	Supports coordination and exchange of resources between Aspects of a Science Discipline, data, and hard infrastructure categories of e-Infrastructure.	<ul style="list-style-type: none"> <li>Observe effects of infrastructure development, such as transfer, adaptation and path dependencies.</li> </ul>
Organization	Refers to two classes of organization interacting with Aspects of a Science Discipline: organizations that are changing and organizations that are emerging.	Observe change in the environments (internal or external to a science discipline): Change in their funding levels and change in composition of their human resources, such as repeated declines in the number of scientists employed in the organization.
Governance	Governance includes the structure of an organization, and rules by which it operates and works towards achieving its goals.	Compare governance structure for changes over time.; e.g., structural and functional changes in an organization over time indicate changes in its governance structure mostly likely to address environmental pressures.
<b>Co-evolution Relationship of a Science Discipline and e-Infrastructure Development Component (4)</b>		
Path between a science discipline and an e-Infrastructure	A Path refers to a linkage between a community in a science discipline and an e-Infrastructure resource provider (organization), which supports access and exchange of one or more resources.	Change in number of paths created over time between Aspects of a Science Discipline component and the e-Infrastructure development process component.
Linkages	Linkages are created between organizations to provide access to resources.	Compare agreements between organizations over time.
<b>Scientific Discovery Component (5)</b>		
Discovery Events	A “discovery event” marks the occurrence of a discovery. A “discovery” commences with the awareness of an	Compare discovery events at different time intervals.



	anomaly. The Scientific Discovery component marks the path between components (5) and (6).	
<b>ICT Investments Stimulus Component (6)</b>		
Stimuli of ICT investments	Observe stimuli of ICT investments intended for either the development of e-Infrastructure or the environment of a science discipline.	Compare U.S. government investments on an annual basis across different investment categories.

**Table 4 Key Concepts, their Definitions and Observations**

## CHAPTER V: Research Design

### 5. Research Design

The purpose of a research design is to help the investigator understand the actual structure of the study, and to plan the study and execute it. Maxwell (2005) describes research design in a qualitative study as “an ongoing process that involves tacking back and forth between the different components of the study, assessing the implications of goals, theories, research questions, methods, and the validity threats to one another (Maxwell, 2005).”

The goal of this study is to construct theory about the process of e-Infrastructure development and its impact on scientific progress. As we described in Chapter 4, our theory building process starts with an initial conceptual lens. This lens represents our ideas about ICT investment stimulus, e-Infrastructure development, the Aspects of a Science Discipline and its impact on scientific discovery. The research design represents the components of our theory building process and the components of the empirical research process to test and validate the theoretical framework.

For this study, we adopt the “Interactive Model of Research Design” by Maxwell (2005). This model represents a research design encompassing the following five components: Goals, Conceptual Framework, Research Questions, Methods and Validity. Maxwell (2005) describes the model as providing “a structure that is interconnected and flexible,” whereby it supports interaction between each of the components. The model represents the components of the theory development process, the empirical research process, and their interconnections.

#### **Components of the Interactive Model of Research Design:**

In this section, we shall describe the components of Maxwell’s interactive model of research design (Maxwell, 2005). With an understanding of the function of a

component, we will describe each component of the model for our research design. Finally, we will show a graphical representation of the model developed for our research design.

***Goals:***

The Goals component of the research design model details the purpose for conducting the study. Goals can help guide other design decisions to ensure that the study is worth doing, and that something of value will result (Maxwell, 2005). As a guide, goals can be thought of as a leader that pulls the investigator along while being guided in making decisions to reach the desired destination of the study. Goals can serve to justify the reasons for doing a study, and can be personal, practical, or intellectual (scholarly) (Maxwell, 2005).

The goals of this study are to increase understanding of the process of e-Infrastructure development and how it impacts scientific discovery. From the perspective of a science discipline, we also want to increase our understanding of how the requirements of a science discipline shape the development of e-Infrastructure.

***Conceptual Framework:***

The Conceptual Framework component frames what the researcher's ideas or thoughts are about what is going on with the issues, settings, or people the researcher plans to study. The literature, theories, beliefs, and prior research findings the researcher will draw upon to guide or inform the research study.. Chapter 4 described the tentative theory and conceptual framework of this study. We drew upon literature and theories in order to develop this conceptual framework and Concept Map. The map interconnects the concepts of e-Infrastructure development and the Aspects of a Science Discipline.

***Research Questions:***

At the core of a research design are the research questions. The research questions state what it is the researcher who wants to understand by doing the study. The research questions direct us towards gaining a greater understanding of the process

of e-Infrastructure development, its mutual shaping relationship to a science discipline, and how the interaction between infrastructure development and science discipline can lead to dramatic increases in scientific progress.

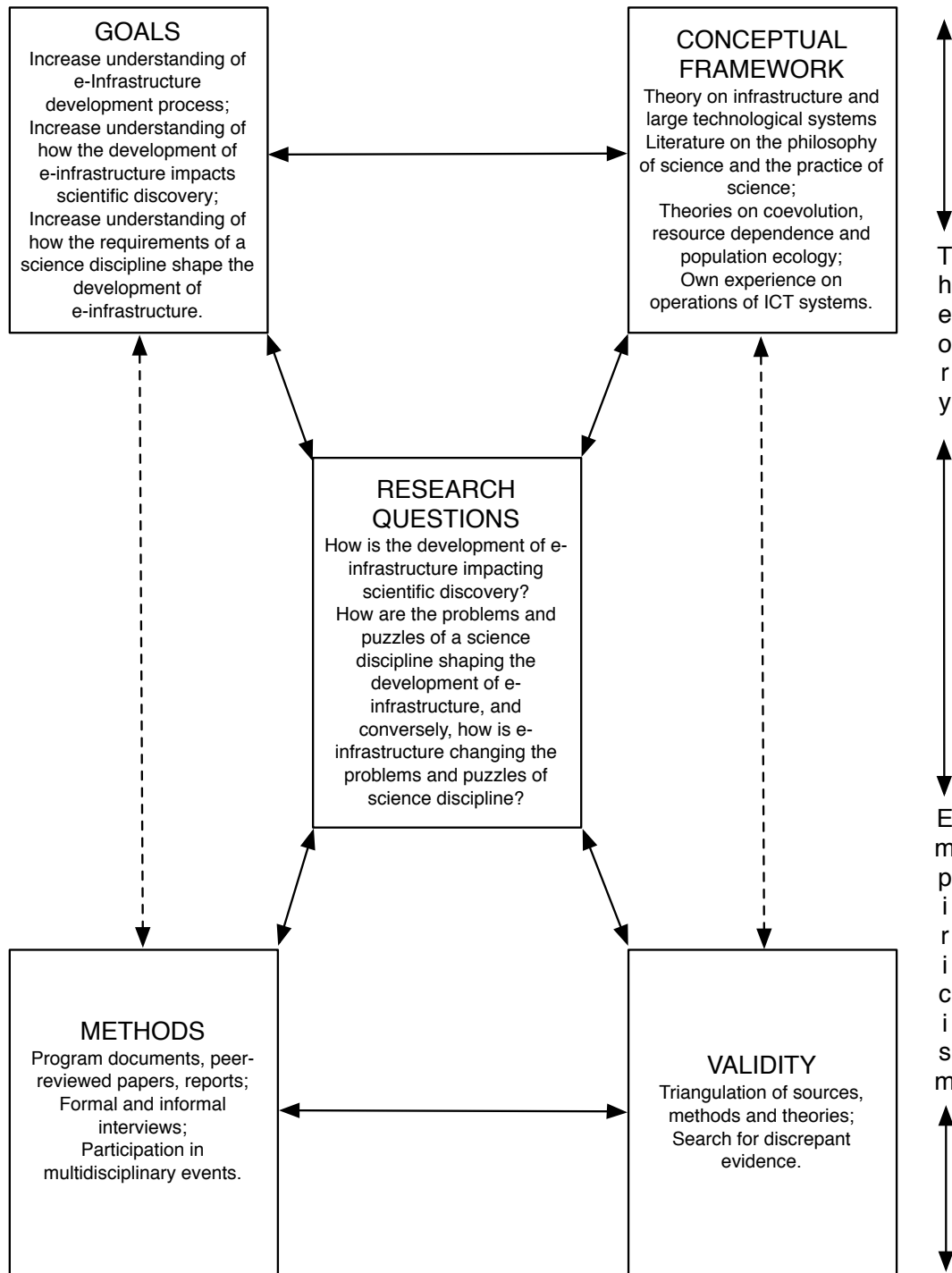
***Methods:***

The Methods component describes the approaches and techniques the researcher plans to use to collect and analyze data. Since the process of e-Infrastructure development and the initiatives of Cyberinfrastructure and e-Science are emerging, their impact is not well understood. As a result, the approach chosen for this study is exploratory. The plans for data collection and data analysis are detailed in the next chapter on Research Methodology.

***Validity:***

The Validity component describes the strategies the researcher uses to identify and rule out potential threats to the results or conclusions of the study. Our approach to testing for Validity threats is through triangulation and searching for discrepant evidence (Maxwell, 2005).

Figure 6 illustrates the research design of this study, showing how the study's major components are connected, based upon the Interactive Research Design Model by Maxwell (2005).



**Figure 6 Major components of this study based upon the Interactive Research Design Model**

We have described the research design of our study based upon the Interactive Model of Research Design by Maxwell (2005). We used this model to describe the theory building and empirical research components of our study.

## CHAPTER VI: Research Methodology

### 6. Research Methodology

The aim of this research is to acquire understanding of the development of e-Infrastructure as a stimulus to increase scientific discovery. Our perspective is that e-Infrastructure development is a form of stimulus towards increasing scientific discovery by creating a continuum of technology and socio-organizational resources that support transformative research.

A qualitative study, using an interpretive grounded theory research approach, will guide the research methodology for data collection and analysis. Findings will be grounded in the data (Strauss and Corbin, 1998) and used iteratively to develop the conceptual framework (Maxwell, 2005).

Our perspective, as represented in the initial Concept Map, is that the process of e-Infrastructure development is a form of stimulus for science disciplines. The purpose of this stimulus is to induce change in the environment by creating a phenomenon that will hopefully increase revolutionary scientific progress and dramatically increase discovery. This research utilizes a strategy of a biological sciences case study with two embedded sub-cases: biodiversity and genomics. The aim is to generate a descriptive and explanatory theory for understanding how the stimulus of the process of e-Infrastructure development impacts scientific progress.

#### 6.1 Research Stance: Interpretivism

An interpretive approach is used to guide research into generating meaning about the impact that the e-Infrastructure development process is having on Aspects of a Science Discipline and scientific discovery. For instance, an interpretive approach will guide the research process towards making sense of issues involving Aspects of a Science Discipline, such as a goal to increase comparative research, and the impact from e-Infrastructure providing resources in response to the requirements of that goal.

The objective of this study is to construct theory from the findings of rich case studies of two life sciences disciplines: biodiversity and genomics. Using a qualitative research methodology, these cases studies will guide the research process to focus and highlight the unique characteristics of each discipline: its community of scientists, problems and puzzles, methodology, and resources. A qualitative case study research methodology is suggested in order to express a holistic treatment of the disciplines and e-Infrastructure development process. The purpose of this is to explore interactions in multiple contexts: temporal and spatial, historical, political, cultural, etc. (Stake, 1995).

## **6.2 Connecting with a Research Paradigm: Interpretive Research**

Myers (1997) informs us that the underlying assumptions of the Interpretive Research paradigm are: (1) Knowledge of reality is gained through social constructions, such as language, consciousness, shared meanings, documents, tools, and other artifacts; and (2) the understanding of phenomena comes from the meanings that people assign to them. These meanings are derived from analyzing written and/or oral texts by the participants in the phenomenon being researched, or text-analogues. Myers (1997) explains a text-analogue as “an object, which the researcher comes to understand through oral or written text.” Text, in the form of reports, studies, proceedings, program solicitations, etc., and oral interviews of informants, will provide the data sources of this study. These text and oral sources will derive from different communities with differing perspectives on the role of e-Infrastructure development and its effects on science disciplines. Interpretive research methodology guides the investigator in making connections across different sources in order to make sense of what is going on (Myers, 1997).

## **6.3 Case Study Research Methodology**

As described earlier, the objective of this study is to gain understanding of the development of e-Infrastructure as a stimulus to increase scientific discovery. We will do this through constructing theory from the findings derived from the observations of the interactions between the Aspects of a Science Discipline and the

process of e-Infrastructure development. This section describes conditions under which a case study research methodology is suggested.

*Understanding complex social phenomena:* A case study research methodology is recommended when there is a desire to understand complex social phenomena (Yin, 2003). E-Infrastructure development embodies complex social phenomena, previously characterized as a form of stimulus towards increasing scientific discovery by creating a continuum of technology and socio-organizational resources that support transformative research.

*Observing phenomena that are contemporary:* An e-Infrastructure development process as a stimulus is a contemporary phenomenon, where its events and conditions created through interactions in the environment of a science discipline are not under the control of the investigator (Yin, 2003). Observations of these events are gathered through interviews of scientists and representatives of resource providers interacting in the phenomenon. This investigator, guided by using a case study methodology, can then develop meaningful descriptions of the events, and then connect them to the concepts in the conceptual framework.

*Contextualizing the phenomenon:* The case study is regarded as an appropriate research strategy when the phenomenon under study is not readily distinguishable from its context (Yin, 2003). The context of this study is a small number of science disciplines that are studied through inquiry of issues, happenings or events, involving the interactions of e-Infrastructure development with Aspects of a Science Discipline. On the level of a science discipline, context is driven by pressures on a science discipline to achieve revolutionary scientific progress. In this context, the phenomenon is studied from the perspective of a science discipline acquiring the resources it needs, either from within or external to the discipline. On the level of e-Infrastructure development, investments in e-Infrastructure development to stimulate revolutionary scientific progress drive the interactions with a science discipline. In this context, the phenomenon is studied from the perspective of resource providers (e-Infrastructure developers, stakeholders) stimulating the



environment with resources to enhance the capabilities of a science discipline to achieve increases in discovery.

The sections that follow describe the other components of the case study research methodology. These components are the (a) Unit of Analysis, (b) Data Gathering, (c) Data Analysis, and (d) Interpretation of Observations.

#### **6.4 Unit of Analysis**

The unit of analysis defines what the “case” is and establishes its boundaries (Yin, 2003). A case refers to a “contemporary phenomenon within a real-life context (Yin, 2003; Miles and Huberman, 1994),” that can be defined as a “bounded system” with working parts (Stake, 1995). “The case is, in effect, the unit of analysis (Miles and Huberman, 1994).” The definition of the research questions plays a role in guiding how to define the boundaries of the unit of analysis (Yin, 2003). As a result, defining the unit of analysis is a key component of the case study research design (Yin, 2003).

The unit of analysis is the process of e-Infrastructure development aimed at effecting revolutionary scientific progress. The process starts with a stimulus (investments in technology) aimed at providing scientists with new approaches to solve Problems and Puzzles.

Reasons for starting off with a broad unit of analysis, such as the process of e-Infrastructure development, are the following: First, to consider the varied forms of interactions between a science discipline and e-infrastructure development. As described in section 4.1, Aspects of a Science Discipline and e-Infrastructure coevolve as they exert selective pressure (change) on the other. On the level of a science discipline, the environment of a science discipline exerts pressures that bring about change: Lack of funding impacts resources that are necessary to conduct research; e.g., the creation of a new instrument or the operation of an existing instrument. On the level of e-Infrastructure, pressure is exerted on its components (or artifacts). For example, components of an e-infrastructure are exposed to pressures during transfer to a new environment. These pressures can be placed

upon components from technological, social, political, legal, economic, and other variations.

## 6.5 Data Gathering

Data gathering involves deciding what data to collect, the source of the data, and how to collect it (Walsham, 1995). Inductive interpretive case study research produces data that is part of an iterative process of data gathering and analysis (Walsham, 2006). This data are an informant's interpretations (Walsham, 1995) from sampling of people, settings, events and processes (Maxwell, 2005). The outcome of this iterative process between data gathering and interpretation can then lead to theories being expanded, revised or abandoned (Walsham, 1995).

Our decisions on data gathering are guided by three theories and methodology: (1) Walsham (2006) on carrying out fieldwork when doing interpretive research. (2) Yin (2003) for his guidelines on sources of evidence from data collected in case studies, and principles of data collection, and (3) Eisenhardt (1989) who guides our decisions on data gathering, such that the outcome of the research is theory.

The next section describes our data gathering approach.

### 6.5.1 Methods for Data Collection

Methods for data gathering are organized along the following principle activities suggested by Walsham (2006) and Yin (2003): (1) choosing a style of involvement, (2) gaining and maintaining access, (3) collecting field data, (4) interviewing methodology, (5) direct observation (6) publications.

*Choosing a style of involvement:* Walsham (2006) views involvement as a "spectrum" that changes over time. With a spectrum in mind, he characterized involvement as having a "neutral observer" at one end of the spectrum, and at the other end of the spectrum is a "full action researcher." "Neutral" refers to people in the field situation who do not perceive the investigator as being aligned with a particular individual or group within the organization, or as being concerned with making money as a consultant, or as having strong prior views of specific people, systems or

processes based on previous work in the organization. Walsham (2006) refers to a “full action researcher” as someone with close involvement who consciously and explicitly works at changing things in the way that he/she feels is best.

The spectrum is a good analogy for characterizing this study’s different styles of involvement for data collection. On the close involvement side of the spectrum, the investigator of this study has participated in fieldwork and engaged faculty and students through workshops or meetings structured for the purpose of inquiry about e-Infrastructure and its role in their discipline. Towards the neutral zone of the spectrum, the investigator has been gathering and synthesizing data from interviews, reports and scholarly literature of the science discipline, in a more traditional interpretive style of inquiry (Myers, 1997).

*Gaining and maintaining access:* Walsham (2006) explains that it is critical in an interpretive research study for an investigator to gain and maintain good access to appropriate organizations and informants for the investigator’s fieldwork. An advantage in conducting an interpretive study is that this investigator works in an academic setting and is able to gain and maintain access to local as well as geographically diverse researchers from a variety of science disciplines.

*Collecting field data:* Interviews are recognized as the predominant data collection methodology for creating the interpretations of informants in the field (Walsham, 2006; Myers, 1997). Inquiry normally takes the form of guided conversation (Yin, 2003), in which the stream of questions is fluid (Rubin and Rubin, 1995). Questions asked that are unbiased and in an open-ended manner help guide informants to respond with facts and their opinions about events (Yin, 2003). It is through interviews that the investigator “can best access the interpretations that informants have regarding the actions and events which have or are taking place, and the views and aspirations of themselves and other informants (Walsham, 1995).”

*Interviewing Methodology:* The interviewing methodology for this research will be based on a qualitative interviewing approach. Qualitative interviews involve an exchange of information, in the form of an interactive dialogue between an

investigator and an informant. This normally takes the form of an informal discussion (Rubin and Rubin, 1995). The type of qualitative interviewing we will use to gather data for this research is the semi-structured format, containing open-ended questions. A semi-structured interview is a thematic, topic-centered interview, with a set of open-ended questions employed when an investigator wants more specific information (Rubin and Rubin, 1995; Robson, 2002). The semi-structured format with open-ended questions helps elicit answers about communities of scientists in science disciplines, Problems and Puzzles in those communities, and contemporary events and pressures currently being faced.

*Direct Observation:* Direct observation is another form of fieldwork, where data is collected from a site where aspects of the phenomenon can be observed (Yin, 2003). In case study research, observational evidence is considered useful in supplementing information about the case (Yin, 2003). We include direct observation as one of our data collection methods because the investigator made site visits to observe the deployment of e-Infrastructure and interview scientists in the field working with e-Infrastructure technologies.

*Publications on the domain context of the informants:* In an interpretive study, interviews should be supplemented with other forms of field data, such as publications on the domain context of the informants, be they organizations or people (Walsham, 2006; Myers, 1997). Walsham (2006) makes a distinction between publications and what he calls “internal documents.” Internal documents include strategies, plans, evaluations, proposals, etc. The case studies for this empirical research rely on publications, internal documents, and online sources of information to gather and organize data over periods of evolution involving a science discipline and technology. Synthesis based upon interview data and document data will support our aim to achieve generalizability (Walsham, 2006).

In this section, we have described our approach to performing data collection for an interpretive research study based upon Walsham (2006) and Yin (2003). We described sources and methods for data collection as a subset of the activities

Walsham (2006) describes within the overarching activity he calls, “Carrying out fieldwork.”

## 6.5.2 Sources for Data Collection

In this section, the sources of data for the data collection process are described. The primary sources of data are the Informant and Documents.

### 6.5.2.1 *Informants*

Informants are subject matter experts in a particular community of science, a representative in an organization involved with the process of e-Infrastructure development (e.g., a resource provider), or a representative from an organization that provided a stimulus into the environment of a science discipline (e.g., representative of a government funding agency).

Informants most likely will be senior scientists who have been practicing science in their community for the past 20-plus years. In addition to the importance of the length of time practicing is the kind of work they have been doing — work that has allowed them to experience changes brought about by technology. They are knowledgeable about the evolution of their discipline from the aspects of both the science and the technology. They have the background to understand the objective of this research, meaning that they should have a conceptual understanding of e-Infrastructure and Cyberinfrastructure programs. They can provide concrete examples that support their conceptual understanding (Rubin and Rubin, 1995).

### 6.5.2.2 *Documents as Sources of Data*

**Criteria for Documents:** Documents of interest should provide sources of information with its scientific work on and approach toward solving its Problems and Puzzles, and building upon its achievements. The historiography<sup>9</sup> of science classifies different types of documents for sources of scientific information. Kragh (1987) categorizes these sources as scientific work viewed as creative, intellectual activity. “A source is an objectively given, material item from the past, created by

---

<sup>9</sup> The historiography of science is the study of the history and methodology of the history of science, as a sub-discipline of history (Kragh, 1987).

human beings; as in a letter or a clay pot (Kragh, 1987).” The material items, or artifacts, we are interested in are documents. For documents to function as source material, they “must be capable of being utilized to give some of the information they contain in a latent form (Kragh, 1987).” Documents can help corroborate and augment evidence from other sources, such as information from which inferences can be made to find clues worthy of further investigation (Yin, 2003). On the other hand, documents can also offer contradictory information, which could provide insights into further inquiry (Yin, 2003).

Kragh (1987) defines two types of sources: symbolic sources and non-symbolic sources. Symbolic sources contain written, textual information. Symbolic sources can also consist of images, such as maps, films, illustrations, photographs, television programs, movies, etc. Non-symbolic sources can consist of physical artifacts, such as buildings and laboratories; instruments, machines and apparatus; chemicals, herbaria, natural history collections. Our primary sources will be of the symbolic type, primarily textual documents. Herein we shall use the term symbolic document to refer to documents that consist of symbolic sources of information (as described above). We further classify symbolic documents that are of particular interest as sources. Types of symbolic documents include: Strategic, Programmatic, Peer-reviewed and Online.

**Strategic Documents:** These documents normally provide information that articulates the vision or strategic goals of a scientific discipline. These types of documents focus on challenges the discipline faces, and describe short-term and long-term solutions. These documents could be in the form of reports, bulletins or policies (Lober, 1997) from scientific institutions. For example, the U.S. National Science Board publishes memorandums and reports that announce actions or policy recommendations that authorize the NSF to fund programs that can have short-term or long-term impact on programmatic activities that effect science disciplines.

Strategic documents also contain information that introduces change into a community or discipline. These documents normally focus on long-term issues,

such as decadal surveys of a discipline describing priorities over a 10-year period, based on collective input from its community. They could also be visionary documents published by a sponsorship/funding agency, such as the NSF. For example, the NSF published the Cyberinfrastructure Vision for 21st Century Discovery (NSF, 2007). Strategic documents can also address short-term issues that have long-term implications. For example, the collection of workshop reports on the Biodiversity Observation Network (BON) addressed the short-term issues of organizing a community effort and a peer-review process. The process helped secure funding for a program that would result in longer-term strategic goals for the biodiversity discipline.

**Programmatic documents:** Collectively, these types of documents describe the social and institutional environment of a science discipline (Kragh, 1987). Often called program solicitations, programmatic documents solicit proposals from the community of interest to address Problems and Puzzles with solutions that apply methods of particular interest to the sponsor. Program solicitations state requirements the proposer must adhere to. Proposals submitted to program solicitations are peer reviewed to ensure compliance with solicitation requirements. These types of documents are important, because they can aggregate interventions into a discipline through significant amounts of funding.. This has the potential for long-term consequences (Woolgar et al, 2006).

**Peer-reviewed documents:** Peer-reviewed documents (also called academic papers) are normally found in journals and publications produced by the members of an academic discipline. Alongside subject-matter-expert information, peer-reviewed documents will be an important source for information on a discipline's body of knowledge: its problems, puzzles, and solutions. Conference proceedings are another form of peer-reviewed document that can provide information about what domain scientists are doing in their discipline of intellectual and scholarly work (Cooper, 1998). For example, the proceedings from the International year of Astronomy 2009 documents intellectual and scholarly activity of the international

astronomy community, as well as goals of fostering education and capacity building throughout the world.

**Informational documents:** Informational documents refer to sources used to record information concerning events, activities, discussions, decisions, outcomes, etc. Examples of informational documents can be committee reports, minutes of meetings, newsletters, online information, and documents about a project. Informational documents may also convey information with a specific intention, such as stimulating discussion in a particular scientific community on an opinion or topic that is contentious.

## 6.6 Data Analysis

Data analysis concerns itself with how the collected data is analyzed. This includes describing the strategy and the methodology used for data analysis.

For data analysis in an interpretive inquiry, we adopt the strategy suggested by Eisenhardt (1989) and extended by Walsham (1995) of approaching data analysis as an interplay between the field data and the conceptual framework. As a methodology for building theory from interpretive case study research, the data gathering process should be flexible enough to allow for the overlapping of data gathering and data analysis processes, and allow for the investigator to make adjustments (Eisenhardt, 1989; Walsham, 1995). Eisenhardt (1989) recognizes this flexibility as a key feature of theory building in case study research. For example, adjustments can involve adapting cases in response to particular themes that emerge (Eisenhardt, 1989).

Theory provides a way to view data (Walsham, 1995). The conceptual lens developed in Chapter 4 provides an initial theory, encapsulating the investigator's notions and pre-understandings<sup>10</sup> of the phenomenon. Consistent with the

---

<sup>10</sup> Pre-understanding, as a concept, is grounded in hermeneutics – a process humans use “to understand a complex whole from preconceptions about the meanings of its parts and their interrelationships (Klein and Myers, 1999).” When scientists debate over their claims, then come to a shared understanding, they are engaged in hermeneutics — “the art of interpretation” (Manicas, 2006).



principles of hermeneutics, the data analysis process uses the conceptual lens as an initial theory, which when combined with data, creates an iterative process between field data and analysis. The initial theory is refined through this iterative process, resulting in a refinement of the conceptual lens. Theory can be used to analyze data. Walsham (2006) refers to this development between theory and data, in an explicit form, as the “data-theory link.” As the study progresses from the initial case to the next case, the conceptual lens will be refined with emergent insights from data-theory linkages, which will then feed back into the iterative process.

This section discusses our approach to data analysis in this interpretive case study research. Our approach consists of the following three steps, adapted from Walsham (1995), in reporting methods for data analysis: (1) recording field interviews and data, (2) analyzing field interviews and data, and (3) developing the iterative process between field data and theory, or the data-theory link. These steps are consistent with the traditional data analysis sequence suggested by Miles and Huberman (1994, pg. 85).

#### **6.6.1 Recording Field Interviews and Other Data**

Organizing and documenting data collected are methodological considerations for how data will be recorded, and then used for reporting. Yin (2003) recommends the creation of a database for the recording of case data, so that case data remains a distinct object from the case study report. “Every case study project should strive to develop a formal, presentable database, so that in principle, other investigators can review the evidence directly and not be limited to the written case study reports (Yin, 2003). As a result, a case study database increases the reliability of the entire case study.”

*Case Study Database:* A case study database was created to record notes used to capture themes and insights from each interview. For this study, notes primarily are the result of interviews, observations and document analysis (Yin, 2003). “Case study notes must be stored in a manner that other persons, including the investigator, can retrieve efficiently at some later date (Yin, 2003).” Microsoft Excel

was used to store case study notes and to create the case study database. Microsoft Excel allows for field notes to be organized and recorded in a tabular format, consistent with the organization of the conceptual framework, and facilitating easy retrieval at a later time.

Interviews were recorded using a digital voice recorder. The voice file was then transcribed verbatim into a Word document that could be easily used by the investigator for coding, annotating and developing field notes. The process of transcribing increases familiarity with the data, which facilitates the process of linking data to theory (Walsham, 1995; Lapadat, 1999). Themes, issues and insights were culled and synthesized from the transcription in an iterative manner, then added to the case study database (Eisenhardt, 1989; Walsham, 2006).

A final point of clarification regarding the case study database with respect to case study notes and documents: An important function of the case study database is to provide a mechanism to store, retrieve and in general, manage the case study notes. This is considered a data management function (NSTC, 2009). Data management functions allow for a defined set of operations to be performed on the case study notes to confirm findings by others at a later time (Yin, 2003). For this empirical study, the case study database will store both source data and coded data.

