Technovation 100 (2021) 102180

Contents lists available at ScienceDirect

Technovation

journal homepage: www.elsevier.com/locate/technovation

Coevolution of cyberinfrastructure development and scientific progress

Yan Yu [a,*], Julio E. Ibarra [b], Kuldeep Kumar [c], Vasilka Chergarova [d]

[a] School of Information, Renmin University of China, 59 Zhongguancun Street, Haidian, Beijing, 100872, China
[b] Technology Augmented Research, Center for Internet Augmented Research and Assessment, Florida International University, Miami, FL, 33199, USA
[c] College of Business, Florida International University, 309 Sanctuary Dr., Key Largo, FL, 33037, USA
[d] Center for Internet Augmented Research and Assessment, Florida International University, Miami, FL, 33199, USA

ARTICLE INFO

Keywords:
Cyberinfrastructure
Data sharing
Scientific discipline
Coevolution
Longitudinal case study

ABSTRACT

Many nations have been investing significant funds in cyberinfrastructure for sciences (commonly known as "CI" or "e-science") in the last few decades. The purpose of these investments has been to advance scientific progress in the supported scientific disciplines such as biodiversity and ecological research. However, there is scare research on how these CI investments impact scientific progress. We conducted a longitudinal case study of CI development and scientific progress based on the Long Term Ecological Research (LTER) program in the United States. We contribute to filling in the research gap by finding CI investments result in impacts both internal and external to the particular discipline of biodiversity. We also find a feedback effect from the discipline to CI development, such that the science and its supporting CI coevolve. The coevolution of CI development and the scientific discipline hinges on the "linking-pin" of data infrastructure. The large-scale data sharing within the discipline and across disciplines, in turn, is contingent on CI governance and how data are structured. The identification of the synergistic coevolution of CI and science enriches coevolutionary theories as well as CI development research. Accordingly, further research and policy changes relating to why and how to fund CI development for sciences, are necessary.

1. Introduction

Over the last few decades, nation states such as the US and confederations of nation-states such as the European Union, have been investing increasingly large amounts of money and research resources in developing the Cyberinfrastructure (CI) for sciences (EC, 2013; Farley et al., 2018; NSF-US, 2006). CI for sciences, also known as e-science, refers to the socio-technological infrastructure that integrates information technologies (IT) with human resources and organizations to advance the discipline and practice of these sciences. These infrastructures are typically designed for the creation, dissemination and preservation of scientific data and knowledge in the "digital age" (Atkins et al., 2003). The investments in CI are being made with the hope that they will transform scientific disciplines and result in revolutionary scientific progress that may ultimately help humanity find solutions to the substantial challenges it faces (e.g., the looming ecological crises).

However, at this time not much is known about the effectiveness of these CI investments in bringing about changes in scientific disciplines, much less about the impact of CI development on the nature of any discipline's progress. Scientific progress comes from a combination of increased rates of scientific discovery, and an increase in building cross-disciplinary connections. Cross-disciplinary connections are supposed to lead to further increases in the rate of scientific discovery, as interdisciplinary research requests a complex process of knowledge integration (Avila-Robinson and Sengoku, 2017). But, the link of increases in intra- and cross-disciplinary connections leading to increases in the rate of scientific discovery, at present, is only a conjecture, based upon preliminary anecdotal data, and have yet to be confirmed through empiricism. Therefore, our fundamental research question is *"How does CI development impact scientific progress in the particular discipline such as biodiversity and ecological research?"*

The extant literature on IT for sciences is mainly concerned about research productivity at an individual level. On the one hand, IT investment and development can elevate scientific productivity (usually measured by published papers, patents, etc.) (Ding et al., 2010; Hesse et al., 1993; Winkler et al., 2010); on the other hand, such IT-enhanced productivity is shown still limited in the Internet era (Kabo et al., 2014; Vasileiadou and Vliegenthart, 2009). The continuing debate about the "productivity paradox" requires further investigation as to the impact of IT investment on science. More importantly, CI for sciences is much

* Corresponding author.
E-mail addresses: yanyu@ruc.edu.cn (Y. Yu), Julio@fiu.edu (J.E. Ibarra), kuldeep.x.kumar@gmail.com (K. Kumar), vchergar@fiu.edu (V. Chergarova).

https://doi.org/10.1016/j.technovation.2020.102180
Received 13 August 2020; Accepted 19 August 2020
Available online 9 September 2020
0166-4972/

more complex than the discrete IT tools adopted by scientists. Thus, CI impact should go beyond improving the individual productivity. In particular, Lane et al. (2015) explicitly claim that the value of integrative data infrastructure on the science of science policy is remarkable. Hence, it is imperative to reveal the impact, as well as the influencing mechanism of CI investment and development, at a discipline level, so as to offer important implications for digital infrastructure development and science policy making.

Meanwhile, the extant studies on IT infrastructure and digital infrastructure are focused on the value of IT for organizations and rely on an evolutionary view to explain the infrastructure development (Henfridsson and Bygstad, 2013). This implies the necessity to investigate IT for science as well as the necessity to take alternative perspectives. IT infrastructure has been viewed as a strategic asset for organizations and can exert strategic impacts when the infrastructure development is aligned to strategic imperatives in organizations (Broadbent et al., 1999). The value of digital infrastructure is shaped by improved organizational agility, responsiveness, coordination capability, and innovation (Sambamurthy et al., 2003; Urbinati et al., 2019). Also, advanced technologies such as big data, cloud computing, artificial intelligence, Internet of things, and so on have been recognized having substantial impacts on transforming sciences (Chen et al., 2012; Yang et al., 2017). For instance, a computing grid infrastructure at CERN (European Organization for Nuclear Research) enabled digital coordination within the particle physics community (Venters et al., 2014). Yang et al. (2011) claimed that spatial computing would become a core technology that drives fundamental physical science in the 21st century. Big data and cloud computing have pointed to potential solutions for various digital earth problems in geoscience and relevant domains such as social science, astronomy, and business and industry (Yang et al., 2017). The cloud enabled spatial data infrastructure architecture is proposed to resolve the challenges such as data sharing and management, inter-operability, and global collaboration in geospatial community (e.g.,Tripathi et al., 2020; Yang et al., 2019). In bioscience, technological solutions include the development of open code- and data-sharing platforms, flexible statistical models that can handle heterogeneous data and sources of uncertainty, and cloud-computing delivery of high-velocity computing to large-volume analytics (Farley et al., 2018). In the spirit of open science, the ECO-OP (ECOsystem and interOPerability) project in US, the European iMarine project, and the eReefs project in Australia aimed to develop cyberinfrastructure for marine IEAs (Integrated Ecosystem Assessments) that involved analysis of natural and socio-economic information based on diverse and disparate sources of data, requiring collaboration among scientists of many disciplines and communication with other stakeholders (Beaulieu et al., 2017). In the epidemiology discipline, Ribes and Polk (2014) emphasized the "flexibility" that makes the long-term research infrastructure of MARC (Multicenter AIDS Cohort Study) responsive and resilient to sociotechnical, technoscientific and institutional changes. Although technological advancement is impressive, more solid research to uncover how the development of cyberinfrastructure exerts value for science is definitely needed.

Further, IT infrastructure evolution has been characterized by complex interdependencies and interactions between socio-technical elements (