### **6.6.2 Analyzing Field Interviews and Data**

This section describes the within-case analysis approach for each case. Our approach to within-case analysis is based upon methods from Eisenhardt (1989), Miles and Huberman (1994), and Yin (2003) to first build a coded field data file as a result of our coding process, and then to write a detailed case study report. The coded field data file and the case study report of each case are then added to the case study database.

As described earlier, field data refers to data collected from interviews, direct observation, and document analysis (Walsham, 2006). Interviews were transcribed after they were conducted. Transcriptions were verbatim, capturing the flow and nuances of the conversation (Walsham, 1995). For each site, transcriptions, notes

and documents are coded to produce a coded field data file for each case. This process of building the coded field data file for a case is called “within-case analysis” (Eisenhardt, 1989; Miles and Huberman, 1994). Building a coded field data file as part of within-case analysis satisfies “best practices” in data management (Eisenhardt, 1989) by generating detailed case study reports (Eisenhardt, 1989), and constructing a chain of evidence (Yin, 2003).

Some additional information about the coded field data file bears mentioning, since it performs an important intermediary function towards the development of case study reports. A coded field data file will contain coding categories for concepts and descriptive notes. This is the result of the coding process. Codes are tags or descriptive labels that assign meaning to descriptive or inferential information gathered from fieldwork (Miles and Huberman, 1994). The objective of the coding process is to find meaning in the data that will lead to the generation of insight (Eisenhardt, 1989). Insight can result when data is linked to a concept and assigned meaning about its significance in a given context (Miles and Huberman, 1994). The objective is to link data into concepts, which then lead to insights that reveal themes and patterns (Miles and Huberman, 1994). Themes and patterns are also coded. We refer to this as the “data-concept linkage” — a process by which to abstract concepts from the data. Strauss and Corbin (1994) refer to this process as “open coding.” In building and reusing the coded field data file, we adhered to the following constraint from Yin (2003): “The assignment of codes and coding categories shall organize the data and case study notes in a manner that allows for retrieval and reuse at some later date by other investigators.”

The final step of the within-case analysis is to finalize the detailed case study report. As the coded field data file was filled in with coding categories, descriptive notes, analytical notes (such as those for data-concept linkages), we also extracted themes and trends to simultaneously write the detailed case study report. This was done while the information was fresh. This iterative process involved moving back and forth between the coded field data file, as well as the source data. The strategy is to perform this iterative process between these components of the within-case

analysis, allowing for patterns to emerge, and for the case to sufficiently harden into a stand-alone entity (Eisenhardt, 1989).

### 6.6.3 Developing the Iterative Process Between Field Data and Theory

Embedded in the iterative process of writing the detailed case study report is revising parts of the conceptual lens that embody our notions about the phenomenon (Miles and Huberman, 1994). We refer to our conceptual lens to guide us in the coding process, such as performing a data-concept linkage, and selecting coding categories. We also refer to our conceptual lens for guidance in data analysis and linking coded data themes and patterns into our conceptual framework. We referred to this iterative process between data and theory as the data-theory linkage (Walsham, 2006).

The iterative process between field data and theory refinement will have us move back and forth between the “coding process” and the “analysis process.” The Analysis process focuses on how evidence is to be analyzed. Yin (2003) refers to the analysis of evidence as a component of “an analytic strategy.” The analytic strategy defines priorities for what it is to analyze and why. The analytic strategy encompasses a broader strategy, which can address the need for doing analysis at the level of the whole case, versus a subset of it (Yin, 2003). Yin (2003) recommends the development of a strategy for the analysis of evidence prior to starting a case study. Yin (2003) gives the following reasons for developing an analytic strategy prior to data collection: (1) helps the investigator treat the evidence fairly, (2) helps the investigator produce compelling analytic conclusions by guiding the investigator to look at data in many divergent ways (Eisenhardt, 1989), and (3) helps rule out alternative interpretations.

Two analytic strategies from Yin (2003) are (a) relying on the study’s Theoretical Framework/Conceptual Lens, and (b) defining and testing rival explanations. These two strategies are described below:

*Relying on a Theoretical Framework/Conceptual Lens:* In qualitative data analysis, use of the conceptual lens is the preferred analysis strategy, because it provides

guidance for tracing back to a set of research questions, the objectives and design of the case study, and concepts grounded in the literature review.

*Defining and testing rival explanations:* Rival explanations strategy can be paired with the first strategy, which relies on the conceptual lens, to define and test rival explanations. Defining and testing rival explanations involves, first, being aware (ahead of time) of a rival relationship, and second, looking at the data in “divergent ways” to collect evidence of possible “other influences” (Yin, 2003; Eisenhardt, 1989). In our study, there is a possible rival explanation between e-Infrastructure as a stimulus and e-Infrastructure as a resource. We inquire if e-Infrastructure is working as a stimulus and introduced into the environment of a science discipline through external influence, such as a funding agency. We also inquire if e-Infrastructure develops within a science discipline as part of its practices; i.e., e-Infrastructure is brought in as a resource, instead of a stimulus.

The objective is to build theory. Building theory is done through a process of theory refinement or shaping theory (Eisenhardt, 1989). It’s important to remember that this study started with an initial theory, defined in Chapter 4 – our conceptual lens; it is for this reason we build theory through a refinement (shaping) process. Results from the within-case analysis, along with analysis strategies and tactics used in case study research, themes, patterns and relationships should begin to emerge (Eisenhardt, 1989). Building theory involves the sharpening of the constructs of the conceptual lens by (1) refining definitions of the constructs, and (2) building a chain of evidence through repeated refinements of the conceptual lens (Eisenhardt, 1989; Yin, 2003). Eisenhardt (1989) describes the shaping process as “occurring through constant comparison between data and constructs, so that accumulating evidence from diverse sources converges on a single, well defined construct.” The result of the shaping process helps the case study to strengthen construct validity (Yin, 2003).

*Summary of the Data Analysis methodology:* We have described our approach to data analysis as an interplay between field data and the conceptual framework towards

the objective of building theory. Theory that closely fits the data is considered good theory (Eisenhardt, 1989). This approach is based mainly on the scholarly works on case study research methodology of Eisenhardt (1989), Miles and Huberman (1994), Walsham (1995, 2006), Yin (2003) and Maxwell (2005).

A data analysis process was described, consisting of the following three sub-processes adapted from Walsham (1995): (1) recording field interviews and data, (2) analyzing field interviews and data, and (3) developing the iterative process between field data and theory. Sub-process 1, recording field interviews and data, described how interviews were recorded and documents annotated to corroborate data from other sources. Sub-process 2, analyzing field interviews and data, described our within-case analysis approach. Sub-process 3, developing the iterative process between field data and theory, described the interplay between field data and theory refinement, and how we used the conceptual lens to guide analysis, and conversely, how coded data helped shape theory.

An analysis process was described aimed at theory building. Analytic strategies for the analysis of evidence, the definition of priorities for what to analyze, and the “why” of what is happening, were discussed. Analytic techniques, working alongside the analytic strategies, were also discussed.

## 6.7 Summary

This chapter described the research methodology of this study. An interpretive approach to guide the research was described. A case study research methodology to guide the empirical inquiry of this study was described. The data gathering and data analysis methodologies for this study were also described.

## CHAPTER VII: EMPIRICAL RESEARCH

### 7. Case Studies: Introduction

In Chapter 4, our initial Concept Map was developed based on a review of the literature. Figure 5 in Chapter 4 shows the initial Concept Map. As was described in Chapter 4, the initial Concept Map consists of six primary components: Aspects of a Science Discipline, Environment of a Science Discipline, e-Infrastructure Development Process, Coevolution Relationship, Scientific Discovery, and ICT Investments Stimulus. This chapter is the start of the empirical work of this study. It describes data on Information and Communications Technology (ICT) investments to start building the foundation for the evidence to support the hypothesis of this study.

#### **Stimulus: Information and Communications Technology Investment**

As described in Chapter 1, nations have made and continue to make significant investments in large-scale Information and Communications Technology (ICT) infrastructures with the intentions of stimulating revolutionary scientific progress. It is hoped that this progress will result in major breakthrough discoveries and innovations. Our Concept Map represents these investments as ICT Investments Stimulus (Figure 5). ICT investment stimulus, as represented in our Concept Map, works on effecting change on the e-Infrastructure development process and Aspects of a Science Discipline (represented as directed arrows).

Our hypothesis is that a significant majority of the ICT investments stimuli were allocated for technological e-Infrastructure development, with the assumption that these investments in technology were going to dramatically lead to increasing scientific discovery. Thus, the impact of the stimulus on the scientific disciplines and the rate of discovery were supposedly indirect and mediated through the e-Infrastructure development process. Our Concept Map depicts the direct effect of the stimulus on e-Infrastructure development with a thick arrow directed from

stimulus to the technology side of the e-Infrastructure development process. The development of e-Infrastructure was conceptualized into seven categories.

To show the extent of the ICT investment stimulus, we analyzed public budget reports for U.S. federal agencies expenditures on networking, IT research and development. In the remainder of this section, we describe U.S. federal agencies' investments in the creation and support of a shared national e-Infrastructure, supporting Basic Research and Development (R&D), and Applied Research. These are referred to in Chapter 1, as components of an innovation ecosystem.

In 1991, Congress passed the High-Performance Computing Initiative<sup>11</sup> (HPCCI), setting forth requirements for federal agencies to establish goals and priorities for federal high-performance computing research, development and networking. Prior to 1989, federal investments in computing and communications were partitioned into three separate streams (Blumenthal, 1998): R&D aimed at high-performance computing, including hardware and software; R&D aimed at computer networking; and R&D aimed at computational<sup>12</sup> science, or the joint advance of computing with natural scientific disciplines (e.g., physics or genomics). Beginning in 1989, informal interactions between program managers at the Defense Advanced Research Projects Agency (DARPA), the National Science Foundation (NSF), the Department of Energy (DOE), and the National Aeronautics and Space Administration (NASA) resulted in the recommendation of combining these three streams. This led to the creation of the HPCCI in 1991.

We collected data on investments in the U.S. from reports of the federal agencies involved in the support of computing and communications for science research. Data were collected from reports archived by the Networking and Information

---

<sup>11</sup> Also known as the High-Performance Computing Act of 1991.

<sup>12</sup> HPCCI started a trend of increasing investments in computational science to stimulate advancements in scientific discovery. Computational science refers to the application of software, mathematical models, and algorithms and advanced computing capabilities towards understanding and solving complex problems (Benioff et al, 2005).



Technology Research and Development (NITRD<sup>13</sup>) Program. These reports span from 1992 through 2011. They report on investments in a broad spectrum of technologies intended to support scientific advancement, economic competition and national security. *Ex-Post*, based upon our analysis of the ex-ante patterns of investment in different categories, over 20 years, we will show, that these investments can be grouped into three broader categories. (1) Investments in technology for R&D, applied research, and the development of a shared national ICT infrastructure were grouped into the ***Technological Infrastructure Investment (TII)*** category. (2) Investments in data management systems and processes were grouped into the ***Data Infrastructure Investment (DII)*** category. (3) Investments in workforce development (education, training, and curriculum development), social and organization management systems can be grouped into the ***Socio-organizational Infrastructure Investment (SOII)*** category. Figure 7 below graphs the investment curves for TII, DII and SOII categories from 1992 to 2011.

Starting in 1992 through 1996, these reports indicated that investments made at an average rate of approximately \$730M over four years were aimed at stimulating growth in high-performance computing capabilities, advanced software technologies and algorithms, and computer networking on a national scale. Education, training and curriculum (SOII) development received significantly less investment at an average of \$122M per year over four years. From 1992 through 1994, the cumulative percentage investment increase was 66% in Technology Infrastructure investments. From 1995 through 1996, there was a cumulative percentage decrease of 23%.

---

<sup>13</sup> The [NITRD program](#) is characterized as the U.S.' "primary source of federally funded revolutionary breakthroughs in advanced information technologies such as computing, networking, and software." Additionally, the NITRD program is further characterized as "a unique collaboration of U.S. federal research and development agencies. "The NITRD Program stems from the [High-Performance Computing \(HPC\) Act of 1991](#) (Public Law 102-194) as amended by the [Next Generation Internet Research Act of 1998](#) (Public Law 105-305). These laws authorize Federal agencies to set goals, prioritize their investments, and coordinate their activities in networking and information technology research and development."

In 1997 through 2000, we found investments in systems supporting digital data, but in the context of “knowledge repositories” and remote visualization of environmental data. Although investments in high performance computing, storage and communications technologies declined from 1995 to 1997, appropriations for them were noticeably greater than investments in human resources, education and training. The average investment in Technology Infrastructure was \$1.2B per year over four years, an increase of 64% from the previous \$730M per year over the previous four years. The average investment in human resources and workforce development was \$49M per year over four years, a drop of 60% from its previous amount of \$122M per year over four years. This data shows the shift in priority from the already low investment in social-organizational infrastructure to an even lower investment. From 1997 to 2000, the rate of investment in Technology Infrastructure averaged a 15% increase per year.

Starting in 2001, investments in high performance computing were directed at two specific components: high-end computing for research and development, and high-end computing for infrastructure and applications. These investments fluctuated slightly between 2001 and 2005. From 2004 to 2007, investments in high performance computing infrastructure increased on average by 6.85%. From 2007 to 2008 these investments increased further by 16%. As a result of the Stimulus funding, from 2008 to 2009, investments in high performance computing increased by approximately 48%. In 2010, investments declined by 3.7%. Then, in 2011, they declined dramatically by 25%.

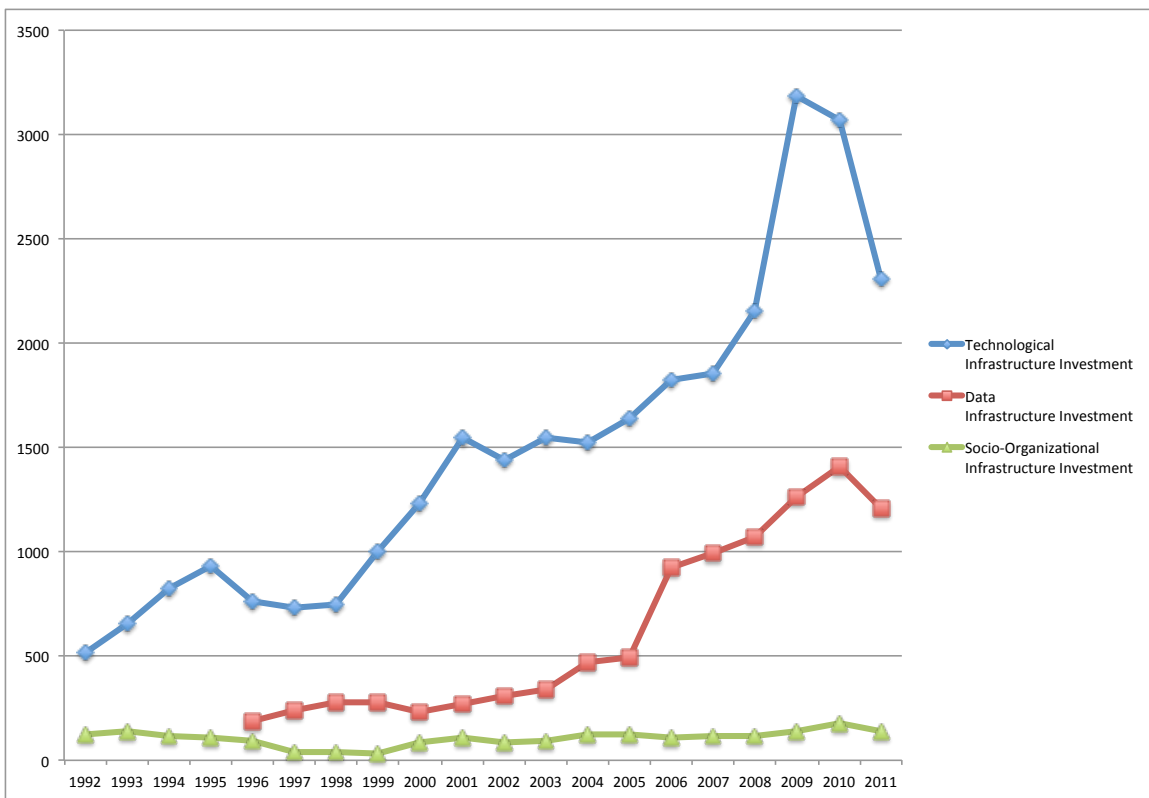
In 2001, an investment sub-component for data management technologies was included in the infrastructure and applications investment component of the fiscal year 2002 budget. This finding is a bifurcation in the infrastructure investments pattern that indicated a rise in priority for the development of e-Infrastructure for data. Investments in information management and data sharing increased in priority due to increasing demand to make large amounts of data easily available and useful across multiple communities of interest (NITRD, 2001). Between 2001

and 2005, investments in information management infrastructure increased on average by 17%.

Between 2005 and 2006, investment in information management and data sharing increased dramatically by 86%. This was a result of a new program on Cyber Security and Information Assurance. Between 2006 and 2010, investments increased by 11.16%, and in 2011, investments decreased by 13.8%.

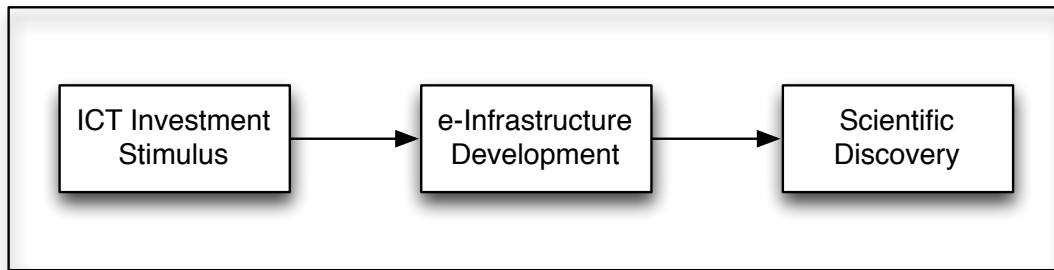
Investments in social, economic and workforce development between 2001 and 2011 were kept relatively flat, with a minimum investment of \$85M and a maximum investment of \$179M. The annual average investment was \$121M.

Figure 7 shows that federal investments in technological infrastructure have increased significantly over the past 20 years. Starting in 2000, investments towards developing a data and information infrastructure have increased.



**Figure 7 U.S. federal investment in information and networking technology over 20 years**

Figure 7 is consistent with our theory that nations are making significant investments in technological infrastructure to create a stimulus towards increasing scientific progress and discovery. While the funding agencies also realize that scientific progress would require investments in human resources and workforce development, the comparative amounts spent in this category, and its stable allocation over the years seems to indicate that the funding agencies are primarily focused on funding technology. Investment in human and workforce development was secondary. Our intuition leads us to Figure 8. Figure 8 basically says that the observed trend in stimulus spending has been favoring investments in ICT infrastructure over the development of data and human resources. Over time, as ICT investments develop into a shared national technological e-Infrastructure, conditions and issues can arise that influence investments in other areas; e.g., towards the development of a data infrastructure.



**Figure 8 Intuition of ICT investment as a stimulus**

The term e-Infrastructure development encompasses investments, not only in ICT, but also in other types of infrastructure that supports science research, such as data infrastructure. Scientific progress then becomes the outcome of investments in e-Infrastructure development.

Investment in ICT to stimulate science research is growing (NITRD, 2012). However, the impact of these investments on scientific progress across disciplines is not well understood. Although E-Science and Cyberinfrastructure programs are relatively new forms of stimuli, for many years communities of scientists have used

ICT investments as stimuli towards generating dramatic improvements in scientific progress to advance their disciplines.

For this case study, we chose to study the impact of ICT investments stimuli on the biodiversity discipline. While other disciplines such as physics or astronomy were just as suitable, we chose biodiversity for the following reasons: First, its research activities concern issues global in scale, such as global change and conservation (Callahan, 1984; Michener et al, 2011). Yet, it was unclear if the ICT investments stimuli being promoted in programs, such as e-Science and CI, were providing a material benefit for this discipline. Second, increasing demand for comparative research was increasing complexity and challenges for biodiversity researchers, particularly those involved with data management. We wanted to understand which innovations from ICT investments were being adopted and if they were having an impact on advancing scientific progress in the biodiversity discipline. Third and finally, sources of information by which to conduct an inductive qualitative study were available either through literature or subject matter experts from which a conceptual framework could be developed that would contribute towards increasing understanding of the relationship between ICT investments and scientific progress.

## 8. Biodiversity Case: The U.S. Long Term Ecological Research Network

### 8.1 Setting the Case

#### 8.1.1 What is Biodiversity?

Biodiversity<sup>14</sup> (or biological diversity) is the continuum of the living and fossil organisms of Earth, ranging from species at the micro level (genomes) to the macro level (ecosystems), and the ecological and evolutionary processes that sustain them. Biologist E. O. Wilson, in the 1980s, introduced the term “biodiversity” to encompass the taxonomic and functional diversity of living organisms (Sugden and Pennisi, 2000). Moreover, the term helped to raise awareness of the rate at which human society growth is altering and destroying the environments that have fostered the diversity of life on Earth for more than a billion years (Wilson, 1988).

Since Darwin, the diversity of life has always been a central theme of biology (Ehrlich and Wilson, 1991). Ehrlich and Wilson (1991) characterize biodiversity studies as “the systematic examination of the full array of organisms and the origin of this diversity, together with the methods by which diversity can be maintained and used for the benefit of humanity.”

Increased awareness of the presence of a close linkage between the conservation of biodiversity and economic development, and the urgency to understand the relationship of human action on the extinction of wild species and ecosystems gave rise to biodiversity studies as a discipline. Conservation of biodiversity is the scientific study of Earth’s biodiversity, with a focus on how to protect species, their habitats, and ecosystems from excessive rates of extinction by human activities or

---

<sup>14</sup> Other definitions of biodiversity can be found in the following sources: “Report of the Second Workshop on the Biodiversity Observation Network” (1999). This report defined nature and scope of biodiversity “to include taxonomic composition and phylogenetics, genomic traits, species interactions, ecosystem function, and landscape patterns that characterize life on earth.”

other agents (Soule, 1985; Sterling et al, 2003). A goal of biodiversity conservation is to provide principles and tools for preserving biological diversity (Soule, 1985)<sup>15</sup>.

### 8.1.2 The U.S. Long Term Ecological Research Network

To study the process of e-Infrastructure development as a stimulus, we focus on the research activities of biodiversity science within a long-term ecological research program created by the National Science Foundation. The program is called the U.S. Long-Term Ecological Research (LTER) Network (hereafter US-LTER).

Why are we studying biodiversity science within a long-term field ecology research program? Biodiversity has been characterized as part of ecosystem function (Hobbie, 2003). Biodiversity gained interest by 1980 when a rise in awareness occurred on deforestation and species extinction that brought global issues into sharper focus (Wilson, 1988). The US-LTER program includes scientists researching changes in biodiversity at sites supported by US-LTER.

We chose to study the process of e-Infrastructure development within US-LTER for three reasons: (1) Its duration provides an opportunity to understand the effects of technology use on the science, and conversely, the requirements of the science shaping technology. (US-LTER is the only federally funded biology program whereby a retrospective study can be conducted for a period of 30 years.) (2) To inquire how technology investments have changed data and data management practices over the history of US-LTER. (3) To inquire how changes from Technology Infrastructure development and data infrastructure development over multiple decades changed the rate of scientific progress in the LTER Network. For the purposes of this study ecological science encompasses biodiversity science.

#### **Events that led to long-term ecological research:**

Prior to 1960, data for biodiversity-ecological research were collected primarily by single or small groups of investigators using ground plots of less-than-or-equal-to

---

<sup>15</sup> While not stated explicitly, these definitions of Bio-Diversity also carries with them an implied “normative,” “ethical,” or “idealistic” stance that environment and species are “natural” phenomenon and should not be disturbed or altered by human action. However, taking sides in this debate is beyond the scope of this dissertation.

one meter squared, over short periods of time, usually every one to three years (Brown, 1994; Gosz et al, 2010). This severely limited scientists' abilities to conduct long-term science research, especially when the effects of the phenomena of interest typically accumulate over long periods of time (Brown, 1994; Michener et al, 1997). By the 1960s, interest grew in society and science disciplines for a greater understanding of issues involving global change, biodiversity and sustainability. This caused scientists to question how biodiversity-ecological patterns and processes vary in time and space, and to inquire about the causes and consequences of this variability (Levin, 1992).

In 1969, the U.S. government passed the National Environmental Policy Act (NEPA), which mandated policy for the government to “create and maintain conditions under which man and nature can exist in productive harmony.” This report and several others that followed raised the demand for long-term ecological research dependent upon the preservation of long-term data sets. These reports described the importance of investment in long-term ecological research at all levels of government and in most sectors of the private economy (Callahan, 1984). Moreover, they recognized the need for greater predictive capability to improve the quality impact analyses through the use of long-term data. More will be described later about types of stimulus.

Between 1977 and 1979, the NSF conducted three workshops that resulted in a solicitation for proposals in the area of long-term ecological research. In 1980, the NSF awarded six proposals and launched the US-LTER program with six initial sites. The goal<sup>16</sup> of the US-LTER program was to stimulate long-term and comparative research on ecological processes among a network of geographically distributed and ecologically diverse field sites (Callahan, 1984). “Long term” in the context of ecological time scales can span decades to centuries. Creation and preservation of long-term data sets are important, because they “provide a context to evaluate the pace of ecological change, interpret the effects of ecological change, and forecast

---

<sup>16</sup> A formal description of the mission and aims of the US-LTER Network can be found in Hobbie et al (2003).



biological responses to change (Gosz et al, 2010).” The program has evolved into 26 research sites to date, involving a wide range of ecosystem types distributed across the U.S. and other territories.

#### **8.1.2.1**      *Characteristics of the US-LTER Network*

In this section, we describe the physical aspects of the US-LTER Network, the Problems and Puzzles studied by its organization, and its governance.

##### ***Physical Aspects of the US-LTER Network:***

The US-LTER program is represented by 26 research sites that cover a wide-range of ecosystem types, spanning broad ranges of both environmental conditions and degrees of human domination of the landscape (Gosz et al, 2010). Ecosystem types include agricultural lands, alpine tundra, barrier islands, coastal lagoons, cold and hot deserts, coral reefs, estuaries, forests, freshwater wetlands, grasslands, kelp forests, lakes, open ocean, savannas, streams, and urban landscapes. The geographic distribution of sites representing ecosystem types is very wide and diverse, spanning from Alaska to Antarctica, and from the Caribbean to French Polynesia.

Collectively, the 26 sites constitute the LTER network. NSF 12-524 characterizes the LTER Network as “a collaborative effort among more than 1,500 scientists and students investigating ecological processes over long temporal and diverse spatial scales.” Moreover, the LTER network is characterized as a resource that enhances the opportunities and capabilities of the individual sites to enable comparative research across sites (NSF, 2012).

##### ***Problems and Puzzles:***

NSF’s investment in the US-LTER program is to fund research to address questions that cannot be resolved through short-term observations or experiments (NSF, 2012). Five Core Areas of research were defined at the start of the US-LTER program in 1980, and have persisted for 30-plus years, providing a contextual

framework for long-term ecological research. Each site develops Problems and Puzzles in research programs within the five core areas:

- Pattern and control of primary production<sup>17</sup>.
- Dynamics of populations of organisms selected to represent trophic<sup>18</sup> structure.
- Pattern and control of organic matter accumulation in surface layers and sediments.
- Patterns of inorganic inputs and movements of nutrients through soils, groundwater and surface waters.
- Patterns and frequency of disturbances.

These five core areas of research influence the data collection process at the sites. Data sets produced within the Five Core Areas are referred to as “Core Data.” Core Data provides the foundation for testing major ecological concepts and theories, for challenging Problems and Puzzles in existing paradigms in ecology and ecosystem science, and for developing new paradigms (NSF, 2012).

***Organization:***

The LTER Network Office is a central resource of the LTER program that “provides leadership in developing and implementing data and information management standards and protocols for the LTER Network, as well as for the broader community of environmental scientists (NSF, 2012).”

The LTER Network Office has a network-wide responsibility to facilitate and mobilize network science (Gholz interview, 2012). “Network Science” refers to the ability of exchanging data, information, collaborating, performing experiments, and other related activities involving conducting science research and education over a

---

<sup>17</sup> “Primary production” is the production of chemical energy in organic compounds by living organisms. It represents a globally important flow of carbon between the atmosphere and the biosphere (Roxburgh et al, 2005).

<sup>18</sup> “Trophic structure” refers to the way in which organisms utilize food resources, and depicts where energy transfer occurs within an ecosystem (Williams and Martinez, 2004). Also referred to as “food chains.”

common, shared network infrastructure interconnecting the LTER sites (Michener et al, 2011). In addition to its charge to facilitate research and education, the Network Office also facilitates governance activities for the LTER Network.

***Governance:***

Over its 30-plus year history, the governance structure of the US-LTER has evolved to represent all 26 sites and to facilitate working as a coordinated Network. Gosz et al (2010) describes the governance structure for US-LTER as consisting of “an elected Chair and an Executive Board comprised of nine rotating site representatives and one member selected to provide expertise on information management.” There are eight standing committees that support and inform the governance process on a broad set of strategic activities: Climate, Education, Graduate Students, Information Management, International, Network Information System, Publications, and Social Science.

## **8.2 Sources of Information for Studying e-Infrastructure Development and Biodiversity and Ecological Research in the US-LTER Network**

Sources of information for studying biodiversity and ecological research in the US-LTER program come from source documents and informant interviews. In this section, we describe the types of source documents we used in our study. We shall then describe the informants who we interviewed and why we interviewed them.

### **8.2.1 Source Documents**

In Chapter 6 we described the following four types of documents: Strategic, Programmatic, Peer-reviewed and Informational. We used these four types of documents for the classification of source documents in our study of the US-LTER Network. In general, they could be used for the study of source documents of a science discipline.

*Strategic documents* were described as types of documents that normally provide information expressing a vision or strategic goals that focus on challenges facing science or a particular science discipline. Strategic documents can also contain

information on investments projected to fund programs to conduct research. The Decadal Plan for LTER (Robertson et al, 2007) and the Data Management for LTER: 1980-2010 are examples of strategic documents for the LTER Network.

We analyzed the Decadal Plan for data on the vision, goals and challenges of the LTER Network. The Data Management for the LTER report is a strategic document that focuses on the secondary use; i.e., the use of data sets by users not associated with their original collection (Robbins, 2011). It provided data on the contentions between the goals of producing long-term shareable data sets and who gets credit for discovery.

The Cyberinfrastructure Vision for the 21<sup>st</sup> Century (NSF, 2007) is a strategic document that was used for data on the vision and the goals of the NSF for the advancement of Cyberinfrastructure and its use in transformative approaches to scientific and engineering discovery, as well as learning across all disciplines. Similar to the previous strategic document, the NITRD documents (described above) report on U.S. federal agencies' coordinated investments in advanced ICT research and development. NSF budget reports<sup>19</sup> provided data on investments for ICT to support research and fund research programs.

*Programmatic documents* were analyzed to identify goals of funded programs. Program solicitations or competitions are programmatic documents that describe the goals and requirements of a program aimed at achieving a particular outcome. An example of this is a programmatic document that funded proposals to stimulate increases in cross-site comparative science. We analyzed and studied several program solicitations of the LTER program for data supporting concepts of an e-Infrastructure development process and the Aspects of a Science Discipline in our conceptual lens.

*Peer-reviewed documents* were one of our primary sources of information. We used them to study the LTER Network using our conceptual lens, its community of

---

<sup>19</sup> NSF budget reports are archived at <http://www.nsf.gov/about/budget/>

scientists, Problems and Puzzles, and a methodology based on a body of knowledge. Data from these documents were used to develop interview questions for informants. Data from peer-reviewed documents were compared with data collected from informant interviews for checking validity and for generating thicker descriptions of events and causal relationships.

*Informational documents* were studied to learn of events and issues affecting the LTER Network. Minutes and reports of committees in the LTER Network were analyzed to learn about the LTER from different perspectives, such as coordination between sites, and issues involving governance and funding. LTER Databits contain informational documents written primarily by information/data managers about work they are doing at their sites. LTER Databits discuss the use of technology across different time periods, giving us an indication of the technologies that were adopted to advance the LTER Network. The LTER web site was studied for informational documents. We found information that documented significant events and milestones of the history of the LTER program. The NITRD and NSF web sites archived informational documentation, which were studied for information on investments into programs supporting ICT and science research.

### **8.2.2 Informants**

Informants are subject matter experts either in the LTER Network, the process of e-Infrastructure development, or both, as described in Chapter 6. Technologists or senior scientists who have witnessed the evolution of the LTER program or witnessed investments in technology aimed at providing a stimulus to scientific progress met our criteria for selection as informants. Since 2007, the investigator has interviewed subject matter experts in biodiversity and e-Infrastructure concerning technology, instrumentation, data and processes to support biodiversity research. The number of informants interviewed in biodiversity and the LTER program was seven. The number in ICT investments was two.

### 8.3 Data Analysis: Linking Data and the Conceptual Framework

Our study is exploratory, because we're exploring a phenomenon (e-Infrastructure development or Cyberinfrastructure), which is relatively new and is not well understood. Exploratory research is appropriate when the phenomenon involves a new area, and the study of it aims to build or emerge a theory about it (Miles and Huberman, 1994). The problem area to be solved is either unstated or an ambiguous problem, which has to be framed and reframed as we go (Miles and Huberman, 1994).

Data were coded using an inductive and grounded approach (Strauss and Corbin, 1998). Source documents were analyzed by creating codes to organize and assign data into categories. We coded using words that reflect action (Charmaz, 2006). Informant interviews were recorded, and then transcribed verbatim into a Word document, which was then coded.

To record and organize coded data, we created a case study database (Yin, 2003). Data from analysis of source documents and informant interviews were recorded in the case study database. The case study database is a distinct object from the case study report (Yin, 2003).

Coded data were assigned to categories for comparison with concepts, themes and issues. For example, we coded data for issues and assigned them to a category for issues in the case study database. Data associated with that issue, such as a technology stimulus, was assigned to a category for stimuli for comparison with the issues category. Maxwell (2005) refers to these categories as organizational categories, because they facilitate organization of coded data for further analysis.

We started the data analysis process with an initial Concept Map (Miles and Huberman, 1994); see Figure 5, Chapter 4. Our objective was to make connections from stimulus to scientific discovery, building a chain of evidence through repeated refinements of the conceptual lens (Eisenhardt, 1989; Yin, 2003). Connections refer to events, issues, actions, processes, activities, etc., which create conditions whose effects can originate from a stimulus and result in scientific discovery.

The LTER Network provides the setting of our case study data. We start by looking at the stimulus of participating in the LTER Network.

### 8.3.1 Stimulus: ICT Investments

What ICT investments occurred to stimulate research involving global change, biodiversity and sustainability?

Data on events resulting from ICT investment stimuli were ordered chronologically, leading up to and into the 1980s at the start of the LTER program. Data analysis revealed two levels of ICT investment stimuli: those external to the LTER community, and those internal to the LTER community. External ICT Investments Stimuli refer to ICT events that lead towards the creation of a common national infrastructure for science research, development and education. For example, in 1985, NSF provided stimuli to network super computing centers to increase access to and sharing of high-end computing resources for researchers around the country, and for all science and engineering disciplines (NSF, 2003). External ICT investments can be thought of as stimuli at a macro level that can be leveraged by science disciplines to provide internal stimuli.

Internal ICT investments stimuli refer to ICT events internal to the LTER community only. For example, in 1988, the LTER program provided stimuli to adopt remote sensing data, combined with GIS<sup>20</sup> technology to enhance cross-site comparative science and the analysis of ecological information (Robbins interview, 2012).

We will now describe technologies that evolved from external ICT investment stimuli and were adopted by the LTER program over its 30-year history.

#### **External ICT Investments Stimuli:**

External ICT investments became significant ICT events during the period of 1980 through 1990 — the first decade of the LTER program. ICT systems (or artifacts) developed from external ICT investments. These technology components

---

<sup>20</sup> GIS technology is described below as a technology external to LTER.

introduced external stimuli into the LTER Network as they were adopted to address its science research issues.

***Advances in Computer and Network Technologies:***

In 1981 through 1986, IBM introduced a series of personal computers, starting with 8-bit microprocessors (Intel 8088) and evolving to 32-bit microprocessors (Intel 80386). These personal computers included storage capacity, starting with floppy disk drives, which were augmented with hard disk drives up to 30 megabytes. The biodiversity-ecological research community viewed the 1980s as a period of convergence in capabilities of micro-, mini- and mainframe computers. This decade witnessed computer-processing speed approximately double each year — the phenomenon known as Moore’s Law. Moreover, we saw decreases in computer memory prices by approximately 50% every two years; increases in hard disk drives capacities, along with decreases in prices per megabytes, and increases in the capacities and proliferation of communications networks (Stafford et al, 1994; Robbins, 2011).

Moore’s Law would stimulate further investments in computing, storage and networking, while dramatically reducing the cost, and increasing the performance of the technology (faster and cheaper). Buying power would also grow at a much faster rate, causing price elasticity (Gallaugher, 2008). Price elasticity and buying power are affected when an external ICT investment stimulus, as in the R&D of a new microprocessor for high-performance computing, results in a new computing platform that’s faster and cheaper than the previous generation. We witnessed the effects of a stimulus and its results affecting buying power in the transformation of computing from mainframes (1960s), to minicomputer (1970s), to microcomputers (1980s), to Internet computers (1990s), to ubiquitous computing using sensors such as RFID, or radio-frequency identification (2000).

***Remote Sensing and Geographic Information Systems:***

As computing, storage and communications networks increased in capacity and decreased in cost, adoption of remote sensing and Geographic Information Systems



(GIS) technologies dramatically increased in science applications. Remote sensing refers to techniques that use sensor devices to detect and record signals emanating from target(s) of interest not in direct contact (thus, at a distance) with the sensor (Short, 2010). Use of the term "remote sensing" was associated with "aerial photography." Today, it's associated with "satellite imagery," in particular high-resolution satellite imagery. Earth sciences, such as biodiversity and ecology, use data from remote sensing systems to acquire information about the Earth, identify its patterns and better understand its processes (Short, 2010; Stafford et al, 1994).

GIS refer to "a class of computer system that supports collection, storage, manipulation and query of spatially referenced data. The GIS system normally includes a graphic interface for displaying geographic maps and optional tools for acquisition and validation of data (Fortuner, 1993)." The availability of lower-cost desktop GIS systems increased the adoption of these systems by the biodiversity-ecological science communities in order to geo-reference their data (Shugart et al, 1998)

#### ***Database Management Systems:***

Database management systems (DBMS) in the 1980s were characterized as having shortcomings to support the demands of the science communities facing dramatic increases in size and complexity of their data sets. ICT investments in the 1980s aimed research at developing general purpose DBMS with greater capabilities than relational DBMS (RDBMS) for representing complex spatial data (Stafford et al, 1994). In 1985, University of California Berkeley developed POSTGRES, a successor to Ingres, to address problems with precursor contemporary DBMS.

#### ***Integration of Electronic Mail:***

Scientists were early adopters of electronic mail (email). Prior to the 1980s, email came as an application in a host computer system — called "host-based email systems" (Partridge, 2008). Scientists working in research organizations or universities used email gateways to communicate and collaborate with colleagues using incompatible host-based email systems (Van Vleck, 2001). For example, the

LTER Network Office created an email gateway linking all of the disparate email systems used by scientists participating in the LTER Network (Robbins interview, 2012). In the 1980s, proposals to establish email standards were adopted that would eventually result in a global standard for email message formats (Partridge, 2008).

### ***Internet and Web Technologies:***

Internet online systems, such as File Transfer Protocol (FTP), Gopher and a variety of tools leveraged the power of the Unix operating system and its many tools to give researchers the ability to work with the Internet. Unix's powerful client-server paradigm gave rise to many different servers on the Internet, of particular importance was HTTP servers, also known as Web servers. Combined with Web client technology standards, the Web encouraged a set of standards to support document exchange (Abiteboul et al, 2000). Many LTER sites implemented Gopher servers to discover sites with available data sets. With the Web and browsers, such as in Mosaic, LTER sites were able to publish information about their data and provide access to it. The Web provided a layer of abstraction that greatly simplified discovery and access to data that were more complex at the Unix level.

### ***XML for Metadata Management:***

The eXtensible Markup Language (XML) was designed to transmit structured data, and to reconcile heterogeneous data found in diverse data sources (Abiteboul et al, 2000). The integration of XML with database management systems motivated the use of metadata management. Before XML became a standard in 1998, metadata management in the LTER was left up to the site information manager to determine how metadata would be captured, stored and structured for users (Michener et al, 2011). XML allows many different types of tools to be used to create and work with metadata. XML transformation systems can be used to transform LTER metadata documents into other metadata formats, thus facilitating systems using other metadata standards to represent LTER data in other systems (Michener et al, 2011).

### ***Sensor Networks:***

Environmental monitoring data, such as precipitation and soil temperature, are traditionally collected using sensors, in an automated fashion, at meteorological stations. External ICT investments have provided stimuli resulting in innovations in a variety of low cost sensors that have led to adoption in the biodiversity and ecological research discipline (Michener et al, 2011).

Sensor network devices have been enhanced with other technologies. Wireless technologies, connected to networks like the Internet, enable real-time data transmission to analysis sites (Collins et al, 2006). The availability of data storage devices in a small form factor, and at decreasing cost, has been coupled into sensor network devices. This has increased the use of sensor networks to measure environmental variables over broader spatial scales (Sheldon, 2008). Enhancing technologies through external linkages to other technologies can provide a capability to the discipline (Boyd, 1990), which advances the e-Infrastructure development process as well as the discipline itself.

Sensor networks are being used to collect new types of data, such as sound. They are increasingly used to collect unobtrusive data like motion detection or observation of an animal without disruption. (Porter et al, 2005).

Several external ICT investments that provided stimuli to the LTER Network and the biodiversity-ecological research community were discussed. Each of these external ICT investments provided stimuli that resulted in a number of technologies for general use. These technologies were designed and implemented to benefit society and all science disciplines. Conversely, as these general use technologies were adapted and linked into the general national science infrastructure, they also benefitted biodiversity as a discipline (just as the highway infrastructure benefits all who travel on the highway).

In the next section, starting with external ICT investment, we will step through the e-Infrastructure development process, using our conceptual lens. Using the LTER Network as our case, we will describe what happens when the object of an external ICT investment is introduced as stimuli into the LTER Network.

### 8.3.2 Participation in the LTER Network

Becoming an LTER Site involves a commitment to the long-term studying of phenomena through collecting and managing long-term data sets (Robbins, 2011). NSF periodically invites the scientific community to submit proposals for grants to become LTER sites.

Linking into our conceptual framework, a grant consisting of funding, for a Site to participate in the LTER Network, is considered a stimulus. A proposal for a site to join the LTER Network must satisfy prerequisites imposed by the LTER program. An applicant for LTER funding is supposed to have these pre-requisites for qualifying for consideration to LTER. Being prerequisites, these are additional stimulus for the applicant investing in underlying infrastructure to support LTER. The philosophy behind LTER is that it does not directly fund these investments; it encourages the applying institutions to search for and obtain funding from other sources to build these infrastructures. Thus, it tries to leverage funding and investments from other sources.

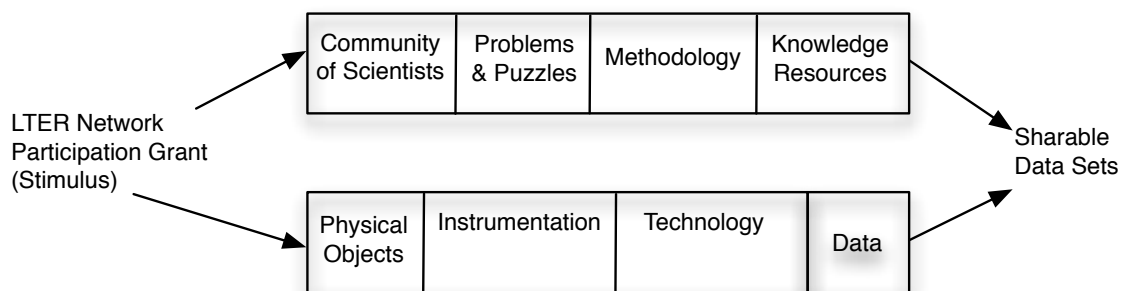
First, there is an operational infrastructure, consisting of: (a) physical (bricks and mortar) facilities to house scientists, staff, specimens, etc., for the conduct of science research; (b) technology for computation, storage and communications for the site to produce excellent local science (Robbins, 2011); and (c) instrumentation for the production of data. This infrastructure of physical, technological and instrumentation components must be present as a functional substrate upon which the stimulus can be used to achieve its goals (Gholz interview, 2012). In the case of the LTER program, its goals include the creation and preservation of long-term data sets.

Second, there is an established science community associated with the site. Linking to our conceptual framework, we characterize a site as a community of scientists, a set of Problems and Puzzles of interest to the LTER Network community, and methodologies of its community of scientists for producing excellent site science.

This science community is supported by the aforementioned operational infrastructure.

Third, LTER sites and their community of scientists were expected to produce sharable data sets in addition to providing support for their use by other sites in the Network. This was important in order for "others to carry on the work, at other places, and at other times (Robbins, 2011).

The LTER program significantly leverages the infrastructure of participating sites (Gholz interview, 2012). As a result of participation in the LTER program, at least two infrastructure components of a site are significantly leveraged: its operational infrastructure and its community of scientists. Linking back to our conceptual framework, participation in LTER acts as a stimulus on a site's operational infrastructure and its science community. As a result, this stimulus acts upon the site, enabling its infrastructure components, to produce sharable data sets.



**Figure 9 Stimulus acting upon Site infrastructure to produce sharable data sets**

The goal of the LTER program was to increase production of sharable long-term data sets by creating a network of autonomous sites that would facilitate the sharing of data sets to conduct cross-site comparative science.

Using our conceptual lens to interpret the effects of the investments (stimuli), an e-Infrastructure was being constructed in the LTER Network. This e-Infrastructure was being created on at least three levels: (1) technological, (2) science discipline, and (3) the interactions and inter-dependencies between the two. In the next section, we are seeking to understand what the effects were of e-Infrastructure

development on a science discipline by studying the e-Infrastructure development in the LTER Network.

### 8.3.3 E-Infrastructure Development Process and the LTER Network

Keeping figure 9 in mind, our objective is to understand how the e-Infrastructure development process affects scientific discovery. Using our conceptual lens as a guide, scientific discovery would be an outcome that results from the interactions between the e-Infrastructure development process and the biodiversity-ecology discipline.

Based on literature, we conceptualized Aspects of a Science Discipline as consisting of four broad categories: (1) Community of Scientists, (2) set of Problems and Puzzles, (3) Methodology, and (4) Resources. From these categories, we conceptualized how science is done. We added Philosophy of Science as a fifth category to conceptualize how science is thought about.

An inductive analysis of the data was performed to find issues or activities with a discipline-wide impact. Inquiry was focused on investments in ICT and technological infrastructure development that provided capabilities needed in the biodiversity-ecology discipline. Our conceptual lens was used to link data with concepts to increase understanding of the effects of e-Infrastructure development on the biodiversity-ecology science discipline. Data from source documents and interviews on the LTER network were also used to identify issues, before linking them to aspects within the categories of a science discipline.

In this section, we will explain the development of e-Infrastructure in the biodiversity-ecological research discipline by studying the US-LTER program. The term “e-Infrastructure” is less than 10 years old. However, we are tracing the process of e-Infrastructure development over a 30-year period when we study the US-LTER program.

**LTER Network — Decade I (1980-1990):** Prior to the 1980s, the lack of adequate data and analytical tools severely limited the research that could be conducted

involving diverse spatial and temporal scales. Into the early 1990s, the community of scientists involved with long-term ecological research argued for the capability to increase the scale of inquiry from the small area of a sample plot, habitat patch and small watershed of traditional ecology to the landscape, geographic region, continent, ocean and entire Earth (Brown, 1994). By the 1980s, the availability of new data (e.g., remote sensing data), new technology (GIS systems) and tools for analysis and modeling enabled scientists to increase the scale of biodiversity-ecological research (Brown, 1994; Stafford et al, 1994).

Investments in GIS technology led to dramatic increases in data by transforming ecological data into geo-referenced data. Stimuli on the LTER Network and ecology discipline were coming from at least two sources: government and technology. Government was concerned with global change issues, biodiversity and sustainability. Technology caused the creation of an enormous supply of primary data. This led to further investments in ICT for the development of tools to analyze primary geo-referenced data into ready-for-science data use.

By 1985, geo-referenced data was growing at an alarming rate, because of new technology capabilities, remote sensing data available from external sources, and additional stimulus coming from the NSF and other sponsors. The role of data management in LTER and in the ecology discipline started changing. Historically, data management within scientific organizations involved task oriented work (e.g., data entry, archiving, security, and quality assurance). This work was largely viewed as scientific custodial services (Stafford et al, 1994; 1996).

By 1986, the role of data management was being integrated into the research process and was rising in importance (Stafford et al, 1986). This was largely the result of a publication of a workshop volume in 1986 that described a variety of methods for data management and the development of metadata (Michener, 1986). Yet, the goal of sharing long-term data sets between sites in the LTER Network and other disciplines was not progressing.

Porter (2010) attributed the notion of sharing data as foreign to most researchers in the LTER network in the 1980s and into the 1990s. Data sharing is not a natural condition in the culture of most science disciplines (Robbins, 2011). "Most researchers are loathe to part with their data. After all, data are the raw material out of which scientific discovery, and thus scientific careers are made (Robbins, 2011)." Another reason for reluctance was that sharing data came with the responsibility of active participation in the preparation of documentation (metadata) in order for the data to be usable by others (Michener et al, 1997; Porter, 2010).

In the latter part of 1989, the NSF commissioned a study to find out what each LTER site would need to get connected to the NSFNET. Due to the network infrastructure at many universities, many sites were within close proximity to a connection; however, most were unaware of a local connection or why they would want to use it (Porter, 2010).

By 1990, a socio-organizational condition was impacting the data sharing goals of the LTER community. Stimuli coming from ICT investments had resulted in changes to data in two ways: added geo-referencing attributes, because of integration with remote sensing data; and increased size and complexity of data sets, because of spatial and temporal dimensions added from geo-referencing attributes. Robbins (2011) characterized this condition as a "social problem that called for an active social engineering solution." The LTER Governance committee was convened to address the issues of this condition. With substantial input from data managers and investigators, the committee developed guidelines for information policies at individual sites (Porter, 2010). Because excellence was of primary importance, it was feared that a dictatorial sounding policy might alienate or be rejected by investigators. This could ultimately impact the quality of local science.

An important achievement in 1990 was the publishing of the first LTER-wide core data catalog in a printed format.(Michener et al, 1990). The catalog consisted of summary descriptions of data sets, but not the data themselves (Porter, 2010).



Without a catalog through the 1980s, data users were largely unaware what sorts of data sets were being maintained by any LTER site (Porter, 2010). “For the first time it was possible to explicitly identify which data were being collected where and by whom (Michener et al, 2011).” This catalog had the affect of an internal stimulus that helped alert data users and the research community to the data resources and the potential of these resources to address a broad array of cross-site research questions (Porter, 2010).

In the late 1980s and into the 1990s, the following technology events occurred that provided a stimulus to facilitate data sharing:

- The NSFNET backbone was upgraded to 1.5 megabytes
- Microsoft Excel was widely adopted for scientific data management
- The microcomputer industry took off, bringing advanced computational capabilities to the desktop.

By the end of decade I (1989-1990), ICT investments had enabled sites in the LTER network to produce data sets that were geo-referenced. This furthered the capabilities of the community of scientists to conduct cross-site comparative research. Enhanced data sets were larger and more complex, which created challenges in data management. The process for creating metadata content and format were ad hoc, with decisions left mostly up to the information managers at each site (Michener et al, 2011).

**LTER Network — Decade II (1991 - 2000):** At the start of the 1990s, various technology events were significant to LTER and other science communities. Use of the Internet began to accelerate with the launch of Internet-based online systems (FTP, Gopher, and subsequently the World Wide Web). These systems were being widely adopted at academic institutions, industry, and government. — all of them organizations capable of connecting to the nascent Internet for searching and discovery of information (Porter, 2010). LTER sites were connecting to the Internet as well, though some scientists were reluctant to allow access to their data sets online, because of fear their data could be stolen (Porter, 2010).

Prior to 1991, research methods used at LTER sites were primarily hypothesis<sup>21</sup> driven (Gholz interview, 2012). In 1991, the marine science community joined LTER and established the Palmer site in Antarctica. Marine scientists used a data-driven research approach to develop models in response to the Five Core Areas of research (described earlier). They conducted observations, including remote sensing data, and constructed models before developing hypotheses based on empirical results. Integration with the marine ecology community was changing LTER into a more diverse and complex network.

In 1991, the NSFNET's acceptable use policy permitted limited commercial use (Robbins, 2011). As described in section 7.1, Congress also passed the High-Performance Computing and Communications Initiative (HPCCI). However, electronic mail systems were still largely incompatible, making it very difficult for LTER users to communicate with each other using the Internet. A technology solution was used to bridge the incompatibility. The LTER community simplified the use of email, which increased the demand for use of the Internet.

In 1992, the LTER Information Management Committee (IMC) funded a project to develop standard ways of exchanging metadata that would be both human and machine-readable (Michener et al, 2011). The federal appropriation for high-performance computing and communications increased by approximately 30%, and by 27% the following year.

In 1993, Mosaic — the first WWW browser — was released by NCSA. By 1993, the LTER Network had grown to 18 sites nationwide. Many LTER sites had implemented their own Gopher servers and other Internet-based servers. Internet-based technology was spreading organically across the sites, the universities, and the broader academic community.

---

<sup>21</sup> Hypothesis-driven science is typically based on situations in which very specific questions can be addressed within tightly bounded spheres of inquiry (O'Malley et al, 2009).

Demand for access to data sets increased by 1993. Sites started to adapt a sample policy developed in 1990 that met the criteria listed in the guidelines for data access. Recognizing that scientists had little experience writing data-access data-sharing policies, people involved with the governance of LTER (the LTER Network Office) and select representatives from the community of scientists, wrote a sample policy for future use. This sample policy recognized the need for scientists who shared their data to be rewarded, either by receiving scientific credit through acknowledgement, citation or co-authorship, or by receiving financial remuneration (e.g., royalties, potential future grant funding). Porter (2010) contains an example of such a policy.

In 1994, an initial metadata content standard was developed based on metadata conventions and common elements surveyed from each site (Porter interview, 2012). It was later adopted by the IMC (Michener et al, 2011).

Recognizing these Internet-based technology advancements and the growth of the LTER community, the NSF announced a special competition for cross-site comparisons and synthesis. Sites began to compete with one another to build up their capability and fitness for producing shareable data sets, in time for review of their proposals for funding. Using this special competition (solicitation for proposals) the NSF was working on at least two levels: First, it was leveraging its investments in the NSFNET and the super computing centers to stimulate communities like the LTER Network to utilize these resources for research and education. Second, it was providing social engineering of LTER investigators to publish their data sets and make them available online, thereby increasing data set sharing (Porter, 2010). As a result, LTER Governance mandated that each site should have at least one data set available online by the end of the calendar year. By then, there was sufficient buy-in by the community of scientists at each site, and stimulus coming from the NSF, for the Governance committee to take this bold step.

In 1995, 13 of the 18 LTER sites were awarded grants to enhance cross-site comparative research and synthesis. What had started in 1994 as a trickle of online

data sets, turned into a data deluge of online data sets in subsequent years (Porter, 2010). In parallel, there was a growing awareness to the development and implementation of archival metadata standards, as costs of ready-for-science data set production were increasing (Gross et al, 1995). Costs were increasing, because of the absence of a metadata standard for ecological data, resulting in sites using their own conventions. As an example, the Federal Geographic Data Committee (FGDC, 1995) created metadata standards as part of the National Information Infrastructure efforts (Baker et al, 2000). The Future of Long-Term Ecological Data (FLED), a working group of the Ecological Society of America (ESA), also expanded and formalized this early metadata standard of the IMC, then published in 1997 (Michener et al, 1997).

In 1996 the XML markup language was developed. Its adoption for the exchange of data in communities of science was an event for LTER and the National Center for Ecological Analysis and Synthesis (NCEAS), of the Partnership for Biodiversity Informatics (PBI), which then enabled them to adapt the content standard for LTER metadata into a machine and human parsable XML schema (Michener et al, 2011).

By 1997, LTER had adopted a network-wide policy for publishing and sharing their site-based data sets. This LTER Network policy evolved from individual sites' data-access data-sharing policy guidelines, developing since 1990. Michener et al (2011) and Porter (2010) both emphasize the importance of allowing the data policy to evolve over time as a necessary condition to get buy-in from researchers, and to “allow development of an emerging set of ethical principles surrounding data reuse (Michener et al, 2011).” In 1997, the LTER Network expanded to include sites in urban areas and land margins. This led to increasing diversity of the ecological systems in the network’s local and cross-site science and data sets.

In 1998, Windows 98 was released with imbedded browser capability (Robbins, 2011). Additional sites added to the LTER program from the Land Margin proposal competition. Congress also passed legislation for the Next Generation Internet Research Act. The act served to promote investments in the development of

technologies to advance Internet capacity and capabilities. Investments in large-scale networking jumped 14% in 1999, then by 3% into the new century.

In 1999, a Social Science Committee was created to facilitate the integration of social sciences into the LTER. The integration of the social sciences into long-term ecological research was recognized as an urgent priority (Redman et al, 2004). Traditionally, ecological and social scientists had worked in isolation within the boundaries of their disciplines pursuing answers to fundamental questions about pattern and process in the ecological and human world (Low et al, 1999; Redman, 1999a; Redman, 2004). Recognition of this need to integrate brought together social, earth, and life sciences researchers to increase understanding of the human dimensions of ecological change for the LTER network.

In 2000, funding was provided for a project to further develop the work and usability tools of the FLED working group on developing the metadata standard into an XML schema. The project was called The Knowledge Network for Biocomplexity<sup>22</sup> (KNB). KNB was a project of the Partnership for Biodiversity Informatics<sup>23</sup> (PBI). The PBI partnership involved the National Center for Ecological Analysis and Synthesis (NCEAS), the Natural History Museum and Biodiversity Research Center at the University of Kansas, the Network Office of the Long Term Ecological Research (LTER) Network, the San Diego Super Computing Center (SDSC) and the California Institute for Telecommunications and Information Technology (Cal-IT2) at UCSD.

By the end of decade II (1999-2000), access and sharing of data sets significantly increased, largely because of the Internet and its online tools that enhanced searching and discovery of information. While the Internet provided the technological stimulus to increase demand for access and sharing of data sets, it was the sites and the community of scientists' willingness to adopt guidelines, and then later establish policies, for access and sharing of data sets that resulted in notable

---

<sup>22</sup> <http://knb.ecoinformatics.org/index.jsp>

<sup>23</sup> <http://pbi.ecoinformatics.org/projects.html>

increases in cross-site comparative research. Alongside the increased access and sharing of data sets was improvement towards metadata standardization, which resulted from XML becoming adopted for the exchange of heterogeneous data on the Web.

**LTER Network — Decade III (2001 - 2010):**

In 2001, the NSF commissioned a 20-year review of the LTER program. The report contained findings and recommendations to guide the LTER Network into its third decade, or “the decade of synthesis science.” Clearly, data and the processes for its management were the missing links to support a complete vision of synthesis science <sup>24</sup>. The committee recommended that “LTER science” become “multidisciplinary, multidimensional, scalable, information driven, predictive and model based, education oriented, and increasingly virtual and global (Robbins, 2011).”

Another recommendation was made to implement a “systemic information infrastructure.” Reviewers of the LTER program recognized the potential emergence of an infrastructure to automate parts of the process of data reuse. Funding for the development of a “systemic information infrastructure” was provided years later in 2009.

In decade III (by 2002), a stimulus perceived by the LTER scientific community was that scientific inquiry was becoming more multi-disciplinary, driven by global issues. Demand was increasing for the LTER Network to make its primary data more directly usable by other communities (Brunt et al, 2002). The community of scientists responded with a plan to stimulate investment by calling for the development of an active, globally integrated information network. This network

---

<sup>24</sup> Synthesis science in the context of LTER refers to a process in which the data and knowledge gained over the past twenty years of the LTER Network program, plus current studies, are brought together to reach new levels of understanding of long term ecological patterns and processes (LTER White Paper: Priority Setting in the LTER Network, 2001).

would include the capacity to discover, access, interpret and process data for facilitating the integration and synthesis of primary data (Brunt et al, 2002).

In 2003, the LTER formally adopted the Ecological Metadata Language (EML) version 2.0 as the LTER metadata standard. Available database systems were quick to be utilized for the management of metadata (Michener et al, 2011). Demand increased for the management of EML documents. Specialized database systems for the management of XML documents were adapted to manage EML metadata (Michener et al, 2011).

In 2003, LTER organized a mini symposium on the integration of geosciences and social science with the LTER. This was consistent with the 20-year review recommendation that LTER science be more multidisciplinary, partnering with the social science discipline. In subsequent years, LTER organized more mini symposiums: LTER Research information land management (2004); Coastal research in LTER (2005); LTER and Global Change (2006); Cycles of change in socio-ecological systems: perspectives from long-term ecological research; Social-ecological systems in a changing world (2008); Ecological Connectivity in a changing world (2009); Ecosystem services in a changing world (2010).

LTER Network data policy was modified in 2005. It defines the responsibilities of the "Data Collector." The policy dictates how long data access can be restricted (two years after collection). It also identifies special conditions that may allow additional restrictions for a more extended period (e.g., locations of endangered species, human confidentiality). Additionally, the policy outlines the properties of the required metadata.

With a metadata standard established, LTER focused efforts on facilitating discovery of comparable data sets across sites. In 2005, a challenge secondary data users were facing was that data creators were not consistent in the application of keywords they used to characterize datasets (Porter, 2010). This exacerbated the process of discovering data at LTER sites. The LTER Information Management Committee (IMC) established an ad hoc "Controlled Vocabulary Working Group" to

study the problem of uncontrolled use of keywords (Porter, 2010). The outcome of the committee's work will be reported later in this section in the events for 2009.

LTER's Technological and Socio-organizational Infrastructures were developing, as a result of its adoption of technological innovations for the Web and its governance leadership. The change that was fostered stimulated increases in scientific progress. More sites were providing datasets online, and the demand for tools for online exploration was increasing. With guidelines and policies facilitating access to and sharing of datasets, sites began to collaborate in the process of online exploration of datasets. These data sets can be integrated, and then be used for cross-site analysis (Michener et al, 2011).

Discovery and integration of climate and hydrologic datasets from multiple LTER sites and the USDA Forest Research sites was one of the first applications that leveraged the established Technology Infrastructure (based on the Web) and the Socio-organizational Infrastructure (based on data access and data sharing policies). While the sites of each organization provided access to their data and metadata online, researchers found it problematic to locate, access, and assemble data from multiple sites of different organizations (Henshaw et al, 2006). Support was obtained to build an information system called the ClimDB/HydroDB, which leveraged technology, data management and organizational resources from prior investments. The system provided uniform access to common daily streamflow and meteorological data through a single Web portal (Henshaw et al, 2006). The investment in this information system introduced both technological and socio-organizational stimuli towards the development of a "systemic information infrastructure." The technology investment stimulus was considered "an effective bridge technology between older, more rigid data distribution models and modern service-oriented architectures (Henshaw et al, 2006)." The socio-organizational investment stimuli resulted from "scientific interest, organizational and personal commitment, and participation incentives to build ClimDB/HydroDB as an integrated cross-site information product (Henshaw et al, 2006)."



In 2008, as a response to the recommendation of developing a “systemic information infrastructure,” LTER continued enhancing online data exploration and discovery capabilities using new Web services technologies, such as portal technology. Using these technologies, scientists and information managers were then able to integrate heterogeneous long-term data into a common data format, called “derived data.” Using derived data along with new modeling and analysis methods (Henshaw et al, 2006), scientists had access to more tools by which to conduct cross-site comparative research and better understand processes within the Earth’s ecosystems (Servilla et al, 2008).

In 2007, the LTER Cyberinfrastructure Strategic Plan (Robertson et al, 2007) was published. The plan proposed the integration of many of the ideas described in the report on Cyberinfrastructure investments (Atkins et al, 2003), and articulated a vision of how LTER science would be transformed through the use of Cyberinfrastructure. The LTER Cyberinfrastructure strategic plan would provide a blue print for stimulating investments of Cyberinfrastructure technologies and methods in the LTER Network.

In 2009, LTER Network Office received stimulus funding for Cyberinfrastructure development. Funds were approved towards enhancing the Cyberinfrastructure capability of LTER by improving automation of metadata generation and creation of sharable data sets.

In 2010, the end of the third decade of LTER, an e-Infrastructure had been constructed supporting many of the recommendations from the 20-year review of the LTER program. The 20-year review can be characterized as an internal stimulus that came from peers of the biodiversity-ecological research discipline. It established a vision for the third decade as “the decade of synthesis science,” and recommended the implementation of a “systemic information infrastructure” to establish informatics as a core function of LTER for the third decade.

Systems and processes necessary for a systemic information infrastructure were developing. There were significant achievements in the third decade that leveraged ICT investment stimuli:

- Ecological Metadata Language (EML) as an LTER standard for exchanging heterogeneous datasets
- Tools for online data exploration and discovery
- Processes increasing automation of metadata creation, combined with governance policies for data access and sharing,

#### **8.3.4 E-Infrastructure Development Process: Constructing the Data Infrastructure**

Our representation of an e-Infrastructure development process consists of seven interconnected components (Figure 2, in Chapter 2). In the previous section, we described the events in Decade I that led to the enhancement of LTER data with remote sensing data to create geo-referenced data. In this section, we use our conceptual framework to explain the process of enhancing LTER data as an instance of an e-Infrastructure development process.

In our analysis of LTER literature, we found that around circa 1985 the LTER community was under pressure to work with data sets from diverse sources, such as other sites or even from other government agencies. Progress on issues involving cross-site comparative science was lacking. Sites mostly focused on local site-based science (Gholz interview, 2012). The LTER community lacked tools to work effectively in identifying ecological patterns and linking them with corresponding processes (Stafford et al, 1994). Therefore, they explored new technologies that could be applied to enhance progress of scientific issues.

In 1988, an advisory committee on scientific and technology planning provided recommendations to address scientific issues (Shugart et al, 1988). GIS capability was identified as the most urgent technology to add in order to stimulate cross-site science and achieve the LTER community's goal of advancing the state of ecosystem science across the network of LTER sites. The recommendation was a technology

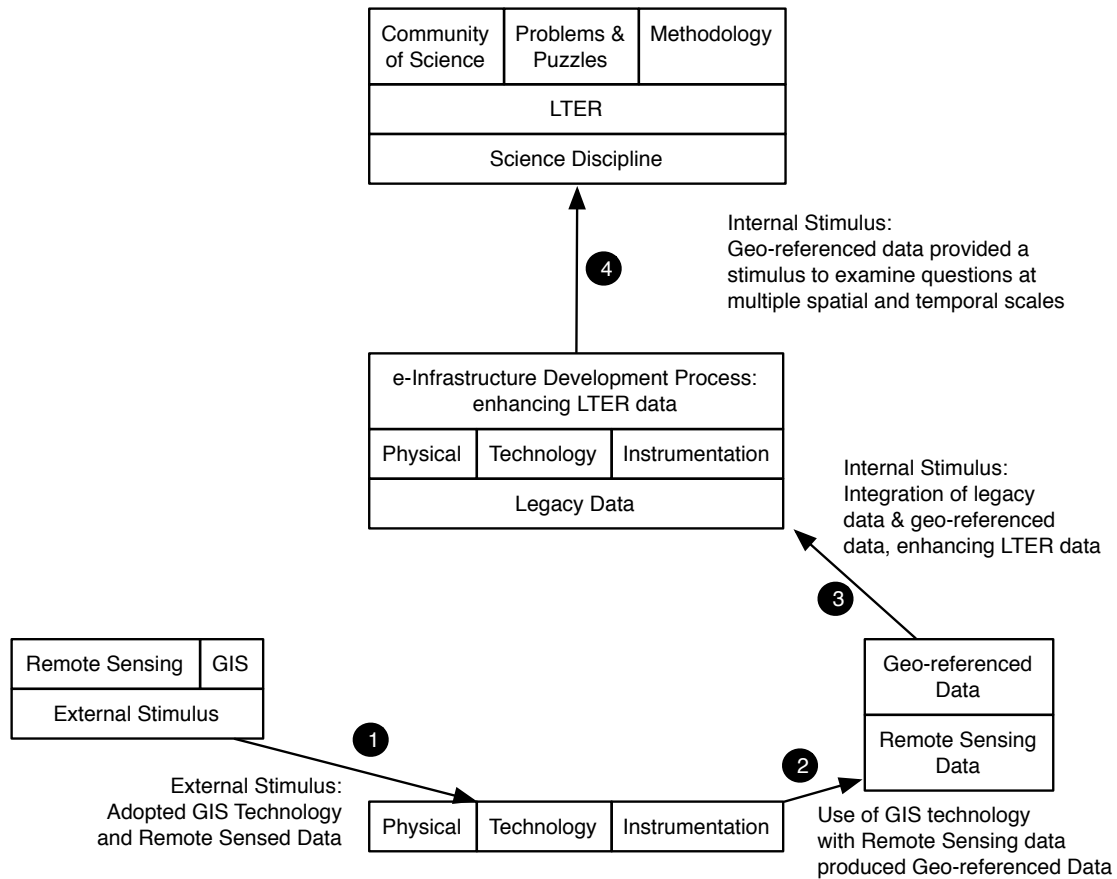
stimulus developed from an external ICT investment. GIS technology and remote sensing data were introduced into the LTER network.

Referring to our conceptual lens (Figure 5), we start our analysis with an external ICT stimulus.

GIS technology and data from remote sensing technologies were adopted as an external ICT stimulus into the LTER community. Physical computer systems with GIS software to process remotely sensed data were deployed at LTER sites. In Figure 10, we show this labeled as (1), which connects the external stimulus with the physical technology. The deployment of the GIS systems, along with the combined use of GIS with remotely sensed data created geo-referenced<sup>25</sup> data in the LTER community (Shugart et al, 1998). In Figure 10, the conceptual lens uses the Technology category to represent the GIS systems. The remotely sensed data is processed through the technology to produce geo-referenced data. This is labeled as step (2) in Figure 10.

---

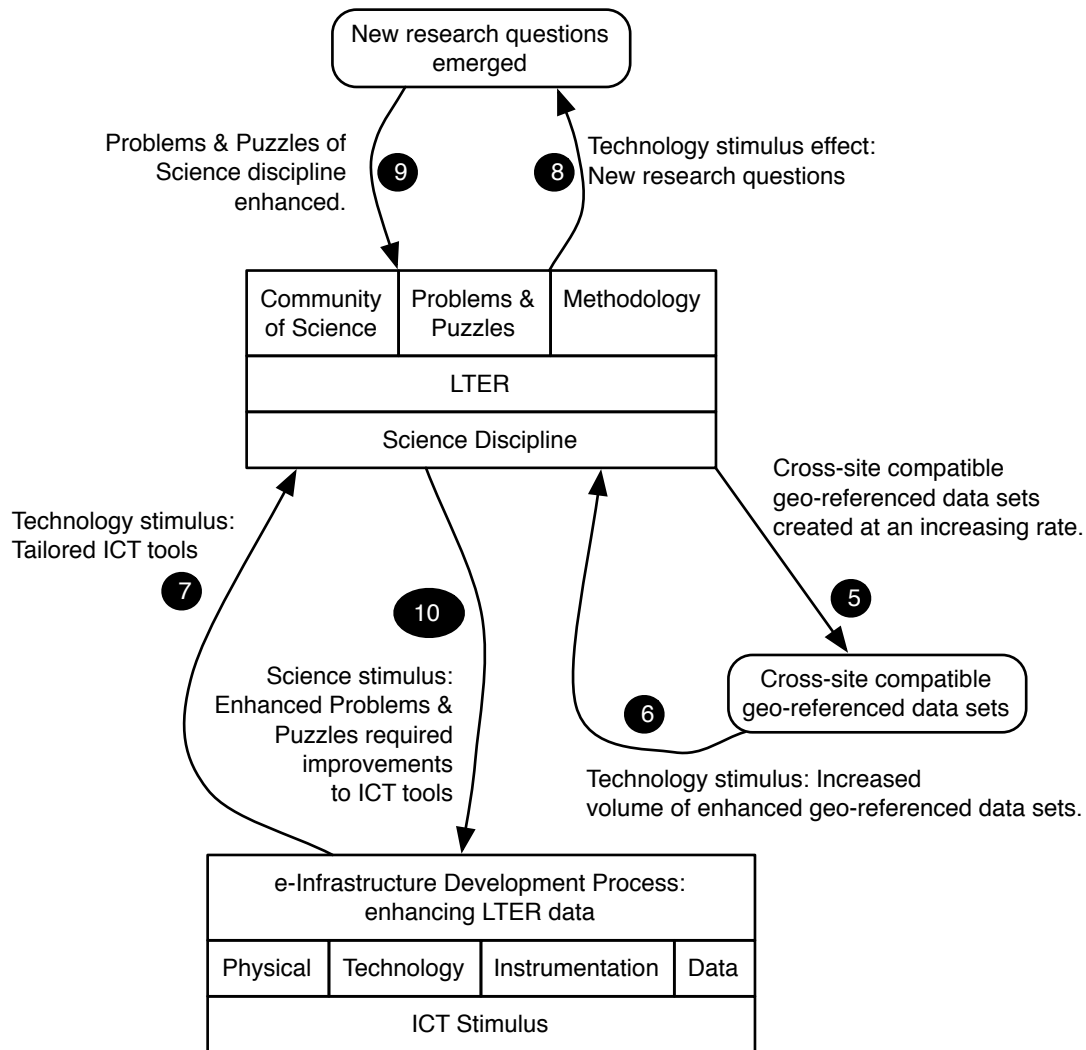
<sup>25</sup> To georeference an object means to define its existence in physical space. Georeferencing provides a locality description that can be used to tag biodiversity occurrences (Hill et al, 2009).



**Figure 10 E-Infrastructure Development Process: Enhancing LTER data, steps 1-4**

LTER scientists and information managers (community of scientists) started to develop tools to integrate their legacy data with new data that was geo-referenced. The integration of geo-referenced data and legacy data enhanced LTER data. Enhancing LTER data is shown in step (3) as an activity of the e-Infrastructure development process that results from linking geo-referenced data and legacy data. Geo-referenced data sets added spatial and temporal resolution to the data, referred to as “data density” (Stafford et al, 1994).

Geo-referenced data sets created an internal ICT stimulus, motivating scientists to examine questions at multiple spatial and temporal scales (Stafford et al, 1994) — step (4) in Figure 10.



**Figure 11 E-Infrastructure Development Process: Enhancing LTER data, steps 5-10**

Cross-site compatible geo-referenced data sets were being created at an increasing rate across the LTER network — step (5) in Figure 11. These sets began to create data sets that were cross-site compatible across the LTER network (Shugart et al, 1988).

Cross-site compatible geo-referenced data sets — step (6) in Figure 11 — were combined with new ICT tools tailored to process the increased volume of enhanced geo-referenced data sets at a faster rate. This created a technology stimulus in the LTER community that combined technology and data — step (7). The effect of this

stimulus was observed as the emergence of new research questions — step (8) in Figure 11. New research questions enhanced the Problems and Puzzles of the science community — shown as step (9).

ICT stimulus in (7) can be either external or internal, because in either, an ICT tool is used with geo-referenced data to address scientific questions. As improvements to ICT tools were required, for example to observe phenomena in geo-referenced data at smaller scales, the LTER community communicated its requirements to the appropriate agent in the e-Infrastructure development process. We show this as step (10) in Figure 11.

Enhancing LTER data has been represented as an e-Infrastructure development process in steps (1) through (10) in Figures 10 and 11, using our conceptual lens. The effects of technology stimuli created from external ICT investments were represented. In the first decade of the program, stimuli from technology played a significant role in the enhancement of LTER data. Data analysis identified the following technology stimuli: GIS technology, technology that combined GIS technology and remotely sensed data to create geo-referenced data sets, and tailored ICT tools combined with cross-site compatible geo-referenced data sets that led to the emergence of new research questions and new Problems and Puzzles.

In this section, we used our case study database (Yin, 2003) and our conceptual lens to explain the stimulus that was applied to enhance scientific data in the LTER Network. This established linkages between the data in our database and our conceptual lens (Walsham, 2006). An external stimulus of data (remote sensing data) and technology (GIS) started a process of e-Infrastructure development that led to enhancing the scientific data of the LTER network. This was a significant achievement in the first decade of the LTER program.

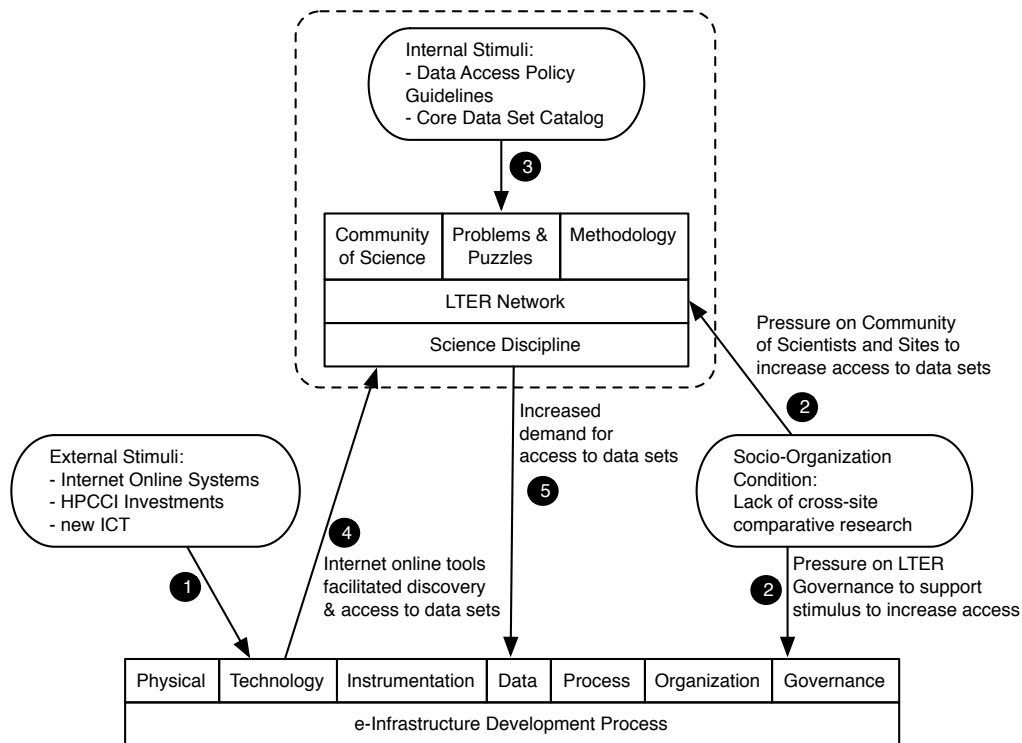
### **8.3.5 E-Infrastructure Development Process: Constructing a Data Infrastructure**

In this section, we continue to explore the e-Infrastructure development process by explaining the evolution of data access in the LTER Network. Data access in the LTER Network was a process that evolved from a confluence of stimuli and a socio-

organizational condition, which led to social engineering by the NSF and the LTER Governance.

We focus our explanation in decade II of the LTER program, which started in 1990. By 1991 several significant technology events occurred that were changing LTER and many science disciplines: Internet online systems (FTP, Gopher, then WWW) were being adopted; investments stimulated by the HPCCI were increasing; and new ICT, resulting from the Internet and increased technology (HPCCI) investments were introducing change and increasing complexity. We represent this as an external stimulus into the e-Infrastructure development process, shown by step (1) in Figure 12.

A socio-organizational condition caused by lack of cross-site comparative research was creating pressure on the community of scientists and the governance of the LTER Network. Pressure from this socio-organizational condition is shown as step (2) in Figure 12 below.



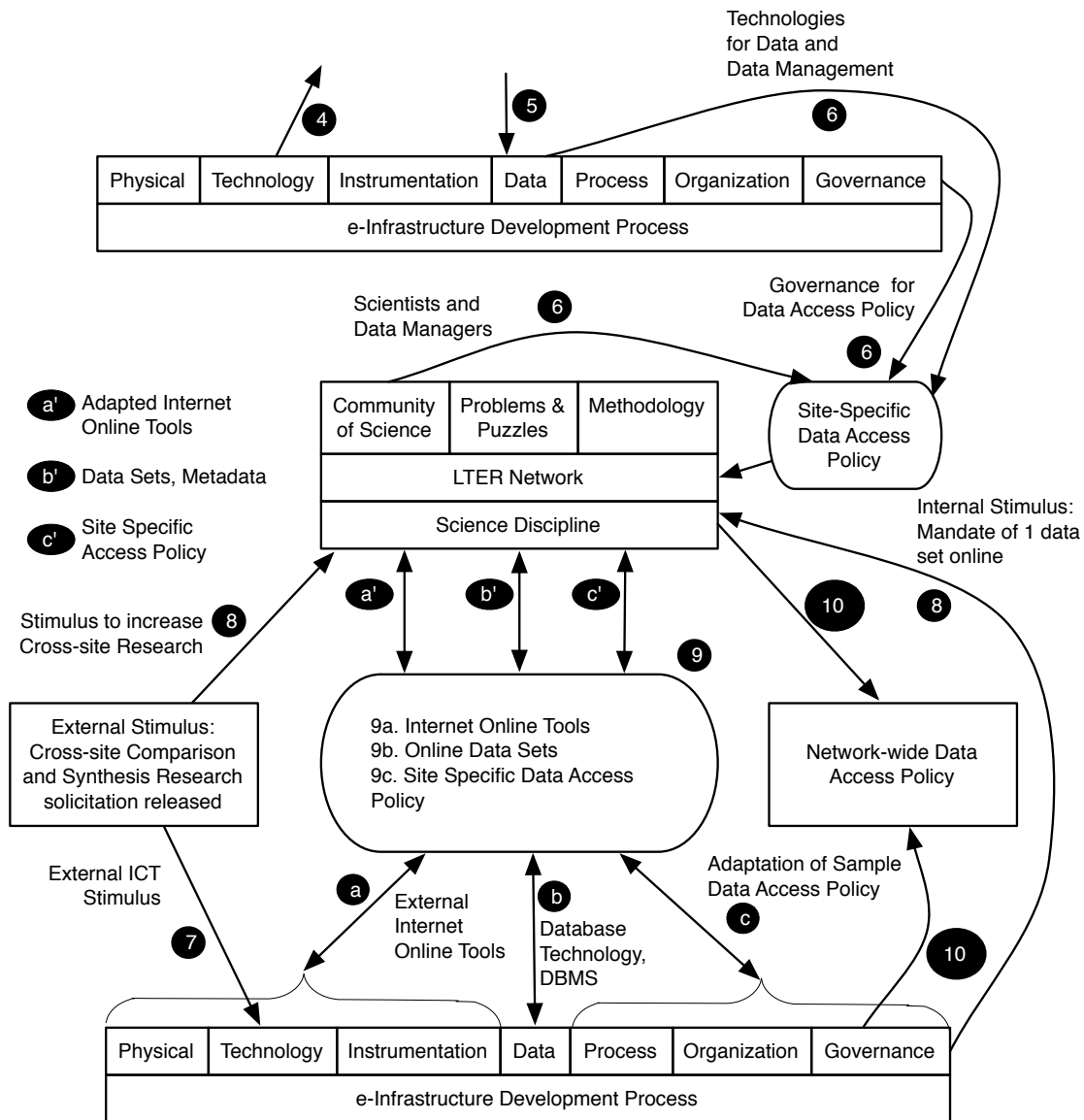
**Figure 12 Access to data sets and e-Infrastructure development process**

The LTER community came to realize two important achievements at the start of decade II: data access policy guidelines and the Core Data Set catalog, previously described in section 8.3.3 about the LTER Network. We characterize these two achievements as an internal stimulus, step (3) in Figure 12, coming from within the LTER Network, because their availability alerted the research community to data resources and their potential to stimulate cross-site questions (Porter, 2010). In Figure 12, we use a dashed polygon to differentiate internal from external stimuli.

Exploration of our data revealed that the data access policy guidelines and the Core Data Set catalog, combined with Internet online tools, facilitated discovery and access to data sets. We represent this outcome as step (4) in Figure 12. Step (4) can be viewed as an extension of the external stimuli, in the form of technology that is adapted for its use within the LTER Network. As sites and scientists from other communities discovered that long-term ecological research data sets were online, demand for access to data sets increased. This condition is represented by step (5), with a directed arrow from the LTER Network to data of the e-Infrastructure development process. This reflects changes to the Data category as a result of increases in access.

Sites adapted the Data Access Policy guidelines into site-specific data-access policies (Porter, 2010). Exploration of the data showed Governance, the LTER community of scientists at each site, Data and Data Managers, representing systems and people, were involved in establishing site-specific data access policies. In Figure 13, we represent this as directed arrows, labeled (6), to the Site-Specific Data Access Policy object.





**Figure 13 Data Access and e-Infrastructure development process**

There was a convergence of Site Specific Data Access Policy, Data Management Guidelines, the Core Data Set catalog, Internet Online Systems, investments supporting HPCCI, and new ICT supporting Internet technologies. This confluence signaled the NSF and LTER Governance to introduce a solicitation, requesting proposals to enhance cross-site comparison and synthesis research. We coded this action by the NSF as an external stimulus, which we represent as an external ICT stimulus, step (7), and a stimulus to increase cross-site research, step (8) in Figure 13. Along with the NSF stimulus, LTER Governance mandated that all sites should

have at least one data set available online coded as an internal stimulus to the LTER Network, shown as step (8).

Exploration of the data revealed events and trends in years 1992 – 1997 that changed the LTER Network. We found these trends and events along three perspectives: technological, data and socio-organizational. Building upon our analysis of the data, and using our Concept Map, we shall now describe events, conditions, trends, etc. that led to the LTER Network establishing a network-wide data access policy.

From a technological perspective, external Internet online tools were becoming increasingly accessible to the LTER Network: the Gopher Information server (1992), the WWW server (1993), Web-based database and information management tools, etc. The LTER community adapted many of these Internet online tools for site-specific or network-wide use: online research summaries (1994); Web-based tools were used to develop an LTER Personnel directory (1995), an automated system for research summaries (1996); a web crawler to harvest then publish climate information (1996); etc. Figure 13 characterizes the technological perspective starting with step (7), an external ICT stimulus, entering the e-Infrastructure development process. Based on data, it emerges as external Internet online tools (described earlier), labeled step (a). The LTER Network adapted external Internet online tools, labeled step (a'). This process is represented as step (9a), Internet Online Tools.

From a data perspective, the number of online data sets was increasing, eventually turning into a data deluge (Porter, 2010). Internet online tools and technologies were a stimulus towards the enhancement of database technologies and database systems (DBMS) for the Internet. The process is shown starting as step (b). LTER sites adopt these Internet-capable DBMS to publish more data sets and their metadata online, shown as step (b'). Representation for the process of increasing online data sets is labeled (9b).

From a socio-organizational perspective, sites had each developed a site-specific data-access policy. The stimulus to increase cross-site comparative research led sites to the publishing of more data sets online. Internet online tools were facilitating the process. Step (c') in Figure 13 represents the process starting from LTER Network sites and their site-specific data-access policy. Processing multiple site-specific data-access policies to access data sets became burdensome. A sample data access policy was adapted, integrating pieces from site-specific policies. Adaptation of the sample data access policy using site-specific data access policies is represented as step (9c).

By 1997, supported by an e-Infrastructure development process, including Internet online tools and increasing online data sets, the sample data access policy had evolved sufficiently, that it was adopted as a network-wide policy for publishing and sharing site-based data sets (Michener et al, 2011). Step (10) represents the adoption of the network-wide data access policy. The adoption of the network-wide data access policy is as far as we go describing the e-Infrastructure development process, in this illustration.

In Figure 13, we have represented the process that led to the development and adoption of an LTER network-wide data access policy. The process started from a set of data access policies specific to each site. The LTER Network had developed sufficiently to tackle socio-organizational issues involving, and technological and social issues concerning metadata development. This will be discussed in the next section.

### **8.3.6 Data Sharing and Governance to Support Data e-Infrastructure**

Prior to the creation of the LTER program, scientists and leaders in the U.S. realized that questions involving global change, conservation, and biodiversity change required data about phenomena measured over long periods of time. “The LTER program was created to enable the study of long-term phenomena that could not be studied effectively over the course of a typical three- or five-year funded project (Robbins, 2010).” Robbins clarifies that “long term” is a relative term with respect

to time, because in the context of ecological research, it could span decades, centuries or millennia (Robbins, 2010). Robbins (2010) reasoned that if the work of LTER is to achieve scientific progress by contributing insights on phenomena spanning multiple decades, or centuries, it will more than likely be from the use of archived data than from published literature. “Thus, the creation and sharing of long-term data sets is clearly an essential part, a *sine qua non*, of the LTER program (Robbins, 2010).”

#### **8.3.6.1**      *Events That Gave Rise to Data Sharing and Its Acceptance in the LTER Network*

In 1992, reuse of LTER data sets required documentation about the data; i.e., metadata. Sites were creating metadata using their own conventions. As a result, it was challenging to process metadata using computers, without first going through a lot of effort to understand a site’s particular metadata (Porter interview, 2012). The LTER Information Management Committee (IMC) funded a project to develop standard ways of exchanging metadata that would be both human and machine-readable (Michener et al, 2011).

In 1994, a process started to collect metadata conventions from each of the sites. This process found common elements, and then compiled them into an initial metadata content standard for LTER (Porter, 2010, personal communication), which was later adopted by the IMC (Michener et al, 2011).

In 1995, the Future of Long-Term Ecological Data (FLED), a working group of the Ecological Society of America (ESA), expanded and formalized this early metadata standard, which was published in 1997 (Michener et al, 1997). In parallel to this development of a metadata standard for LTER, encoding standards were emerging, most notably the eXtensible Markup Language (XML).

In 2000, funding was provided for a project to further develop the work of the FLED working group on the metadata standard into an XML schema and also to develop tools to enhance its usability. The project was called The Knowledge Network for

Biocomplexity (KNB), <http://knb.ecoinformatics.org/index.jsp>. KNB is a project of the Partnership for Biodiversity Informatics (PBI), <http://pbi.ecoinformatics.org/projects.html>

The PBI partnership involved the National Center for Ecological Analysis and Synthesis (NCEAS), the Natural History Museum and Biodiversity Research Center at the University of Kansas, the Network Office of the Long Term Ecological Research (LTER) Network, the San Diego Super Computing Center (SDSC) and the California Institute for Telecommunications and Information Technology (Cal-IT2) at UCSD.

In 2001, the KNB project, started in 2000, developed the metadata content standard into a machine and human parsable XML schema, known as the Ecological Metadata Language (EML) (Michener et al, 2011). EML was sufficiently detailed to allow data users to create automated programs to act on the data (Porter interview, 2012).

In 2003, the LTER formally adopted the Ecological Metadata Language<sup>26</sup> (EML) 2.0 as the LTER metadata standard for the exchange of LTER metadata (Michener et al, 2011).

#### **8.3.6.2**      *Governance Structure of the LTER Network*

The governance structure of the LTER evolved as the Network evolved, in order to satisfy its long-term goals, and to address complex socio-organizational issues, such as data access and data sharing. This section analyzes the changes to the LTER Governance structure over its 30-year history. The analysis reflects the theory of this study that development of technological infrastructure will introduce pressures on aspects of a science discipline. These pressures will result in events and conditions that will motivate a socio-organizational intervention, which may lead to organizational change or policy change. The Governance component in the

---

<sup>26</sup> EML is a metadata specification implemented as a series of XML document types that can be used in a modular and extensible manner to document ecological data, <http://knb.ecoinformatics.org/software/eml/>

conceptual framework tracks such changes in the LTER Network involving intervention by the LTER Governance committee.

A goal early in the LTER program was to achieve comparability of data among projects (Callahan, 1984). It was necessary to communicate and share information, such as sampling of national ecosystems, and tests of regional to national scale hypotheses, on a regular basis among researchers working on different projects. Sites operated autonomously conducting local science. In 1981, at the encouragement of the NSF, a Steering Committee was formed to provide communication and coordination between sites (Callahan, 1984).

Standing Committees in LTER support and inform the governance process. The Climate Committee was formed in 1982. In 1983, the Coordinating Committee replaced the Standing Committee, and the LTER Network Office (LNO) was established in 1983 also. The LNO was created to facilitate and mobilize network science for the overall Network (Gholz interview, 2012). One of the LNO's primary activities was to facilitate cross-site scientific activities. In 1989, a Technology Committee was established as a Standing Committee. They assessed the technology fitness for the production of long-term data sets. This Technology Committee provided recommendations for establishing ICT standards on a per site level (Shugart et al, 1988; Gosz, 1989).

In 1995, an Executive Committee was established. An Executive Board was later formed that replaced the Executive Committee. The Executive Board managed day-to-day governance activities of the Network (Michener et al, 2011). The LTER Network Office is overseen by the Executive Board (Michener et al, 2011).

In 2002, the Network Information System Advisory Committee (NISAC) was formed. Its primary function today is to assure that information management is facilitating ecological research at all levels (Michener et al, 2011).

Several changes occurred in 2007 to the LTER Governance. First, the Information Management Committee (IMC) replaced the Technology Committee. Second, the

Science Council (SC) replaced the Committee on Scientific Initiatives. Third and finally, the Executive Board replaced the Executive Committee. Michener et al (2011) referred to the SC as the “core of the LTER governance system.” The Executive Board includes an elected chair and rotating membership derived from the SC, along with an elected member from the IMC (Michener et al, 2011). Both the SC and IMC consist of a representative from each of the LTER sites, and elect an executive group (IMExec) to conduct day-to-day information management governance activities (Michener et al, 2011). The IMExec and the IMC create ad hoc information management working groups to focus on topics and trends on technologies and practices to enhance LTER information management. Findings are reported back to the IMC at an annual meeting (Michener et al, 2011).

#### 8.4 Findings

We hypothesized at the beginning of the case study that a significant majority of the ICT investment stimuli were allocated towards technological infrastructure development, with the assumption that investments in technology were going to dramatically improve scientific progress. We now use our Concept Map to explore events from decade I (1980 – 1989) in the LTER Network. Next, findings are described, that will lead us to an explanation for revising the conceptual framework.

In decade I of LTER, investments in ICT provided stimuli that resulted in dramatic increases to data volume and complexity, shown as (1) in Figure 14. Analysis of our data showed that investments in remote sensing technologies transformed LTER data so that it could be geo-referenced and used to enhance cross-site comparative science. This led to dramatic increases in volume and complexity of LTER data. Represented as (2) in Figure 14.

For example in (2), investments in microcomputers and Unix workstations increased with advancement in computer and network technologies. In turn, this stimulated investments in GIS technology for microcomputer systems, which provided new technological tools to manipulate and represent geo-referenced data.

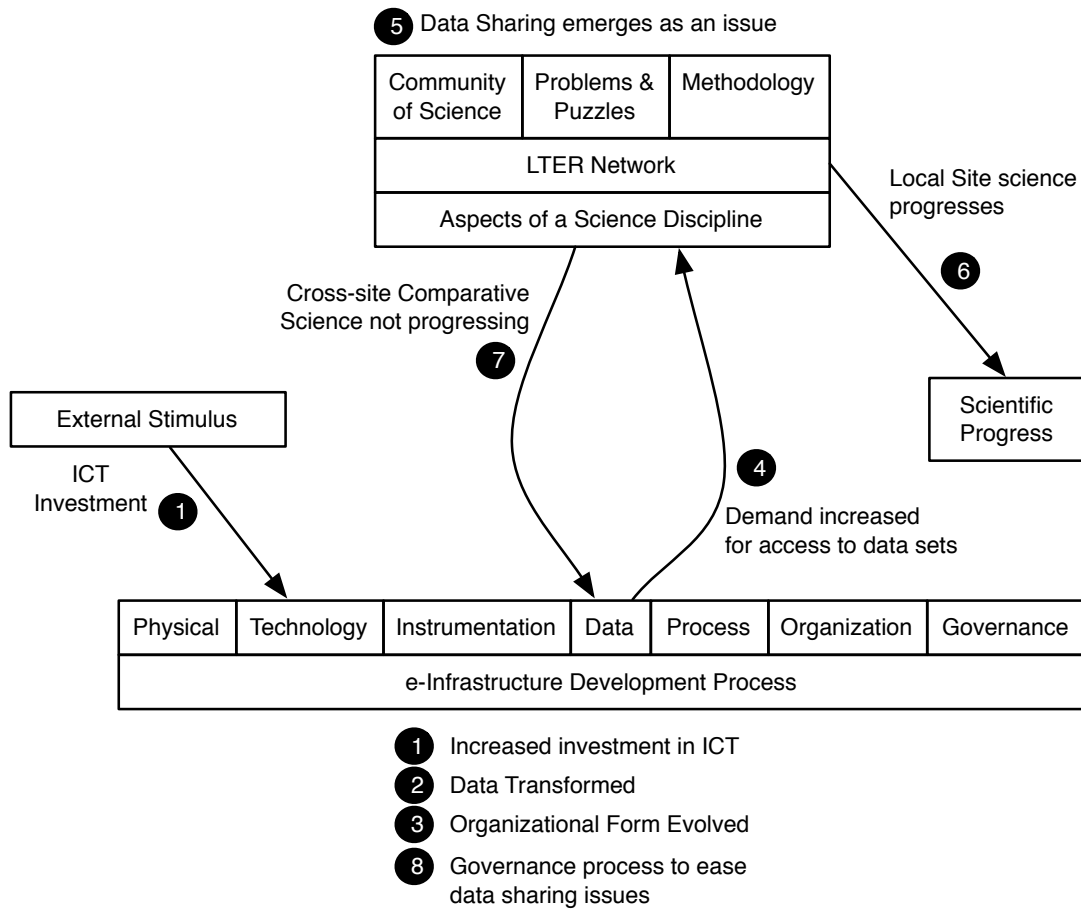
The organizational form of LTER evolved, shown as (3) in Figure 14. Data management processes and the role of data managers changed at LTER sites. The LTER Network Office (LNO) was established as a central entity to facilitate and mobilize network science for the overall Network (Gholz interview, 2012). Nevertheless, the individual sites operated independently and were funded by the NSF based on meritorious science.

Data management was integrated into the research process and rose in importance across the LTER Network, after the influential publication of a workshop volume in 1986 on new methods for data management and the development of metadata (Michener, 1986).

Demand for access to data sets increased, shown as (4) in Figure 14. As geo-referenced data sets increased in numbers, demand increased from within the LTER network and from other communities working with long-term data. Demand for data access raised issues from scientists concerning data sharing, shown as (5) in Figure 14. Data sharing emerged as a barrier to advancing cross-site comparative research. Our analysis showed that data creators were the primary users of the data, and that data was largely used for site-based science. As a result, local site science progressed shown as (6).

Cross-site comparative science, however, was not progressing as intended. Cross-site comparisons were too difficult to execute, because too much effort was required to discover who had what data. We found that a governance process was required to create a “social engineering solution (Robbins, 2011),” to ease the data sharing issues, shown as (7). The governance process chose the development of guidelines, not policy, as an initial step towards bridging the gap between the norms of how data is used by scientists who create it, versus the goals of the LTER program, shown as (8).





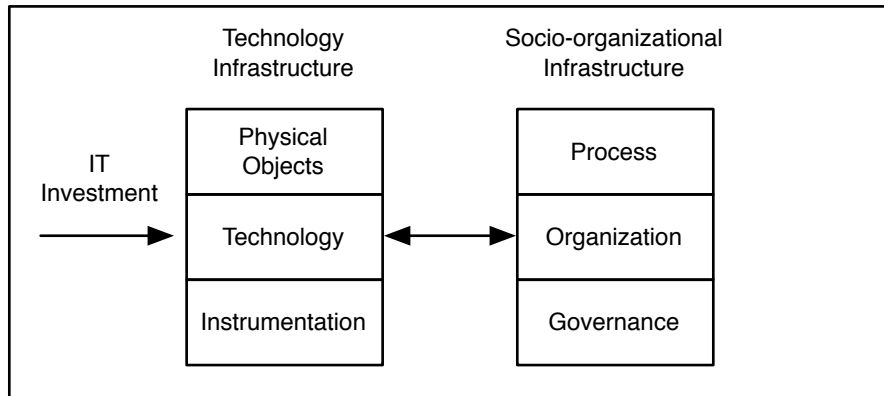
**Figure 14 Concept Map of findings in decade I**

First, using Figure 14 as a lens into our conceptual framework, we generalize that the external ICT investments (1) work as an independent variable.

Second, the e-Infrastructure development process component increased demand for access to data sets. The goal to enhance cross-site comparative research had not progressed. These two conditions were represented as pressure exerted between the e-Infrastructure development component and the Aspects of a Science Discipline component. Based on our analysis of decade I of LTER, we found that increasing ICT investments resulted in dramatic increases in data size and complexity (2).

Third, changes to data, eventually led to changes in processes and organization (3). We generalized those investment stimuli in the development of Technology

Infrastructure led to changes in Data. Eventually, this led to changes in complementary stimuli in the categories for Process, Organization and Governance. As a result, we re-conceptualize the e-Infrastructure development process in the following way, shown in Figure 15:



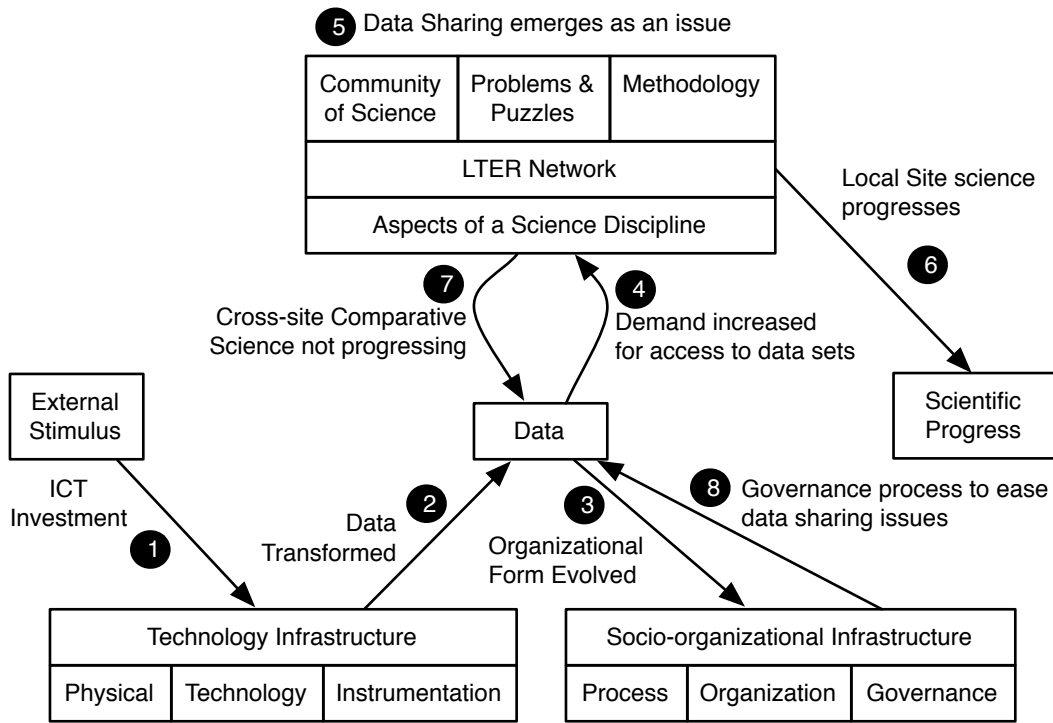
**Figure 15 Complementarities of technology and socio-organizational sides of e-Infrastructure development**

Findings elevated the role of the Data category of the conceptual framework to a mediating<sup>27</sup> variable, between the e-Infrastructure Development component and the Aspects of a Science Discipline component. Findings from the data analysis showed the Data category was present in patterns between e-Infrastructure development and goals of the LTER program, such as increasing cross-site comparative research. Additionally, a similar pattern was observed when pressure between a science discipline and the e-Infrastructure level resulted in a socio-technological intervention. For example, an intervention was introduced into the LTER community to ease issues involving data sharing to stimulate cross-site comparative research. Cross-site comparative research was analyzed as an outcome approaching potential discovery.

<sup>27</sup> A mediating variable describes “how” rather than “when” effects will occur between the independent variable and the dependent variable by accounting for the relationship between the independent variable and the dependent variable (Rudestan and Newton, 2001).

Based on the clustering supported by the data, we shall cluster categories Physical Objects, Technology and Instrumentation into the broader category — Technological Infrastructure. Also, based on the clustering from the data analysis, we shall cluster Process, Organization and Governance into the broader category — Socio-organizational Infrastructure.

Figure 16 shows the clustering we described with the new categories of Technology Infrastructure and Socio-organizational Infrastructure. Also shown are the analytical steps described in Figure 14 transferred to Figure 16.



**Figure 16 Concept Map with clustering of categories for e-Infrastructure Development Process**

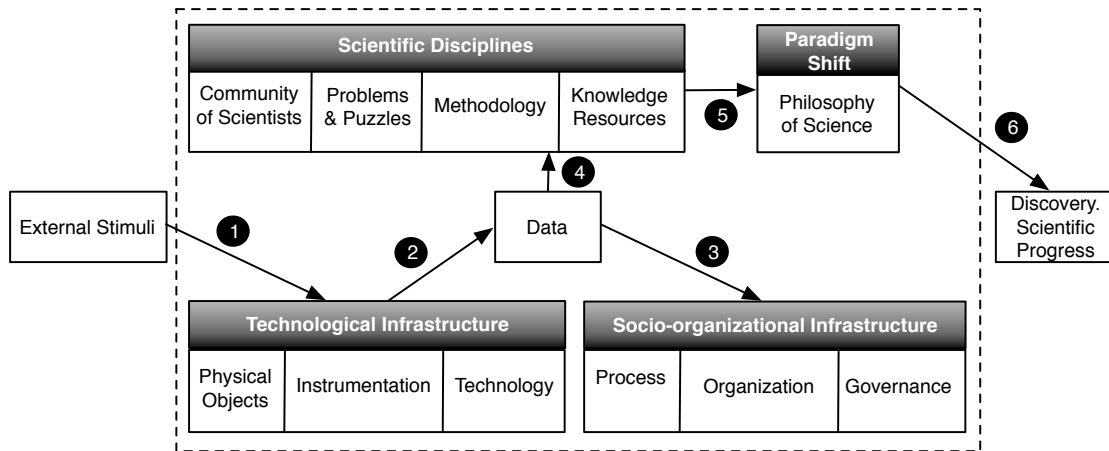
#### 8.4.1 Revision to conceptual framework

We now present a revised Concept Map (Figure 17) of our initial Concept Map (Figure 5). Figure 17 represents what we learned from our analysis and findings. This revised Concept Map represents the e-Infrastructure development process as two major categories: Technology Infrastructure and Socio-organizational Infrastructure. It also promotes the Data category into the role of a linking pin between Technology Infrastructure, Socio-organizational Infrastructure and the

Aspects of Science Discipline components. We characterize this relationship as a “linking pin” to mean how change from Technology Infrastructure affects Data, which then affects the Socio-organizational Infrastructure, and vice versa. Similar movement occurs towards Aspects of a Science Discipline.

The following list corresponds to the numbers in the revised Concept Map. Movement is from left to right, and in order. The descriptions associated with the numbers are inferences that explain the interdependencies between each of the components of the Concept Map.

- (1) Represents that a significant majority of the ICT investment stimuli were allocated towards Technological Infrastructure development.
- (2) As ICT investments increased, it resulted in dramatic changes in data size and complexity.
- (3) As data dramatically changes in size and complexity, it leads to complementary stimuli in processes, organizational form and governance components of a Socio-organizational Infrastructure.
- (4) Data evolves as it mediates stimuli from Technology Infrastructure and Socio-organizational Infrastructure to facilitate sharing of data and to combine data for integrated multidisciplinary research (Brunt et al, 2002).
- (5) New interpretations and synthesis of the data can result in knowledge creation, transformation of theory and paradigms.
- (6) Discovery is achieved, either as revolutionary or evolutionary scientific progress.



**Figure 17 Revised Concept Map: Concept Map 2**

Our analysis of the LTER Network over three decades supports data in this mediating role between Technology Infrastructure development (as a result of external stimuli), Socio-organizational Infrastructure development (as a result of complementary stimuli) and Aspects of a Science Discipline. Recalling, in decade I, LTER data evolved when it was geo-referenced by combining LTER data with remote sensing data. Complementary stimuli occurred in the Socio-organizational Infrastructure category to address data sharing issues. Geo-referenced data was an internal stimulus to the community of scientists to conduct research at multiple spatial and temporal scales.

In decade II, stimuli in Technology Infrastructure development increased with the adoption of the Internet, online data sets, and new search and discovery capabilities. Data's role as a linking pin was manifest. Demand for access and sharing of data sets online increased dramatically. This pushed the socio-organizational side to first develop site-specific access policies, then subsequently into a network-wide policy for sharing data sets. Stimulus to increase cross-site comparative science was a pressure applied to the Science Discipline category from Data and Technology Infrastructure.

Decade III, called "the decade of synthesis science," marked a period when data and the processes for its management was the linking pin between technology, the Aspects of a Science Discipline and Socio-organizational Infrastructure. During this

period, the Technological Infrastructure received stimuli toward the development of Internet technologies, and the Socio-organizational Infrastructure faced an increasing demand for data sharing and improved metadata practices. Another example of data functioning as this kind of linking pin between aspects could be found when scientific inquiry became more multi-disciplinary to address global issues.

#### **8.4.2 Conclusion: Mixed Results**

We set out to find evidence of increases in scientific discovery as a result of stimuli into an e-Infrastructure development process. We find that our results are mixed.

On the one hand, our analysis allows us to build a chain of evidence linking investments in ICT infrastructure development to changes in the practice of science and the scope of scientific inquiry. Increasing cross-site comparative research is just one example of how scientists are able to see deeper and wider into phenomena as a result of increased capabilities provided by e-Infrastructure (Kuhn, 1996; Arthur, 2009).

On the other hand, our chain of evidence does not support our hypothesis that Technology Infrastructure investments are guaranteed to result in dramatic improvements in scientific progress, in particular revolutionary scientific progress. Indeed, scientific discovery, our dependent variable, has not radically changed as a result of e-Infrastructure investments.

In our analysis of the LTER Network, we found that in fact, other fields in adjacent (e.g., evolution, genomics, geology, oceanography, and climatology) and even disparate disciplines (e.g., sociology and economics) were finding the LTER network data useful in answering other sets of questions (Reichman et al, 2011). We infer from our findings that data has begun to cross-disciplinary boundaries. We also found that as data sharing increased between sites, it eventually led to an increase in cross-site comparative research (Johnson et al, 2010).

Thus, if data is being shared between multiple disciplines, such that it enhances comparative research and synthesis, does it not also result in an increase in discovery?

We conclude from this case study of the LTER Network that our findings are mixed concerning the outcome of scientific progress being either evolutionary or revolutionary. Going forward, we want to understand why our results are mixed, because our expectation was that by investing in technology as a stimulus, it would result in revolutionary scientific progress. Furthermore, we present the following proposition to continue the empirical inquiry of this study:

*Combining data from multiple disciplines can lead to increasing discovery.*

Our next step is to return to the literature and to look deeper into what the data is telling us about the outcome of the dependent variables.

## **8.5 Literature Revisited**

In the previous section we found that discovery was not radically changing. In other words, the data did show revolutionary scientific progress that we described earlier as, transformative research. This motivated our next question: Despite all this investment in e-Infrastructure development, why did we not find any clear connection to radical discoveries?

### **8.5.1 IT Productivity Paradox**

We found that the IT Productivity Paradox in the literature has been used to explain similar outcomes involving ICT investment when provided as a stimulus to achieve dramatic productivity improvements (Brynjolfsson and Hitt, 2000).

Productivity is a simple concept. It is a ratio of output produced in relation to inputs required to produce it (Dewan and Kraemer, 1998). Productivity is characterized as “one of the most fundamental measures of business performance (Martinsons and Martinsons, 2002).” Both negative (Roach, 1991; Strassman, 1997; McKinsey and Co., 2002) and positive (Brynjolfsson, 1993; Brynjolfsson and Hitt, 2000) relationships between ICT investment and productivity have been reported.

The IT Productivity Paradox literature says that while there's been massive investments in IT, the rate of productivity, because of IT, has not changed very much (Economist, 1990; Brynjolfsson, 1993). Productivity slowdown was detected in the early 1970s, a phenomenon that coincided with the rapid investment and use of ICT (Brynjolfsson, 1993). The IT Productivity Paradox debate surged when the high expectations from growing ICT investments failed to be reflected in national productivity statistics (Dewan and Kraemer, 1998).

Studies have used the IT Productivity Paradox to assess ICT investments and their impact on productivity both at the organization level and at the national or industry level (Chan, 2000; Park et al, 2007). The scope of our study is ICT investments at both the national level for the development of a national shared e-Infrastructure for science research and education, as well as at the organizational level, such as the LTER network. Therefore, the IT Productivity Paradox provides a framework upon which we can continue to explore the process of e-Infrastructure development and its relationship to scientific progress.

The IT Productivity Paradox literature identified the following four variables to explain how investments in ICT are linked to productivity: Transformation, Complementarities, Transferable ICT products, and Time Lag.

***Transformation:***

Transformation refers to a process brought about by introducing change either at the level of the organization or the industry (nation) through purposeful investments in ICT. Since the focus of this study is at the organization level, we will not discuss transformation at the industry level. See Park et al (2007) for a discussion of transformation at the industry level.

At the level of the organization, changes in work practice, strategy, products, and services are variables used to measure the effectiveness of ICT investment in transforming the organization, supplier and customer relationships (Brynjolfsson and Hitt, 2000). In order to describe transformation in the context of LTER, we will



now look at how purposeful investments in ICT brought about changes in work practice, strategy, products and services

Creating data for secondary use is a persistent goal of the LTER program whose mission it is to study long-term ecological phenomena through site and cross-site comparative research and synthesis. This goal has attracted repeated ICT investments across all three decades of the LTER program (Michener et al, 2011). Work practices have changed from the adoption of ICT to achieve increases in data access and data sharing. Strategy evolved towards increasing cross-site comparative research by leveraging prior ICT investments; for example, LTER leveraged the Internet (investments at the national level) to transform the process of searching and discovering information about data. Products of the LTER are data sets and associated metadata for secondary use. Increasing demand for ready-for-science data products drove ICT investments and changes in publishing and sharing policies. Services in LTER refers to useful<sup>28</sup> ready-for-science data products for secondary use by data users.

***Complementarities:***

Complementarities refer to an organization's ability to leverage ICT investments to create improved work practices or business processes. The value of any ICT investment is measured by its positive impact on one or more aspects of an organization (Brynjolfsson and Hitt, 2000). Complementarities have been used as intermediate variables or dependent variables to measure linkages between ICT investments and business value (Brynjolfsson and Hitt, 2000).

Complementary innovations or complementary organizational investments refer to subsequent investments that leverage ICT investments (inputs), enabling organizations to increase output and leading to increases in productivity. For example, investment in the electric motor provided industrial engineers more flexibility in the placement of machinery in factories. Eventually, this led to

---

<sup>28</sup> Baker et al (2000) described useful data as a "known quality that is well described with metadata."

complementary organizational investments in workflow redesign, which dramatically improved manufacturing productivity (David, 1990; Brynjolfsson and Hitt, 2000).

Complementarities were used to predict the outcome of investment in ICT (Hitt and Brynjolfsson, 1997). In general, the absence of complementarities created by an organization or nation will indicate a decreasing return on its ICT investments. Park et al (2007) linked a nation's capability to exploit its pre-existing intellectual capacity, and leverage its ICT investments as an indicator of a positive return on its investments.

Our revised Concept Map in Figure 18 links changes in data to complementary investments. This link is evidenced by enhanced scientific processes, the creation of new organizational forms, or changes in governance structures. Referring to our revised Concept Map, we can characterize the concept of complementarities in the first three inferences, in the following way: (1) Investments in Technology Infrastructure lead to changes in data, such as in volume and complexity; (2) Changes in data bring about conditions that lead to complementary investments to change practices (methods) or processes (within discipline or across multiple disciplines); (3) The combination of outputs from ICT investments, complementary innovations of enhanced practices or processes, and data, leads to more innovation. This complementary innovation provides value to a science discipline, in the form of new advancements in data sharing and cross-site comparative research and synthesis. Brynjolfsson and Hitt (2000) referred to these interactions as a pattern of complementary innovations.

***Transferable IT Products:***

Transferable IT products establish a relationship between the successful transfer of IT products and technological innovation. Recent empirical studies, using data of ICT investment at the national level, have identified a positive correlation between increased IT investment and productivity growth (Dewan and Kraemer, 2000; Park et al, 2007). These studies use the successful transfer of IT products and their

adoption as conditions by which to measure enhancements to production efficiency and global competitiveness at the national level (Park et al, 2007).

We adapt the concept of “transferable IT products” to “transferable data products.” This then begs the question: Is there a positive correlation between “transferable data products” and scientific progress (productivity)? We would argue that such a positive correlation is present. What does our data say about transferable data products in LTER? By decade III, LTER was leveraging Technology Infrastructure and socio-organizational complementarities enabling scientists and information managers to integrate heterogeneous long-term data into derived data products. Derived data products used a common data format that can be combined with complementary innovations consisting of other knowledge sources, new ICT, and approaches to promote new interpretations and synthesis of the data (Peters, 2010).

***Time Lag:***

Time lag is one of four explanations given for the IT Productivity Paradox (Brynjolfsson, 1993). Time lag basically says that the benefits from ICT investments can take several years to show positive results. An econometric study by Brynjolfsson et al (1994) found lags of two to three years before the strongest organizational impacts of ICT were detected. "In general, while the benefits from investment in infrastructure can be large, they are indirect and often not immediate (Brynjolfsson, 1993)."

Time lag can also occur when there's an issue present that calls for an active social engineering solution. The absence of such an active social engineering solution can impede the adoption of ICT innovation and its complementarities. LTER experienced this condition when comparative research was not progressing due to scientists' unwillingness to share data sets they had created. A time lag of almost ten years occurred before scientists adopted the processes and the ICT tools to share data sets online.

### 8.5.2 Impact of ICT on Scientists' Productivity

Advancements in Technology Infrastructure have changed the ways in which scientists work and conduct research. Arguably, the impact of ICT investments plays a particularly important role in the production of knowledge, given that scientific inquiry is highly dependent on instrumentation, physical materials, knowledge and human resources. Access to these resources is greatly enhanced by ICT. Studies on the relationship between ICT and research productivity generally have found support for the view that ICT enhances productivity in several scientific disciplines, such as oceanographic science (Hesse et al, 1993), life science (Winkler et al, 2010; Ding et al, 2010), philosophy, political science and sociology (Cohen, 1996; Walsh et al, 2000). ICT has impact on scientific productivity by enhancing collaboration and democratization among communities of scientists. This is consistent with the ICT impact on organizations, industries and even nations.

#### ***Enhancing Collaboration:***

ICT offers enhanced collaboration and connectivity among scientists. When combined with socio-organizational complementary innovations, scientists can be supported to achieve higher levels of productivity. The mission of the LTER network was to address long-term ecological phenomena through research at individual sites, as well as comparative and synthetic activities among those sites.

Johnson et al (2010) assessed how the LTER achieved its mission using intersite publications as the measure of collaboration. They recognized characteristics of the LTER mission promote intersite co-authorship: cross-site measurements and comparisons (Hobbie, 2003; Redman and Foster, 2008), information technology transfer (Porter et al. 2005; Brunt and Michener, 2009), documentation of methodologies (Robertson et al., 1999; Greenland et al., 2003; Fahey and Knapp, 2007), and synthesis of ecological concepts (Peters, 2008). The study by Johnson et al (2010) showed over 26 years of the LTER program that the research collaboration efforts of LTER scientists expanded from site-specific studies (no intersite publications) to the production of numerous intersite publications.

***Democratization:***

ICT can also produce a “democratizing” effect, which can benefit underrepresented groups (e.g., scientists at lower tier institutions) thereby leveling the research “playing field.” Several studies using longitudinal data illustrate how lower tier institutions and scientists at lower-tier institutions benefitted relatively more from ICT investments (Agrawal and Goldfarb, 2008; Murray et al., 2008; Furman and Stern, 2009; Winkler et al., 2010; Ding et al., 2010). IT has an equalizing force, providing a greater boost to productivity and more collaboration opportunities for scientists who are more marginally positioned in academe.

The International LTER (ILTER) network was formed in 1993 to facilitate communication and information exchange between international sites conducting comparative research. The ILTER program proposed to facilitate development of LTER-type programs where they do not exist, providing scientists with the opportunity to collaborate and have access to data and knowledge resources. One of the benefits of the ILTER network is the opportunity it provides to all participating sites, regardless of economic status or ranking in the global science community, to evaluate different approaches to interdisciplinary science (Hobbie et al, 2003).

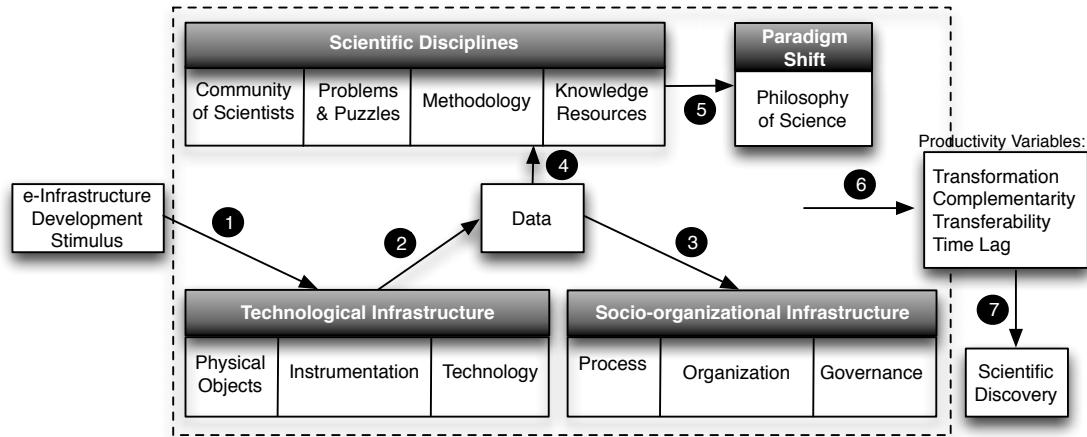
In summary, we examined the literature on the IT Productivity Paradox. We found four intermediate variables of impact that allow us to establish a relationship between ICT investments and scientific progress. Findings were the presented on the four intermediate variables and impacts on enhancing collaboration and democratization in the context of the LTER Network.

We then analyzed the IT Productivity Paradox variables with the data from the LTER Network case study. We adapted the primary research question by changing “scientific discovery” to “scientific progress.” This reframed the primary research to:

*How is investment in e-Infrastructure development impacting scientific progress?*

The scope of our inquiry becomes the impact of e-Infrastructure development on scientific progress. We use the IT Productivity Paradox variables to observe patterns of productivity.

We again revise our Concept Map to incorporate these intermediate variables of impact by which to measure scientific progress derived from ICT investment stimuli.



**Figure 18 Revised Concept Map with Productivity variables: Concept Map 3**

The scientific goals of LTER are to produce excellent site-specific research, and produce sharable data sets to enhance cross-site collaborative research. Findings indicated that scientific progress is more nuanced than either evolutionary or revolutionary progress. Scientific productivity, using concepts of the IT Productivity Paradox, was introduced to measure the impact of ICT investments and their impact on e-infrastructure development and scientific progress.

## 9. Genomics Case

### 9.1 Introduction

Upon completion of the biodiversity discipline case study, we concluded with the following proposition:

*Combining data from multiple disciplines can lead to increasing discovery.*

In this chapter, the scope of our inquiry returns to scientific discovery in order to inquire about data use across multiple disciplines and its impact on scientific discovery.

What disciplines were already combining data with the biodiversity discipline? From our exploration of biodiversity we had hints that genomics was a potential discipline that was sharing data with biodiversity. However, we were looking for a discipline whose area of exploration overlapped with biodiversity, but approached it from a micro level. We recognized biodiversity as being a macro-level discipline.

*What do we mean by these categories of macro-level discipline and micro-level discipline?*

The idea came to us when exploring the relationship between particle physics and astronomy. In very simple terms, both disciplines aim to understand the nature of the Universe, but through different approaches: Astronomers study celestial objects and how they interact with each other, while particle physicists study particles and how they interact with other particles. The discovery of dark energy and dark matter created an intersection for both disciplines to share data. Through its data, astronomy approached this intersection from a macro-level perspective. Conversely, particle physics, through its data, approached this intersection from a micro-level perspective. This is why we conceived categorizing disciplines as macro-level and micro-level. The key factor for the macro-micro level relationship is the presence of an intersection between the two disciplines that supports data sharing.

It is our conjecture that the sharing and combining of data across multiple disciplines will increase or influence discovery.

*What are the properties of biodiversity that make it a macro-level discipline?*

Biodiversity is referred to as a macro-level discipline, because it is the study and classification of organisms. Organisms are unique living things. All living things are unique, because the specifiable properties of individual living things are determined in large part by the particular, frequently contingent historical events that happened to each of their unique ancestors (Robbins, 1996; Fitzhugh, 2006).

Robbins (1996) argues that a characteristic of the micro level that differentiates it from the macro-level is that objects of interest are interchangeable. In other words, objects, such as atoms, particles, electrons, quarks, etc., can be combined and interchanged to form other objects of interest. Organisms, on the other hand, are unique; therefore, not interchangeable.

*What neighboring discipline fits as a micro-level discipline that we can connect with the biodiversity discipline?*

We argue that genomics fits as a micro-level discipline opposite of biodiversity for the following reasons. First, genomics aims to answer similar questions or seeks to understand the same or similar phenomena as biodiversity, but uses different methods and works at the micro level. Second, genomics is the study of the genomes of organisms. A genome<sup>29</sup> is an organism's complete set of DNA, including all of its genes. Each genome contains all of the information needed to build and maintain that organism. DNA<sup>30</sup>, or deoxyribonucleic acid, is the hereditary material in humans and almost all other organisms. Nearly every cell in a person's body has the same DNA. The information in DNA is stored as a code made up of four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T). The order, or sequence, of these bases determines the information available for building and

---

<sup>29</sup> <http://ghr.nlm.nih.gov/handbook/hgp/genome>

<sup>30</sup> <http://ghr.nlm.nih.gov/handbook/basics/dna>



maintaining an organism, similar to the way in which letters of the alphabet appear in a certain order to form words and sentences. DNA bases are interchangeable. Very simply, we can refer to genomics as a micro-level discipline, complementary to biodiversity (it's macro-level discipline), where both involve the study of species identification and discovery.

## **9.2 Confirming Genomics as a Micro-level Discipline Connected to Biodiversity**

We want to find evidence to support our claim that genomics fits as a micro-level discipline connected to the biodiversity discipline. Our conjecture is if biodiversity scientists are using data from genomics and combining it with their data, then it could lead to increasing discovery. To motivate the inquiry, the following research question is posed:

*How is genomics data combined with biodiversity data to produce a result that leads to increasing discovery?*

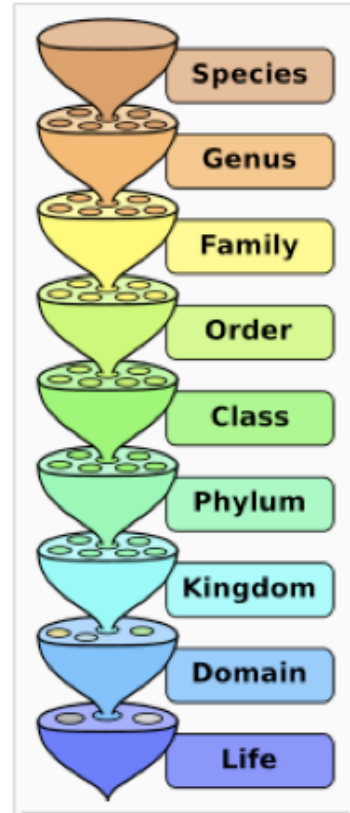
We conducted inquiry on the use of genomics data by biodiversity scientists to confirm biodiversity propositions. The inquiry conducted is supported by two sources of information: (1) peer reviewed papers and reports, and (2) informant interviews.

### **9.2.1 Species Classification and Discovery**

Broadly speaking, biodiversity is concerned with the identification and classification of organisms through examining the variation of life within biological organization (Gaston and Spicer, 2004). They use classification to structure information about organic diversity and make it accessible (Marcus, 1993). "A classification is a division of objects into groups, where the groups have been given names and their distinctive properties stated (Pankhurst, 1993)." Classification orders organisms into groups, reflecting their relationships to other organisms (Sneath and Sokal, 1973). Carl Linnaeus, in the middle eighteenth century, developed a system of classification that continues to be used today, based on a set of ordered ranks: species, genus, family, class and phylum.

***Systematics:*** Systematic Biology (herein “Systematics”) is concerned with the discovery and identification of the diversity of living and fossil organisms (Marcus, 1993). Systematics research is considered the classic way of conducting research in situ in a museum working with preserved collections (Feldman et al, 1992).

Systematists study the relationships among living organisms through time to understand an organism's relationships with other living organisms. “Systematists analyze variation among organisms, patterns of shared common ancestry, and the evolutionary processes that gave rise to both diversity and evolutionary patterns (Chernoff, 1986).”



**Figure 19 Hierarchy of biological classification of the eight major taxonomic ranks**

***Taxonomy:*** Biodiversity scientists use the theory and practice of taxonomy to classify biological diversity. Taxonomy is concerned with the theory and practice of the classification of biological diversity (Chernoff, 1986). This involves formally describing, identifying, classifying, and naming organisms. The core mission of taxonomists is to inventory the species diversity of the globe, to produce a predictive classification of life, and to organize this information into an efficiently retrievable form (Claridge, 2005).

Genomics data is used for species classification and discovery. Taxonomists can use genomic data, such as DNA barcoding data for species classification and discovery (Savolainen et al, 2005).

What is DNA sequencing? DNA sequencing is a laboratory technique used to determine the exact sequence of bases (A, C, G, and T) in a DNA molecule. DNA sequence information is important to scientists investigating the functions of genes. For biodiversity scientists, it was important to know if it would be possible to

distinguish a large number of species using short DNA sequence data (Savolainen et al, 2005).

What is DNA barcoding? DNA barcoding is a diagnostic technique in which a short DNA sequence can be used for species identification (Savolainen et al, 2005).

In 2003, researchers at the University of Guelph in Ontario, Canada, proposed “DNA barcoding” as a way to identify species. Barcoding uses a very short genetic sequence from a standard part of the genome the way a supermarket scanner distinguishes products using the black stripes of the Universal Product Code. DNA barcoding addresses only a limited aspect of the taxonomic process, by matching DNA sequences to “known” species (Savolainen et al, 2005).

DNA barcoding<sup>31</sup> provides biodiversity scientists with a new methodology to accelerate the process of species identification and classification. It increases progress of traditional taxonomic work (Gregory, 2005). Barcodes have the role of filling gaps in species classification information by providing a tool to assign unidentified specimens to already characterized species (Hebert and Gregory, 2005; Hebert et al, 2003a; 2003b).

### 9.2.2 Case Example: Connecting Genomics to Biodiversity via Data

We have established biodiversity as a macro-level discipline and genomics as a neighboring discipline whose focus is at a micro-level.

Our next step is to find evidence that supports the proposition we made earlier. We restate the proposition in the following way:

*As data from multiple fields are combined, scientists can use them across multiple disciplines to increase discovery.*

The objective of this case study is to show how genomics is connected to biodiversity. We will accomplish this by providing empirical evidence, showing that biodiversity scientists use genomics data to confirm biodiversity propositions.

---

<sup>31</sup> <http://barcoding.si.edu/whatis.html>

### **Case Example 1: Biodiversity scientists using genomics data to confirm biodiversity propositions.**

DNA data and barcoding are enhancing the process of classification and discovery in biodiversity research. Traditional methods in taxonomy for classification and discovery of morphology are similar. However, genetically differentiated species called “cryptic species,” have a history of lacking the needed information to classify these cryptic species into the appropriate evolutionary context (Hebert and Gregory, 2005). DNA is useful in identifying evolutionary relationships, such as clades. Discovering cryptic species and filling gaps in biodiversity inventories is recognized as a significant contribution DNA data and barcoding has made to biodiversity research (Blaxter, 2003; Savolainen et al, 2005).

DNA barcoding has successfully identified cryptic species in their natural habitats in marine organisms (Shander & Willassen 2005), including fishes (Mason, 2003; Ward et al., 2005), soil meiofauna (Blaxter et al., 1998; 2004), freshwater meiobenthos (Markmann & Tautz, 2005) and extinct birds (Lambert et al., 2005). In rainforests, rapid DNA-based entomological inventories have been performed so efficiently (Janzen et al., 2005; Monaghan et al., 2005; Smith et al., 2005) that scientists working tropical habitats have been among the most active advocates of DNA barcoding (Janzen, 2004).

Major habitats of biodiversity exist in tropical moist forest regions (Wilson, 1988). Loss of habitat is particularly important at national and local levels, because most ecosystem services are delivered at the local and regional level and strongly depend on the type and relative abundance of species (Duraiappah and Naeem, 2005). The effectiveness of DNA barcoding for the identification and discovery of specimens was tested in a species-rich tropical habitat located at Area de Conservación Guanacaste (ACG) in northwestern Costa Rica. Inquiry was performed to determine whether DNA barcodes provide sufficient resolution to identify specimens in three families of butterflies. Habitats at the ACG site in Costa Rica had been much studied

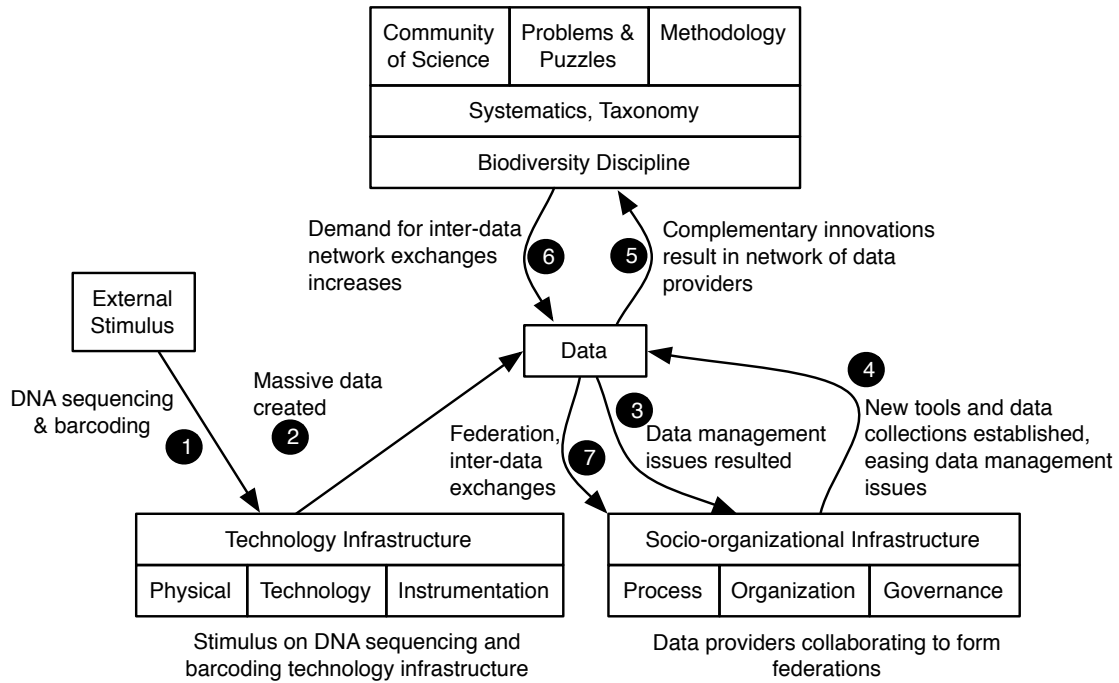
taxonomically for at least two centuries; therefore, the site provided a template against which to test the accuracy of DNA barcoding.

### **Case Example 2: Establishing interconnections between biodiversity and genomics mediated through data**

In this second example, we will use our revised Concept Map to interpret data we gathered.

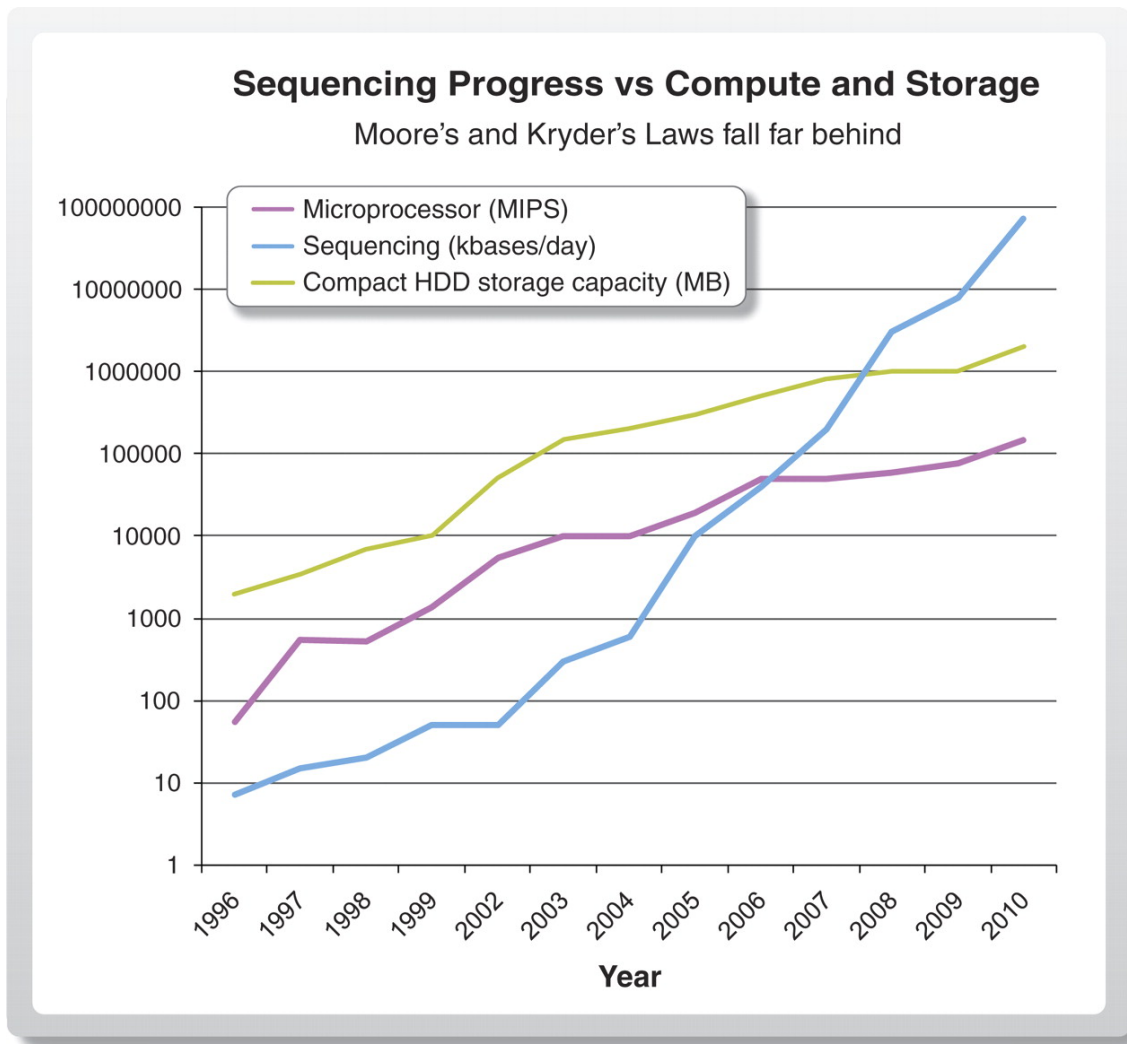
We will explain the way in which data from genomics is being combined and shared with biodiversity data through the use of DNA sequencing. By linking data to our Concept Map, we will describe the connection between genomics and biodiversity.

Investments in DNA sequencing and barcoding technologies provided an external stimulus — shown in Figure 20 as step (1). DNA sequencing and barcoding Technology Infrastructure are evolving. Computational capacity of DNA sequencing has increased dramatically since the success of the Human Genome Project (Shendure and Hanlee, 2008). Improvements in DNA sequencing technology have driven costs down and demand up. DNA sequencing costs have been reduced by several orders of magnitude (Shendure and Hanlee, 2008). High-performance DNA sequencers are now commercially available. Eid et al (2009) demonstrated an increase in the speed of the sequencing cycle by approximately four orders of magnitude. Kahn (2011) reported that there are at least 20 major sequencing labs worldwide that have each deployed more than 10 sequencers.



**Figure 20 Representation of connections between biodiversity and genomics using revised Concept Map**

*Technology Infrastructure to Data:* As investments in DNA sequencing technology increased and these sequencing systems were adopted, biologists were able to generate new data for organism classification and discovery at dramatic rates and volume — shown as step (2) in Figure 20. Referring to the conceptual framework, niches of data (Population Ecology and Resource Dependence theories, Chapter 3) were being created in the environments of genomics and biodiversity disciplines. Biodiversity and other subfields were pressured, facing whether they could deduce from this torrent of molecular data how systems and whole organisms work. “All this information needs to be sifted, organized, compiled, and — most importantly — connected in a way that enables researchers to make predictions based on general principles (Pennisi, 2005).” The rate of raw data output from next-generation DNA sequencers had surpassed Moore’s law for information technology and storage capacity (see Figure 21). From the level of e-Infrastructure development, the volume of data was generating pressures in the environment, enabled by the technology and instrumentation infrastructure.



**Figure 21 from S D Kahn, Science 2001; 331:728-729**

*Data to Socio-organizational Infrastructure:* Many of the tools utilized to process DNA sequenced data and integrate it with other data sources were inadequate, and simply unable to scale with the volume and complexity of the data (Science, 2011). Many of these tools had been developed at individual labs or as part of a project. Often they were not supported beyond the life of the grant that funded them. An imbalance had emerged that caused mutual pressures between the e-Infrastructure level and biodiversity discipline. Stakeholders introduced a socio-organizational intervention into the environment as a way to address the imbalance. Eventually, this imbalance between data and the tools necessary to process it and manage it, created opportunities for complementary innovation by investments in the

development of new organizations supported by new processes and technological systems, shown as step (3). Population Ecology and Resource Dependence theories (Chapter 3) guide our interpretation. Investments created the conditions for resource providers to be formed and to enter the environment to establish linkages between the level of the science discipline and the level of the e-Infrastructure. Resource providers addressed the imbalance by creating linkages to niches between data users and data producers. Referring to the concept of Complementarities of the IT Productivity Paradox, the intention of the investment and the socio-organizational arrangement between organizations at the science discipline level and resource providers was to produce complementary innovations, such as ready-for-science data products. Ready-for-science data products can be interpreted as an output of the Complementarities concept, enabling the science discipline to increase productivity.

Two examples are the iPlant Collaborative at the University of Texas Austin, and the National Institute for Mathematical and Biological Synthesis (NIMBioS) at University of Tennessee, Knoxville. The iPlant<sup>32</sup> collaborative develops tools and Cyberinfrastructure for the plant sciences by leveraging new computational science and Cyberinfrastructure solutions. iPlant enables multidisciplinary teams to address grand challenges in the plant sciences. NIMBioS<sup>33</sup> combines new mathematical methods and computational approaches to develop new tools to find patterns in growing heterogeneous biological data and evaluate hypothesis to address challenges linked to natural and social systems.

*Socio-organizational Infrastructure to Data:* The following regional and nationally scoped initiatives have provided new tools and established large domain-specific data collections, easing the data management issues (Reichman et al, 2011) — shown as step (4): iPlant, NIMBioS Global Biodiversity Information Facility specimen records (Chavan et al, 2010), the Knowledge Network for Biocomplexity<sup>34</sup>

---

<sup>32</sup> <http://www.iplantcollaborative.org/about/project-overview>

<sup>33</sup> <http://www.nimbios.org/about/>

<sup>34</sup> [The Knowledge Network for Biocomplexity, http://knb.ecoinformatics.org](http://knb.ecoinformatics.org)



(Andelman et al, 2004), the Dryad repository (White et al, 2008), and the National Biological Information Infrastructure Metadata Clearinghouse (San Gil et al, 2010).

*Data Infrastructure evolves as it mediates complementary innovations of Socio-organizational Infrastructure and Technology Infrastructure:* Complementary innovations that leveraged technology and improved data management infrastructure have resulted in a network of “data providers” serving their communities of interest (Foster, 2005; Reichman et al, 2011). Data providers have domain knowledge, and over time, have developed expertise in working with systems to produce data products as a service to their communities. “Data networks” have emerged between data providers and data users (Michener et al, 2011; Servilla et al, 2008). Data providers facilitated the sharing and combining of heterogeneous data between biology and its subdisciplines, and the collaboration of disciplines in earth and life sciences, social sciences and humanities (Reichman et al, 2011). See step (5) shown in Figure 20.

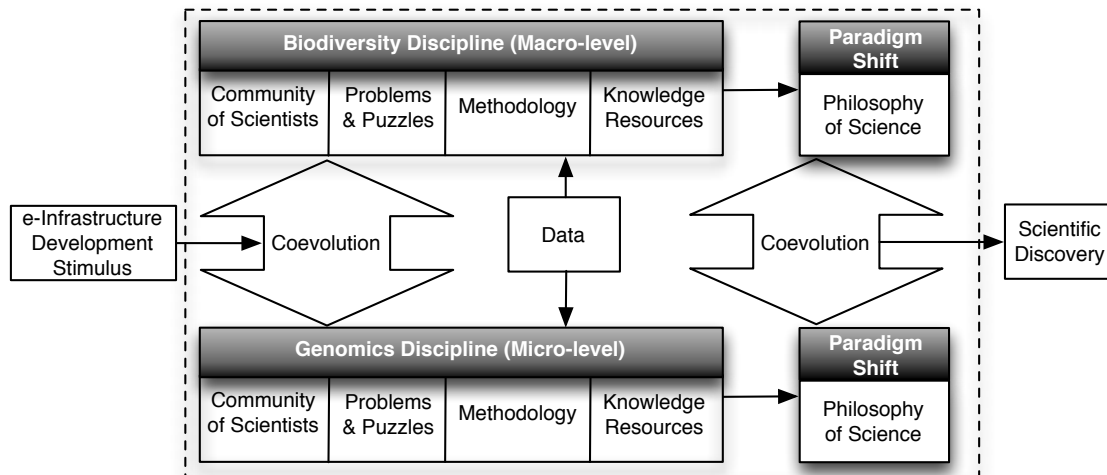
Increases in multidisciplinary research have raised demand for inter-data network exchanges. Reichman et al (2011) found that the holdings and collections in data networks are fragmented due to lack of mechanisms supporting facile search and discovery between data networks. This increase in demand is shown as step (6), Figure 20.

Data providers are exploring “federation” between data networks to support inter-data network exchanges (Reichman et al, 2011). Funded by investments largely from the NSF<sup>35</sup>, Reichman et al (2011) found that data providers are collaborating to form federations that will eventually be cross-linked and interoperable with one another. We show this as step (7), Figure 20.

The Concept Map is revised to show the macro and micro level disciplines and their relationship to the data component (Figure 22 below). A description of Concept Map 4 follows.

---

<sup>35</sup> [http://www.nsf.gov/funding/pgm\\_summ.jsp?pims\\_id=503141](http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503141)



**Figure 22 Macro-Micro level relationship: Concept Map 4**

Concept Map 4 (Figure 22) depicts the macro-micro science discipline relationship. The Technological and Socio-organizational e-Infrastructure components and the Productivity Variables in Concept Map 3 are not shown; however, they are present as substrates in each of the macro-micro disciplines.

Consistent with prior concept maps, an external stimulus is introduced to develop technology e-Infrastructure. The theory posits that investments in technology e-Infrastructure would dramatically improve scientific progress.

Reading Concept Map 4 from left to right, the left Coevolution double arrow represents a set of co-evolutionary pressures from interactions between technological e-Infrastructure development components, and the biodiversity and genomics discipline components. As described in step (2) in Figure 20, niches of data were created at dramatic rates and volume in the genomics and biodiversity environments. This volume of data generated an imbalance in the environment. The Data component in Figure 22 provides a conceptualization of these niches and potential imbalances produced by the interactions between the technology e-Infrastructure components and the socio-organizational components in the e-Infrastructure level.

Niches and imbalances represented by the Data component in Concept Map 4 represent conditions that can lead to stimulating complementary innovations (IT Productivity Paradox concept). Findings indicated that complementary innovations that resulted facilitated the sharing and combining of heterogeneous data between biodiversity and genomics.

The right Coevolution arrow represents co-evolutionary pressures between how science is done and how science is thought about. These two dimensions about Aspects of a science discipline were introduced in the initial Concept Map (Figure 5).

Findings from the biodiversity and the genomics cases supported the proposition that data infrastructure evolves as it mediates complementary innovations. Complementary innovations have resulted in the development of new methods in biodiversity to work with heterogeneous data. Data networks and federation of data networks to support inter-data network exchanges were found to be evolving new scientific methods from complementary innovations. New scientific methods connect the disciplines to the Philosophy of Science components. Ultimately, Concept Map 4 posits that these complementary innovations and new scientific methods combined at the macro-micro discipline levels can lead to scientific discovery.

## 10. Consolidation of Findings and Conclusion

Governments are making significant investments in developing e-Infrastructures aimed at stimulating dramatic increases in scientific discovery. Scientific discovery is recognized as central to achieving key national goals, such as raising living standards, creating good jobs, ensuring national security, strengthening education, improving public health, and protecting the environment (NAP, 1999; NAP, 2007).

Transformative research aims to increase revolutionary scientific discoveries through the application of unconventional or radical approaches to actual problems and scientific puzzles (NSB, 2007). E-Infrastructure is a phenomenon that is driven by government initiatives, aimed at transforming how science is done in order to increase scientific discovery. Development of e-Infrastructure is a response to the goals of Transformative Research. While investments in e-Infrastructures continue to play a significant role as a stimulus towards increasing transformative research, studies to understand their effectiveness are few or do not yet exist.

This research set out to explore the terrain of the phenomenon of e-Infrastructure development and its impact on scientific discovery. The objective of the research study was to:

*(a) Understand how the development of e-Infrastructure is impacting scientific discovery.*

*(b) Understand how the Problems and Puzzles of a science discipline shape the development of e-Infrastructure, and conversely, how e-Infrastructure changes the Problems and Puzzles of a science discipline.*

We constructed a conceptual framework that represented our hypothesis that a significant majority of investments were allocated for technological infrastructure development with the intention that they were going to dramatically increase scientific discovery. The study employed an interpretive grounded theory

exploratory research approach to study the Long-Term Ecology Research (LTER) Network — a particular community of science in biodiversity-ecology discipline.

The remainder of this chapter discusses the findings of the research, and provides answers to the research questions that constituted the objectives of this inquiry. In addition, the chapter discusses overall conclusions based on the findings. The chapter describes implications and recommendations for parties with potential interest to this research (e.g., scientists, managers of science and funding bodies). Limitations of the study and future research directions are discussed.

This concluding chapter is organized as follows: First, section 10.1 discusses the findings and provides answers to the research questions. Section 10.2 discusses the overall conclusions. Section 10.3 discusses implications and recommendations to interested parties. Section 10.4 discusses the limitations of this research. Finally, section 10.5 discusses the future research directions.

## **10.1 Discussion of the Findings**

This section discusses the findings of the research and provides answers to the research questions.

### **10.1.1 ICT Investment Stimuli Impacts Technology Infrastructure Development**

We found ICT investments towards the development of a national e-Infrastructure capability were increasing and were coordinated among U.S. government agencies when the Congress passed the High-Performance Computing Initiative (HPCCI) in 1991. In 1992, total annual ICT investment was \$514M. Meanwhile, investments in social-organizational development — human resources, workforce development, etc. — were dramatically less, with the largest allocation of \$140M in 1993.

By 2000, total annual ICT investment increased to \$1,228M. National e-Infrastructure initiatives, such as Cyberinfrastructure and e-Science, started in 2003, helped to account for the largest annual investment of \$3,187M by 2009. By comparison, investments in socio-organizational development reached a maximum of \$179M in 2010.

Our findings were consistent with our hypothesis that significant investments in technological infrastructure were being made to develop a capability, at a national level, towards increasing scientific progress and discovery. We also found that comparative investments in socio-organizational development were significantly less. This seemed to indicate that priorities focused primarily on the development of technological infrastructure; investments on a comparable Socio-organizational Infrastructure were secondary.

Investments in ICT as stimuli to science research are growing (NITRD, 2012). Although e-Science and Cyberinfrastructure programs are relatively new forms of stimuli, they are supporting a trend that started years earlier, and started accelerating in 1991 as a result of the HPCCI.

#### **10.1.2 Technology Infrastructure Impacts the Growth Rate of Data**

We explored the impact of Technology Infrastructure development by performing a case study of the U.S. Long Term Ecological Research (LTER) Network — a particular community of science in the biodiversity-ecology discipline. ICT investments in this community were explored using our conceptual lens to observe impact on transformative research and discovery.

The growth rate of data increased dramatically as a result of the adoption of technologies that became commercially available in the early 1980s. Microcomputers and the Internet were disruptive technologies that enabled scientists to transform how they work, collaborate and share information. Effects from advances in computer, storage and network technologies, as explained by Moore's Law (Stafford et al, 1994) and the concept of Price Elasticity (Gallaughier, 2008), had a dramatic impact on also advancing other technologies, such as Database Management Systems, Remote Sensing and Geographic Information Systems, Collaboration systems (e.g., email), Sensor Networks, etc. Affordability provided access to these technologies, which resulted in dramatic increases in data production. This explosive growth rate in data has come to be known in science as the "data deluge" (Hey and Trefethen, 2005).

Findings from our analysis of the LTER Network over a period of three decades (1980 to 2010) showed patterns of technology e-Infrastructure development and their impacts on the rate of data production. First, production of primary (raw) data grew dramatically as different types of sensor-based technologies were adopted; e.g., remote-sensing technologies, coupled with GIS systems enabled scientists to increase the scale of Problems and Puzzles and enhance cross-site comparative research. Second, automation of processes for the production of data products based on the use of online technologies increased in decade II. In decade III, the development of a systemic information infrastructure was underway that would integrate heterogeneous long-term data into a common data format, called “derived data.”

### **10.1.3 Technology and Data Infrastructures Influence How Science Is Practiced**

Comparative research across spatial and temporal scales increased in priority as technology and data infrastructures facilitated access to and exchange of data resources. The demand to compare data from multiple sources increased. For example, in 1985 the LTER Network was under pressure to work with data sets from diverse sources, such as other sites or even government agencies. Today, the LTER Network is coordinating with other environmental and observatory research networks. As a result, the demand for investment in e-Infrastructure development to enhance comparative research from multiple diverse sources to provide comprehensive, integrated, and synthesized science at regional to continental scales continues to increase (LTER-SIP, 2011).

E-Infrastructure facilitated data and information discovery, data access and the automation of processes across multiple sites. Data sets online increased in number. So too did the demand for access to these data sets increase. Data access and data sharing demands increased as e-Infrastructure developed. Yet, cross-site comparative science was not progressing as intended. Findings showed that a socio-organizational condition was impacting the data sharing goals of the LTER community.

#### 10.1.4 Data Infrastructure influences Change in Socio-organizational Infrastructure

Demand for use of long-term archival data increased to address the goals of the LTER Network program (Robbins, 2011). One goal of the LTER Network was to produce excellent site science, and to increase production of sharable long-term data sets. However, data access raised issues from scientists concerning data sharing. Therefore, data sharing emerged as a barrier rather than a bridge to advancing cross-site comparative research.

Drawing upon our conceptual framework to explain our findings, the goal to share data resources created the conditions to establish linkages between the LTER sites and other sites that were sources of data. Governance mechanisms were needed to establish these linkages to support data sharing. Specifically, in LTER, guidelines for data-access and data sharing were adopted at each site.

We found that data was developing its own infrastructure, and taking on the role of mediating between Technology Infrastructure, Socio-organizational Infrastructure and Aspects of a Science Discipline. Creation and management of shareable long-term data sets evolved into a data infrastructure that affected scientists, and resource providers. As demand for sharing of data increased, a data policy was developed that defined scientists' roles and responsibilities as "Data Collectors" and "Data Users," and how long data access can be restricted. "Data Managers" were responsible for archiving and preserving data and metadata, as well as ensuring that access was provided only to those users with proper credentials (Porter, 2010). Data Collectors, Managers, and Users roles and responsibilities provided findings that linkages and paths were forming between the Technology and Socio-organizational Infrastructures, and the LTER Network science community.

A metadata standard (or gateway) was adopted. This accelerated the customization of technologies, such as database systems for the management of XML documents and included support of metadata language. This had a lock in effect between the LTER community, resource providers of e-Infrastructure, and the processes supported by the Socio-organizational Infrastructure.



We modified our Concept Map to reflect our findings about data in a mediating role. Concept Map 2, Figure 17, represents our findings that Technology Infrastructure development leads to complementary stimuli in Socio-organizational Infrastructure mediated by the data infrastructure. Concept Map 2 provided a finding towards an answer to the primary research question: Data mediates the process of e-Infrastructure development on scientific discovery.

### 10.1.5 Impact to Scientific Progress: Mixed Results

A finding we expected was that stimuli of an e-Infrastructure development process would lead to increases in scientific discovery. Our results were at best mixed. We found that our chain of evidence did not support our hypothesis that Technology Infrastructure investments will ultimately result in dramatic improvements in scientific progress, in particular revolutionary scientific progress. Scientific Discovery, our dependent variable, was not radically changing as a result of e-Infrastructure investments.

We turned to the IT Productivity Paradox to explain similar outcomes involving ICT investments. Studies have used the IT Productivity Paradox to assess ICT investments and their impact on productivity both at the organization level and at the national or industry level (Chan, 2000; Park et al, 2007). The IT Productivity Paradox provided a framework upon which we continued to explore the process of e-Infrastructure development and its relationship to scientific discovery.

The IT Productivity Paradox literature provided the following four variables to explain how investments in ICT are linked to productivity: (1) Transformation, (2) Complementarities, (3) Transferable ICT products, and (4) Time Lag.

We found these variables of the IT Productivity Paradox to be consistent with the components of our Concept Map. Table 5 provides a summary of the IT Productivity Paradox variables, and their role in the Concept Map.

IT Productivity Paradox Variable	Description	Role in Concept Map
<i>Transformation</i>	Introduces change through	Provides a stimulus

	purposeful ICT investments	towards Technology Infrastructure development
<i>Complementarities</i>	Leveraging ICT investments to change aspects of an organization	Socio-organizational Infrastructure leveraging investments in technological infrastructure, mediated through data
<i>Transferable IT Products</i>	Establishes a relationship between the successful transfer of IT products and technology innovation	Adapt “IT products” to “Data products.” Establishes a relationship between the successful transfer of “data products” and “scientific discovery.”
<i>Time Lag</i>	Benefits from ICT investments can take several years to show positive results	Complementary stimuli in Socio-organizational Infrastructure categories taking time to leverage the Technology Infrastructure investment

**Table 5 IT Productivity Paradox variables and relationship to Concept Map**

Concept Map 2 was modified to include these four IT Productivity Paradox variables and create Concept Map 3, figure 18. The primary research question was reframed to consider the impact on “scientific progress”:

*How is investment in e-Infrastructure development impacting scientific progress?*

We found in the biodiversity case study that while the impact to scientific discovery results was mixed, use of the IT Productivity Paradox variables allowed us to adapt the Concept Map so that it can guide us to observe patterns of scientific progress.

### 10.1.6 Data Sharing Across Disciplines

We found during our exploration of investments into e-Infrastructure development that data infrastructure was also a component of increasing investment<sup>36,37</sup>. The demand for data sharing and data preservation is increasing, and the Technology Infrastructure needed to support it is a component of e-Infrastructure development.

Data sharing between disciplines was a finding in the biodiversity case study, which then led to the following proposition:

*Combining data from multiple disciplines can lead to increasing discovery.*

Data sharing was explored between the disciplines of biodiversity and genomics. Moreover, arguments were presented to link biodiversity and genomics as macro-micro level disciplines, respectively. Biodiversity and genomics were identified as neighboring disciplines.

The research question used to motivate the inquiry was the following:

*How does genomics data combined with biodiversity data produce a result that leads to increasing discovery?*

Genomics was found to be connected to biodiversity by showing that genomics data was used to confirm biodiversity propositions. Two case examples were provided where data from DNA barcoding was used in biodiversity experiments. Case example 1 found results where DNA barcoding data was used in biodiversity to enhance classification of species. Discovery of “cryptic species” resulted from the use of DNA barcoding data. As a result, gaps in biodiversity inventories are being filled.

Case example 2 set out to find interconnections between biodiversity and genomics mediated through data. Results of the analysis supported the following

---

<sup>36</sup> Networking and Information Technology Research and Development (NITRD), <http://www.itrd.gov/pubs/>

<sup>37</sup> NSF Budget Requests to Congress and Annual Appropriations, <http://www.nsf.gov/about/budget/>

interconnections represented in Concept Map 2: (1) Technology Infrastructure to Data, (2) data to Socio-organizational Infrastructure, and (3) Socio-organizational Infrastructure to data. Results of case example 2 found evidence of complementary innovations that evolved from the investments in e-Infrastructure, which supported shared data infrastructure between biodiversity and genomics communities.

The case examples were concluded with a revision of the Concept Map — shown as Concept Map 4, Figure 22

## **10.2 Contributions of the Research**

The contributions of this research to theory and practice are now described.

### **10.2.1 Contributions to Theory**

First, this research makes contributions to the literature on the concept and phenomenon of e-Infrastructure. E-Infrastructure and its development are a recent phenomenon; therefore, the scholarly literature on this topic is limited. The theory building process brought together several streams of literature on the history of Large Technological Systems and infrastructure (Hughes, 1983, 1994; Star and Ruhleder, 1995; Edwards et al, 2007), Social Construction of Technology and Infrastructure (Pinch and Bijker, 1994; Law, 1987), and Technology Domains (Arthur, 2009).

Secondly, this research constructs a conceptual framework to link the concepts of e-Infrastructure Development to scientific discovery. The construction of the conceptual framework makes two important contribution: (a) It brings together a set of concepts and establishes connections between them to develop a theoretical framework to further explore the phenomenon of e-Infrastructure and its relationship to scientific discovery; and (b) it brings together several streams of literature on how science is practiced (Kuhn, 1996; Popper, 1963; Graham et al, 2002) and thought about (Churchman, 1971; Popper, 2002; Glass, 2007).

### 10.2.2 Contributions to Practice

This research makes contributions to practice for scientists, managers of science, government agencies supporting science and the development of e-Infrastructure, and resource providers to science R&D. Scientists and managers of science will be able to associate concepts of e-Infrastructure with contemporary programs of Cyberinfrastructure and e-Science. This is possible through the use of the conceptual lens to explore and gain insights on the interactions of the programs with a science discipline. Government agency representatives will find the use of the conceptual framework in the case studies of particular interest in gaining insights about path formation from stimulus to discovery. Resource providers will find insights about their role in the environment of a science discipline and its interaction with e-Infrastructure development.

### 10.3 Limitations of the Study

Two limitations of the study to be aware of concerning the conceptual framework are: (a) the approach taken in its construction, and (b) the role of data.

The approach taken in the construction of the conceptual framework was to take a broad perspective as a guiding principle when connecting concepts into components, then interconnecting components. Our perspective was “broad, not deep.” A deeper treatment of the concepts was not required for this study. Another study might benefit from taking a perspective from deeper into a component, depending on the research question.

The Concept Map evolved to promote data from a concept in the e-Infrastructure development process into an infrastructure component. We learned from the empirical inquiry using a grounded theory approach. This approach adapted the Concept Map in an iterative manner. Using data from a different discipline may yield a different evolution of the Concept Map. For example, in the discipline of molecular biology, data sharing is built in to that discipline’s infrastructure since each publication must link to a data submission (Robbins, 2011; Costello, 2009).

#### 10.4 Future Research Directions

E-infrastructure and data infrastructure development are contemporary national initiatives, making fertile ground for future research opportunities. Descriptions of future research directions presented here are the following: (1) testing the theory with another pair of adjacent disciplines, (2) increasing understanding of the macro-micro level relationship concept between two disciplines, and (3) exploring the relationship between data infrastructure development and how it impacts discovery.

*Testing the theory with another pair of adjacent disciplines:* The conceptual framework and theory of this study should be tested against another pair of adjacent disciplines. The suggested research approach is case studies for each discipline, with cross case analysis against biodiversity-genomics disciplines, testing the theory and the macro-micro level relationship. Astronomy and particle physics are the two adjacent disciplines suggested. Astronomy-particle physics would represent physical sciences, and biodiversity-genomics would represent life sciences.

*Increasing understanding of the macro-micro level relationship concept between two disciplines:* Although the macro-micro level relationship concept was not the focus of this study, it brings together two adjacent disciplines with common goals, but with differing approaches. Using the conceptual lens with the data infrastructure in a mediating role might reveal interesting findings.

*Exploring the relationship between data infrastructure development and how it impacts discovery:* This is a case where inquiry would go deeper instead of broader, as previously described. Data infrastructure is an active area of investment and research from several perspectives; e.g., sharing, reproducibility, longevity and sustainability, ethical and legal implications (NSB-11-79, 2011).

## 10.5 Overall Conclusions

This research examined the phenomenon of e-Infrastructure development and its impact on scientific discovery. Using an interpretive grounded theory research approach to study the Long-Term Ecology Research (LTER) Network — a particular community of science in the biodiversity-ecology discipline — we found that increases in scientific discovery as a result of significant investments in e-Infrastructure development were at best mixed. Reframing our research objective to focus on scientific progress and using the IT Productivity Paradox variables to observe “scientific productivity,” we found results that appeared promising. Although the findings of this exploratory study were mixed, they raised awareness of potential research in several areas outlined previously in section 10.4. We hope that this study will provide a stimulus to future research of the phenomenon of e-Infrastructure and its impact on the different Aspects of a Science Discipline and scientific discovery.

## References

- Aamodt, A., and Nygård, M. (1995). "Different roles and mutual dependencies of data, information, and knowledge: an AI perspective on their integration." *Data and Knowledge Engineering*(16): 191-222.
- Abiteboul, S., Buneman, P., Suci, D. (2000). *Data on the Web: From Relations to Semistructured Data and XML*. San Francisco, CA, Morgan Kaufmann Publishers.
- Agrawal, A., and Goldfarb, A. (2008). "Restructuring research: Communication costs and the democratization of university innovation." *American Economic Review* 98(4): 1578-1590.
- Aldrich, H. (1979). *Organizations and Environments*. Englewood Cliffs, New Jersey, Prentice-Hall.
- Andelman, S. J., Bowles, C., Willig, M. R., Waide, R. (2004). "Understanding Environmental Complexity through a Distributed Knowledge Network." *BioScience* 54(3).
- Anderson, D. P. (2003). *Public Computing: Reconnecting People to Science*. Conference on Shared Knowledge and the Web, Residencia de Estudiantes, Madrid, Spain.
- Arneson, R. (2006). Excerpts from the *Structure of Scientific Revolutions* by Thomas Kuhn.
- Arthur, W. B. (2009). *The Nature of Technology: What It Is and How It Evolves*. New York, NY., Free Press.
- Atkins, D. E., et al (2003). *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*.
- Atkinson, M., DeRoure, D., et al (2004). "Web Service Grids: An Evolutionary Approach." *Concurrency and Computation: Practice and Experience* 17: 377-389.
- Baker, K. S., et al (2000). "Evolution of a Multisite Network Information System: The LTER Information Management Paradigm." *BioScience* 50(11): 963-978.
- Bany Mohammed, A. A., Altman, J. (2010). "A funding and governing model for achieving sustainable growth of computing e-infrastructures." *Annals of Telecommunications* 65: 739-756.
- Bell, G., & Gray, J. (2002). "What's next in high-performance computing." *Communications of the ACM* 45(2): 91-95.



- Bijker, W., & Law, J. (1992). *Shaping Technology/Building Society: studies in socio-technical change*. Cambridge, MA., MIT Press.
- Bijker, Wiebe, Thomas P. Hughes, and Trevor Pinch, eds. (1987). *The Social Construction of Technological Systems*, Cambridge, MA: MIT Press.
- Blaxter, M. L., et al. (1998). "A molecular evolutionary framework for the phylum Nematoda." *Nature*(392): 71-75.
- Blaxter, M. (2003). "Counting Angels with DNA." *Nature*(421): 122-124.
- Blaxter, M., Floyd, R., Dorris, M., Eyualem, A., & De Ley, P. (2004). Utilising the new nematode phylogeny for studies of parasitism and diversity. *Nematology monographs and perspectives*. R. Cook, & Hunt, D. J. Leiden, E. J. Brill: 615-632.
- Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., Abebe, E. (2005). "Defining operational taxonomic units using DNA barcode data." *Philosophical Transactions of the Royal Society B* 360: 1935-1943.
- Blumenthal, M. S. (1998). "Federal Government Initiatives and the Foundations of the Information Technology Revolution: Lessons from History." *The American Economic Review* 88(2): 34-39.
- Bohannon, J. (2005). *Distributed Computing: Grassroots Supercomputing*. *Science Magazine*. 308: 810-812.
- BON (1998). *Final Report: BIODIVERSITY MONITORING WORKSHOP. A NATIONAL BIODIVERSITY OBSERVATORY NETWORK: Recommendations and Rationale*, University of Virginia: 1-6.
- Boyd, B. (1990). "Corporate Linkages and Organizational Environment: A Test of the Resource Dependence Model." *Strategic Management Journal* 11(6): 419-430.
- Braun, Ingo (1994). "Geflügelte Saurier: zur intersystemische vernetzung grosser technische Netze," in Braun, Ingo and Bernward Joerges, eds., *Technik ohne Grenzen*. Frankfurt am Main: Suhrkamp, 446-500.
- Braun, Ingo and Bernward Joerges (1994). *Technik ohne Grenzen*. Frankfurt am Main: Suhrkamp.
- Brown, J. H. (1994). Grand Challenges in scaling up environmental research. *Environmental Information Management and Analysis: Ecosystem to Global Scales*. W. K. Michener, Brunt, J.W., Stafford, S.G.: 21-26.
- Brunt, J. W., McCartney, P., Baker, K., Stafford, S. (2002). *The Future of Ecoinformatics In LongTerm Ecological Research*. *Proceedings of the 6th*

- World Multi-Conference on Systematics, Cybernetics and Informatics, Orlando, FL, The International Institute of Informatics and Systematics.
- Brunt, J. W., and Michener, W. K. (2009). "The resource discovery initiative for field stations: Enhancing data management at North American biological field stations. ." *BioScience* 59: 482-487.
- Brynjolfsson, E. (1993). "The Productivity Paradox of Information Technology." *Communications of the ACM* 36(12): 67-77.
- Brynjolfsson, E., Malone, T. W., Gurbaxani, V., Kambil, A. (1994). "Does information technology lead to smaller firms?" *Management Science*.
- Brynjolfsson, E., and Hitt, L. M. (2000). "Beyond Computation: Information Technology, Organizational Transformation and Business Performance." *The Journal of Economic Perspectives*, 14(4): 23-48.
- Callahan, J. T. (1984). "Long-Term Ecological Research." *BioScience* 34(6): 363-367.
- Campbell, D. T. (1960). "Blind Variation and Selective Retention in Creative Thought as in other Thought Processes." *Psychological Review* 67: 380-400.
- Campbell, D. T. (1969). "Variation and selective retention in socio-cultural evolution." *General Systems* 14: 69-85.
- Carey, C. C., Hanson, P. C., Bruesewitz, D. A., Holtgrieve, G. W., Kara, E. and R. L., K. C., Smyth, R. L., & Weathers, K. C. (2012). Organized Oral Session 43: Novel Applications of High-Frequency Sensor Data in Aquatic Ecosystems: Discoveries from GLEON, the Global Lake Ecological Observatory Network. . *Bulletin of the Ecological Society of America*: 100-105.
- Casanova, H. (2002). "Distributed Computing Research Issues in Grid Computing." *ACM SIGACT News Distributed Computing Column* 8.
- Chan, Y. E. (2000). "IT value: The great divide between qualitative and quantitative and individual and organizational measures." *Journal of Management Information Systems* 16(4): 225-261.
- Chavan, V. S., et al (2010). *State-of-the-Network 2010: Discovery and Publishing of the Primary Biodiversity Data Through the GBIF Network (Global Biodiversity Information Facility)*. Copenhagen, DK, Global Biodiversity Information Facility Secretariat.
- Charmaz, K. (2006). *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. Thousand Oaks, CA 91320, SAGE Publications.
- Chernoff, B. (1986). *Systematics and long-range ecologic research. Foundations for a national biological survey*. K. C. a. K. Kim, L. Lawrence, Kans., Assoc. Systematics Coll.

- Churchman, C. W. (1971). *The Design of Inquiry Systems*. New York, NY., Academic Press.
- Claridge, M. F. (1995). "Introducing Systematics Agenda 2000." *Biodiversity and Conservation* 4: 451-454.
- Claridge, M. F., Dawah, H. W., Wilson, M. R. (1997). *Species: Units of Biodiversity*. London, Chapman and Hall.
- Cohen, J. (1996). "Computer mediated communication and publication productivity among faculty." *Internet Research* 6 (2-3): 41-63.
- Collins, S. L., et al (2006). "New Opportunities in ecological sensing using wireless sensor networks." *Frontiers in Ecology* 4: 402-407.
- Constant, E. W. (1978). "On the Diversity and Co-Evolution of Technological Multiples: Steam Turbines and Pelton Water Wheels." *Social Studies of Science* 8(2): 183-210.
- Constant, E. W., II. 1980. "The Origins of the Turbojet Revolution", Baltimore: Johns Hopkins University Press.
- Cooper, H. (1998). *Synthesizing Research: a guide for Literature Reviews*. Thousand Oaks, CA, Sage Publications.
- Costello, M. (2009). "Motivating Online Publication of Data." *BioScience* 59(5): 418-427.
- Council on Competitiveness. *Going Global: The New Shape of American Innovation*. Washington, DC, Council on Competitiveness, 1998.
- Coutard, Olivier (1999). *The Governance of Large Technical Systems*. New York: Routledge.
- Coutard, Olivier, Richard Hanley, and Rae Zimmerman (2004). *Sustaining Urban Networks: the social diffusion of large technical systems*. New York, NY: Routledge.
- Cummings, J. N., Finholt, T., Kesselman, C., Lawrence, K. A. (2008). *Beyond Being There: A Blueprint for Advancing the Design, Development, and Evaluation of Virtual Organizatons*. Arlington, VA, National Science Foundation.
- Darwin, C. (1859). *On the Origin of Species by means of natural selection, or, the preservation of favoured races in the struggle for life*. London, UK., J. Murray.
- David, P. A. (1990). "The Dynamo and the Com- puter: A Historical Perspective on the Modern Productivity Paradox." *American Economic Review Papers and Proceedings* 1(2): 355-361.

- DeSanctis, G., & Poole, M. S. (1994). "Capturing the Complexity in Advanced Technology Use: Adaptive Structuration Theory." *Organization Science* 5(2): 121-147.
- Dewan, S., and Kraemer, K. L. (1998). "International Dimensions of the Productivity Paradox." *Communications of the ACM* 41(8): 56-62.
- Ding, W. W., Levin, S. G., Stephan, P. E. Winkler, A. E. (2010). "The Impact of Information Technology on Academic Scientists; Productivity and Collaboration Patters." *Management Science* 56(9): 1439-1461.
- Duraiappah, A. K., and Naeem, S. (2005). *Ecosystems and Human Well-being: Biodiversity Synthesis. Millennium Ecosystem Assessment.* Washington, DC, World Resources Institute.
- Economist (1990). *Working Harder, Doing Less.* The Economist. 315.
- Edwards, P. N., Jackson, S. J., Bowker, G. C., Knobel, C. P. (2007). *Understanding Infrastructure: Dynamics, Tensions and Design. Report of a Workshop on History and Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures.* Arlington, VA, National Science Foundation.
- e-IRG (2007). *A SUSTAINABLE GRID INFRASTRUCTURE FOR EUROPE: Executive Summary of the e-IRG Open Workshop on e-Infrastructures.* Heidelberg, Germany.
- e-IRGSP (2005). *e-Infrastructures Roadmap.*
- Ehrlich, P. R., and Raven, P. H. (1964). "Butterflies and Plants: A Study in Coevolution." *Evolution* 18(4): 586-608.
- Ehrlich, P. R., and Wilson, E. O. (1991). "Biodiversity Studies: Science and Policy." *Science* 253(5021): 758-762.
- Eid, J., et al (2009). *Real-Time DNA Sequencing from Single Polymerase Molecules.* *Science Magazine, AAAS.* 323: 133-138.
- Eisenhardt, K. M. (1989). "Building Theories from Case Study Research." *Academy of Management Review* 14(4): 532-550.
- Fahey, T. J., and Knapp, A. K. (2007). *Principles and Standards for Measuring Primary Production.*, Oxford University Press.
- Feldman, R. M., and Manning, R. B. (1992). "Crisis in Systematic Biology in the "Age of Biodiversity"." *Journal of Paleontology*, 66(1): 157-158.
- Fitzhugh, K. (2006). "DNA barcoding: an instance of technology-driven science?" *BioScience* 56(6).

- Fortuner, R. (1993). *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision*, John Hopkins University Press.
- Foster, I. (2005). Service-Oriented Science. *Science*. 308: 814-817.
- Foster, I., and Kesselman, C. (1999). *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann Publishers.
- Foster, I., and Kesselman, C. (2004). *The GRID2: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann.
- Fox, G., & Walker, D. (2003). e-Science Gap Analysis, [www.grid2002.org/ukescience/gapresources/GapAnalysis30June03.pdf](http://www.grid2002.org/ukescience/gapresources/GapAnalysis30June03.pdf).
- Friedlander, A. (1995). *Emerging Infrastructure: The Growth of Railroads*. Reston, VA, Corporation for National Research Initiatives.
- Furman, J. L., Stern, S. (2009). "Climbing atop the shoulders of giants: The impact of institutions on cumulative research." Working paper, Boston University, Boston.
- Gallaughier, J. M. (2008). *Moore's Law & More: Fast, Cheap Computing, and What it Means for the Manager*.
- Gaston, K. J., and Spicer, J. I. (2004). *Biodiversity: An Introduction*, Wiley and Sons.
- Giddens, A. (1984). *The Constitution of Society: Outline of the Theory of Structuration*, University of California Press.
- Glass, D. J. (2007). *Experimental Design for Biologists*. Cold Spring Harbor, Cold Spring Harbor Laboratory Press.
- Gosz, J. R., Gage, S., Inglis, M., Melillo, J., Ross, V., Zimmerman, P. (1989). *LTER Technology Report for the LTER Strategic Plan*.
- Gosz, J. R., Waide, R.B., and Magnuson, J.J. (2010). *Twenty-Eight Years of the US-LTER Program: Experience, Results, and Research Questions*. Long-Term Ecological Research (LTER). F. M. e. al, Springer.
- Graham, M. H., and Dayton, P. K. (2002). "On the Evolution of Ecological Ideas: Paradigms and Scientific Progress." *Ecology* 83(6): 1481-1489.
- Greenland, D., Goodin, D. G., and Smith, R. C. (2003). *Climate Variability and Ecosystem Response at Long-term Ecological Research Sites*, Oxford University Press.
- Geregory, T. R. (2005). "DNA barcoding does not compete with taxonomy." *Nature* 434(1067).

- Gross, K. L. e. a. (1995). *Future of Long-Term Ecological Data (FLED)*. E. S. o. America. Washington, DC.
- Hannan, M. T., and Freeman, J (1977). "The Population Ecology of Organizations." *American Journal of Sociology* 82(5): 929-964.
- Hanson, P. (2008). "New ecological insights through the Global Lake Ecological Observatory Network (GLEON)." *Ecological Science* 27(5): 300-302.
- Hawley, A. H. (1988). *Forward. Ecological models of organizations*. G. R. Carroll. Cambridge, MA, Ballinger.
- Hebert, P. D. N., Cywinska, A., Ball, S. L., & De Waard, J. R. (2003a). "Biological identifications through DNA barcodes." *Proceedings of the Royal Society of Biology* 270: 313-321.
- Hebert, P. D. N., Ratnasingham, S., & De Waard, J. R. (2003b). "Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species." *Proc. R. Soc. B* 270: S96-S99.
- Hebert, P. D. N., and Gregory, T. R. (2005). "The Promise of DNA Barcoding for Taxonomy." *Systematic Biology* 54(5): 852-859.
- Henshaw, D. L., Sheldon, W. M., Remillard, S. M., Kotwica, K. (2006). *Proceedings of the 7th International Conference on HydroScience and Engineering (ICHE 2006)*, Philadelphia, USA, Drexel University.
- Hesse, B. W., Sproull, L. S., Kiesler, K., & Walsh, J.P. (1993). "Returns to Science: Computer Networks in Oceanography." *Communications of the ACM* 36(8): 90-101.
- Hey, T., and Trefethen, A. E. (2002). "The UK e-Science Core Programme and the Grid." *Future Generation Computer Systems* 18: 1017-1031.
- Hey, T., & Trefethen, A. (2003). *The Data Deluge: An e-Science Perspective*. Grid Computing: Making the Global Infrastructure a Reality. F. Berman, Hey, T., Fox, G., John Wiley and Sons.
- Hey, T., and Trefethen, A. E. (2005). "Cyberinfrastructure for e-Science." *Science* 308: 817-821.
- Hine, C. (2006). *New Infrastructures for Knowledge Production: Understanding e-Science*, Information Science Publishing.
- Hitt, L., and Brynjolfsson, E (1997). "Information Technology and Internal Firm Organization: An Exploratory Analysis." *Journal of Management Information Systems* 14(2): 81-101.
- Hobbie, J. E., Carpenter, S. R., Grimm, N. B., Gosz, J. R., and Seastedt, T. R. (2003). "The U.S. Long Term Ecological Research Program." *BioScience* 53(1): 21-32.

- Hughes, T. P. (1983). *Networks of power : electrification in Western society, 1880-1930*. Baltimore, MD, Johns Hopkins University Press.
- Hughes, T. P. (1987, 1994). *The Evolution of Large Technical Systems. The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. W. E. Bijker, and Pinch, T. P., MIT Press.
- Jankowski, N. W., and Caldas, A. (2004). *e-Science: Principles, Projects and Possibilities for Communication and Internet Studies*.
- Janzen, D. H. (2004). "Now is the time." *Philosophical Transactions of the Royal Society B*. 359: 731-732.
- Janzen, D. H., Hajibabaei, M., Burns, J. M., Hallwachs, W., Remigio, E., & Hebert, P. D. N. (2005). "Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. ." *Philosophical Transactions of the Royal Society B*(360): 1835-1845.
- Johnson, J. C., Christian, R. R., Brunt, J. W., Hickman, C. R., Waide, R. B. (2010). "Evolution of Collaboration within the U.S. Long-Term Ecological Research Network." *BioScience* 60(11): 931-940.
- Kahn, S. D. (2011). "On the Future of Genomic Data." *Science Magazine* 331(6018): 728-729.
- Kaijser, Arne, and Marika Hedin, eds. (1995). *Nordic energy systems: historical perspectives and current issues*. Canton, MA: Science History Publications/USA.
- Katz, J. a. G., W. B. (1988). "Properties of Emerging Organizations." *The Academy of Management Review* 13(3): 429-441.
- Kiesler, S. (1986), "The Hidden Messages in Computer Networks," *Harvard Business Review*, February, 46-59.
- Kiesler, S., and Cummings, J. (2002). *What Do We Know about Proximity and Distance in Work Groups*. Distributed Work. P. Hinds, and Kiesler, S. Cambridge, MA, MIT Press: 57-80.
- Kirk, E. J. (2007). *Embracing Globalization: Meeting the Challenges to U.S. Scientists and Engineers*. Arlington, VA, National Science Foundation: 67.
- Kragh, H. (1987). *An Introduction to the Historiography of Science*, Cambridge University Press.
- Kuhn, T. S. (1996). *The Structure of Scientific Revolutions*, University of Chicago Press.

- La Porte, Todd R., ed. (1991). *Social Responses to Large Technical Systems: Control or Adaptation*, NATO ASI Series D, Behavioural and social sciences, vol. 58. Boston: Kluwer Academic Publishers.
- Lambert, D. M., Baker, A., Huynen, L., Haddrath, O., Hebert, P. D. N., & Millar, C. D. (2005). "Is a large-scale DNA-based inventory of ancient life possible? ." *Journal of Heredity*(96): 279-284.
- Lapadat, J. C., and Lindsay, A. C. (1999). "Transcription in Research and Practice: From Standardization of Technique to Interpretive Positioning " *Qualitative Inquiry* 5(1): 64-86.
- Lave, J. W., E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge, MA, Cambridge University Press.
- Law, J. (1987). *Technology and Heterogeneous Engineering: The Case of Portuguese Expansion*. *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. W. E. Bijker, and Pinch, T. P.
- Levin, S. A. (1992). "The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture." *Ecology* 73(6): 1943-1967.
- Lincoln, Y. S., and Guba, E. G. (1985). *Naturalistic Inquiry*. Newbury Park, CA, USA, Sage Publications.
- Lober, D. J. (1997). "Explaining the formation of business-environmentalist collaborations: Collaborative windows and the Paper Task Force." *Policy Sciences*(30): 1-24.
- Loehle, C. (1987). "Hypothesis testing in ecology: psychological aspects and the importance of theory maturation." *Quarterly Review of Biology*(62): 397-409.
- Low, B., Costanza, R., Ostrom, E., Wilson, J., and Simon, C. P. (1999). "Human Ecosystem Interactions: A Dynamic Integrated Model." *Ecological Economics* 31(20): 227-242.
- LTER-SIP (2011). *Strategic and Implementation Plan: Long Term Ecological Research Network (LTER): Research and Education*.
- LTER-WP (2001). *LTER White Paper: Priority Setting in the LTER Network, 2001*.
- Marcus, L. (1993). *The Goals and Methods of Systematic Biology*. *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision*. R. Fortuner, John Hopkins University Press.
- Markmann, M., & Tautz, D. (2005). "Reverse taxonomy: an approach towards determining the diversity of meio- benthic organisms based on ribosomal



- RNA signature sequences." *Philosophical Transactions of the Royal Society B*.(360): 1917-1924.
- Markus, M. L., and Robey, D. (1988). "Information Technology and Organizational Change: Causal Structure in Theory and Research." *Management Science* 34(5): 583-598.
- Martinson, M. G., and Martinsons, V. (2002). "Rethinking the Value of IT, Again." *Communications of the ACM* 45(7): 25-26.
- Mason, B. (2003). "Marine survey sees net gain in number of fish species." *Nature*(425): 889.
- Maxwell, J. A. (2005). *Qualitative Research Design: An interactive approach*, Sage Publications.
- Mayntz, Renate and Thomas P. Hughes, eds. (1988). *The Development of Large Technical Systems*. Boulder, CO: Westview Press.
- McKinsey (2002). *U.S. Productivity Report. 1995-2000*: [www.mckinsey.com/knowledge](http://www.mckinsey.com/knowledge).
- Mentis, M. T. (1998). "Hypothetico-deductive and inductive approaches in ecology." *Fundamental Ecology*(2): 5-14.
- Michener, W. K. (1986). *Data management and long-term ecological research. Research Data Management in the Ecological Sciences*. W. K. Michener.
- Michener, W. K., Miller, A. B., and Nottrott, R. (1990). *Long-term ecological research core data set catalog*. U. o. S. C. Belle W. Branch Institute of Marine Biology and Coastal Research. Columbia, SC, USA.
- Michener, W. K., Brunt, J.W., Helly, J. J., Kirchner, T. B., Stafford, S. G. (1997). "Nongeospatial Metadata for the Ecological Sciences." *Ecological Applications* 7(1): 330-342.
- Michener, W. K., Porter, J., Servilla, M., Vanderbilt, K. (2011). "Long term ecological research and information management." *Ecological Informatics* 6: 13-24.
- Miles, M. B., and Huberman, A. M. (1994). *Qualitative Data Analysis*, SAGE Publications.
- Monaghan, M. T., Balke, M., Gregory, T. R., & Vogler, A. P. (2005). "DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers." *Philosophical Transactions of the Royal Society B*.(360): 1925-1933.
- Murmann, J. P., Aldrich, H., Levinthal, D., Winter, S. (2003). "Evolutionary Thought in Management and Organization Theory at the Beginning of the New Millenium." *Journal of Management Inquiry* 12(1): 1-19.

- Murray, F., Aghion, P., Dewatripoint, M., Kolev, J., and Stern, S. (2008). "Of mice and academics: Examining the effects of openness on innovation." Working Paper, Massachusetts Institute of Technology, Cambridge.
- Myers, M. D. (1997). "Qualitative Research in Information Systems." MISQ Discovery.
- NAP (1999). Capitalizing on Investments in Science and Technology. Washington, DC, National Academies Press.
- NAP (2007). Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future, The National Academies Press.
- NAP (2010). Rising Above the Gathering Storm, Revisited: Rapidly Approaching Category 5, National Academies Press. Washington DC.
- Niskanen, W. A. (1991). "The Soft Infrastructure of a Market Economy." Cato Journal(2).
- NITRD (2001). Information Technology: The 21st Century Revolution. Supplement to the President's FY2001 Budget, National Science and Technology Council.
- NITRD (2012). Supplement to the President's Budget: FY 2013. The Networking and Information Technology Research and Development Program.
- NRC (2001). Grand Challenges in Environmental Sciences. Washington, DC., National Academies Press.
- NRC (2005). Assessment of Department of Defense Basic Research, National Academies Press.
- NSB (2007). Enhancing Support of Transformative Research at the National Science Foundation. 2007, National Science Board: 15.
- NSB (2010). Science and Engineering Indicators 2010. Arlington, VA, National Science Foundation. (NSB 10-01): 1-566.
- NSB-11-79 (2011). Digital Research Data Sharing and Management, Task Force on Data Policies.
- NSF (2003). The Internet: Changing the Way We Communicate: The Launch of NSFNET. N. S. Foundation.
- NSF (2007). Cyberinfrastructure Vision for 21st Century Discovery.
- NSF (2010). Data-Enabled Science in the Mathematical and Physical Sciences, National Science Foundation.
- NSF (2012). Long-Term Ecological Research (LTER): Program Solicitation NSF 12-524. D. o. B. Sciences. Arlington, VA, NSF.

- NSF-VOSS (2009). Virtual Organizations as Sociotechnical Systems (VOSS), National Science Foundation.
- NSTC (2009). Harnessing the Power of Digital Data for Science and Society. Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council, National Science and Technology Council.
- O'Malley, M. A., Elliot, K.C., Haufe, C., Burian, R.M. (2009). "Philosophies of Funding." *Cell* 138: 611-615.
- Pankhurst, R. J. (1993). Principles and Problems of Identification. *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision*. R. Fortuner, John Hopkins University Press.
- Park J., S., S. K., Sanders, G. L. (2007). "Impact of International Information Technology Transfer on National Productivity." *Information Systems Research* 18(1): 86-102.
- Partridge, C. (2008). "The Technical Development of Internet Email." *IEEE Annals of the History of Computing* 30(2): 3-29.
- Pazos, F., and Valencia, A. (2008). "Protein co-evolution, co-adaptation and interactions." *European Molecular Biology Organization* 27(20): 2648-2655.
- PCAST (2004). Sustaining the Nation's Innovation Ecosystems: Maintaining the Strength of Our Science and Engineering Capabilities, President's Council of Advisors on Science and Technology (PCAST).
- Pennisi, E. (2005). How Will Big Pictures Emerge From A Sea of Biological Data. *Science, AAAS*. 309.
- Peters, D. P. (2008). "Ecology in a connected world: A vision for a "network of networks." " *Frontiers in Ecology and the Environment* 6: 227.
- Peters, D. (2010). "Accessible Ecology: Synthesis of the Long, Deep, and Broad." *Trends in Ecology and Evolution* 25.
- Pfeffer, J., and Salancik, G. R. (1978). *The External Control of Organizations: A Resource Dependence Perspective*. New York, Harper & Row.
- Pfeffer, J., and Salancik, G. R. (2003). *The External Control of Organizations: A Resource Dependence Perspective*. Stanford, CA, Stanford University Press.
- Popper, Karl, "The Logic of Scientific Discovery", Basic Books, New York, NY, 1959.
- Popper, K. R. (1961, 2002). *The Logic of Scientific Discovery*, Routledge, New York.
- Popper, Karl, *Conjectures and Refutations*, Routledge, London, 1963.

- Popper, Karl, All life is Problem Solving, (1994, 1999) ISBN 0415249929.
- Porter, J., Arzberger, P., et al (2005). "Wireless sensor networks for ecology." *BioScience* 55(7): 561-572.
- Porter, J. (2010). A Brief History of Data Sharing in the U.S. Long Term Ecological Research Network. *Bulletin of the Ecological Society of America*: 14-20.
- Redman, C. L. (1999a). "Human Dimensions of Ecosystem Studies." *Ecosystems* 2: 296-298.
- Redman, C. L., Grove, J. M., and Kuby, L. H. (2004). "Integrating Social Science into the Long-Term Ecological Research (LTER) Network: Social Dimensions of Ecological Change and Ecological Dimensions of Social Change." *EcoSystems* 7(2): 161-171.
- Redman, C. L., and Foster, D. R. (2008). *Agrarian Landscapes in Transition Comparisons of Long-Term Ecological and Cultural Change*, Oxford University Press.
- Reichman, O. J., Jones, M. B., Schildhauer, M. P. (2011). "Challenges and Opportunities of Open Data in Ecology." *Science* 331: 703-705.
- Riedel, M., F. Wolf, et al. (2009). "Research advances by using interoperable e-science infrastructures; The infrastructure interoperability reference model applied in e-science." *Cluster Computing: the Journal of Networks, Software Tools and Applications* - 12(- 4).
- Roach, S. S. (1991). "Services Under Siege: The Restructuring Imperative." *Harvard Business Review* 68: 82-91.
- Robbins, R. J. (1996). *Comparative Genomics: A New Integrative Biology. Integrative Approaches to Molecular Biology.*, J. Collado-Vides, Magasanik, B., and Smith, T. F. Cambridge, Massachusetts:, MIT Press: 63-90.
- Robbins, R. J. (2011). *Data Management for the LTER: 1980 - 2010*, National Science Foundation.
- Robertson, G. P., Coleman, D. C., Bledsoe C. S, and Sollins, P. (1999). *Standard Soil Methods for Long-term Ecological Research*, Oxford University Press.
- Robertson, G. P., et al. (2007). *The Decadal Plan for LTER: Integrative Science for Society and the Environment: A Plan for Research, Education, and Cyberinfrastructure in the U.S. Long Term Ecological Research Network.* Albuquerque, New Mexico, LTER Network Office Publications Series. 24.
- Robson, C. (2002). *Real World Research*. Oxford: Blackwell.

- Romanelli, E., and Tushman, M. L. (1994). "Organizational Transformation as Punctuated Equilibrium: An Empirical Test." *Academy of Management Journal* 37(5): 1141-1166.
- Ross, B. J. (1999). *A Lamarckian Evolution Strategy for Genetic Algorithms. The Practical Handbook of Genetic Algorithms.* B. J. Ross.
- Roxburgh, S. H., Berry, S. L., Buckley, T. N., Barnes, B., and Roderick, M. (2005). "Wath is NPP? Inconsistent Accounting of Respiratory Fluxes in the Definition of Net Primary Production." *Bristish Ecological Society* 19(2): 378-382.
- Rubin, H., and Rubin, I. (1995). *Qualitative Interviewing: the art of hearing data.* Thousand Oaks, CA, Sage Publications.
- Rudestam, K. E., and Newton, R. R. (2001). *Surviving Your Dissertation: A Comprehensive Guide to Content and Process,* Sage Publications.
- San Gil, I., Hutchison, V., Palanisamy, G., Frame, M. (2010). "Metadata Activities in Biology." *Journal of Library Metadata* 10(2-3): 99-118.
- Savolainen, V., Cowan, R. S., Vogler, A. P., Roderick, G. K., Lane, R. (2005). "Towards writing the encyclopedia of life: an introduction to DNA barcoding." *Philosophical Transactions of the Royal Society* 360: 1805-1811.
- Schroeder, R., and Fry, J. (2007). "Social Science Approaches to e-Science: Framing and Agenda." *Journal of Computer-Mediated Communication (JCMC)*(12): 563-582.
- Science (2011). *Challenges and Opportunities : Dealing with Data (Special Section).* Science, AAAS. (331): 692-693.
- Servilla, M., Costa, D., Laney, C., San Gil, I., Brunt, J. (2008). *The EcoTrends Web Portal: An Architecture for Data Discovery and Exploration.* Proceeding of the Environmental Information Management Conference 2008 (EIM 2008): Sensor Networks., Albuquerque, New Mexico.
- Shander, C., Willassen., E. (2005). "What can biological barcoding do for marine biology." *Marine Biology Research*(1): 79-83.
- Short, N. M. (2010). *The Remote Sensing Tutorial.*
- Sheldon Jr., W. M. (2008). *Dynamic rule-based quality control framework for real time sensor data.* Environmental Information Management Conference 2008 (EIM 2008): Sensor Networks, Albuquerque, New Mexico.
- Shendure, J., and Hanlee, J. (2008). "Next-generation DNA sequencing." *Nature Biotechnology* 26(10): 1135-1145.

- Shugart, H. H., et al (1988). Report of the NSF Advisory Committee on Scientific and Technological Planning for Long-Term Ecological Research Projects.
- Siegel, J., Dubrovsky, S., Kiesler, S., McGuire, T. W. (1986), "Group Processes in Computer-mediated Communication," *Organizational Behavior and Human Decision Processes*, 37, 157-187.
- Simon, H. (2001). *The Sciences of the Artificial*.
- Smith, M. A., Fisher, B. L., & Hebert, P. D. N. (2005). "DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar." *Philosophical Transactions of the Royal Society B* 360: 1825-1834.
- Soule, M. E. (1985). "What is Conservation Biology?" *BioScience* 35(11): 727-734.
- Specht, P. H. (1993). "Munificence and carrying capacity of the environment and organization formation." *Entrepreneurship: Theory and Practice* - v17(- n2).
- Sneath, P. H. A., and Sokal, R. R. (1973). *Numerical Taxonomy*. San Francisco, Freeman.
- Stafford, S. G., Alaback, P. B., Waddell, K. L., Slagle, R. L. (1986). *Data Management Procedures in Ecological Research*. Research data management in the ecological sciences. W. K. Michener. Columbia, SC, University of South Carolina Press.
- Stafford, S. G., Brunt, J. W., Michener, W. K. (1994). *Integration of scientific information management and environmental research*. *Environmental Information Management and Analysis: Ecosystem to Global Scales*. W. K. Michener, Brunt, J.W., Stafford, S.G., Taylor & Francis: 3-19.
- Stake, R. E. (1995). *The Art of Case Study Research*, SAGE Publications.
- Star, S. L., & Ruhleder, K. (1995). *Steps Towards an Ecology of Infrastructure: Design and Access for Large Information Systems*. CSCW '94, Proceedings of ACM, 1994 Conference on Computer Supported Cooperative Work, ACM.
- Sterling, E. J., et al. (2003). *Why Should You Care About Biological Diversity*. SSI 2003, Center for Biodiversity and Conservation American Museum of Natural History.
- Stohr, E. A., and Zhao, J.L. (2001). "Workflow Automation: Overview and Research Issues." *Information Systems Frontiers* 3(3): 281-296.
- Strassman, P. (1997). *Computers are yet to make companies more productive*. ComputerWorld.
- Strauss, A., and Corbin, J. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*.

- Sugden, A., and Pennisi, E. (2000). "Diversity Digitized." *Science* 289(5488): 2305.
- Sullivan, A., and Sheffrin, S. M. (2003). *Economics: Principles in Action*. New Jersey, Prentice-Hall.
- Summerton, Jane, ed. (1994). *Changing Large Technical Systems*. Boulder, CO: Westview Press.
- Thompson, J. N. (1994). *The Coevolutionary Process*. Chicago, IL, The University of Chicago Press.
- Van Blerkom, L. M. (2003). "Role of Viruses in Human Evolution." *Yearbook of Physical Anthropology* 46: 14-46.
- Van Vleck, T. (2001) *The History of Electronic Mail*.
- Voss, V., Procter, R., et al (2007b). Sustainability of e-Infrastructures (for the Social Sciences). *e-Social Science 2007*, Ann Arbor, MI.
- Walsh, J. P., Kucker, S., Maloney, N. G., Gabbay, S. (2000). "Connecting minds: Computer-mediated communication and scientific work." *Journal of American Soc. Inform. Science* 51(14): 1295-1305.
- Walsham, G. (1995). "Interpretive Case Studies in IS research: Nature and Method." *European Journal of Information Systems* 4: 74-81.
- Walsham, G. (2006). "Doing Interpretive Research." *European Journal of Information Systems* 15: 320-330.
- Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., & Hebert, P. D. N. (2005). "DNA barcoding Australia's fish species." *Philosophical Transactions of the Royal Society B(360)*: 1847-1857.
- Weick, K. E. (1979). *The social psychology of organizing*. Reading, MA, Addison-Wesley.
- Wenger, E. (2002). *Communities of Practice*, Cambridge University Press.
- White, H., Carrier, S., Thompson, A., Greenberg, J., and Scherle, R. (2008). *The Dryad data repository: A Singapore framework metadata architecture in a DSpace environment*. *Proceedings of the International Conference on Dublin Core and Metadata Applications, Dublin Core Metadata Initiative and Universitätsverlag Göttingen, Berlin*.
- Winkler, A. E., Levin, S. G., Stephan, P. (2010). "The diffusion of IT in higher education: Publishing productivity of academic life scientists." *Economic Innovation of New Technologies* 19(5): 481-503.

- Williams, R. J., and Martinez, N. D. (2004). "Limits to Trophic Levels and Omnivory in Complex Food Webs: Theory and Data." *The American Naturalist* 163(3): 458-468.
- Wilson, E. O. (1988). *The Current State of Biological Diversity*. Biodiversity. E. O. Wilson, and Peter, F. M. Washington, D.C., National Academy Press.
- Woolgar, S., and Coopmans, C. (2006). *Virtual Witnessing in a Virtual Age: a Prospectus for Social Studies of e-Science*. *New Infrastructures for Knowledge Production: Understanding e-Science*. C. Hine, Information Science Publishing.
- Yin, R. K. (2003). *Case Study Research: Design and Methods*, Sage Publications, Inc.
- Yu, J. and X. Fang (2005). "The strategic roles of government in e-infrastructure development: case studies of China and Korea.(Author Abstract)." *Electronic Government* - 2(- 2).



## About the Cover

Life arose from the primordial ooze represented by the multicolor background pattern.

The relationship between Biodiversity and Genomics is captured within the computer display, which represents a component of e-Infrastructure.

The pyramid superimposed on planet Earth represents biodiversity and the macro side of biological sciences.

Both the comet-like image shooting outward and the branching pattern to the right of the pyramid illustrates genomics and the micro aspects of the continuum of life, including chemical bonds and atomic structures.

e-Infrastructure, illustrated by the computer display and computer elements (servers), represents a metaphorical window (or lens) that enables scientists to see deeper into the phenomena of the continuum of life and discover its myriad patterns.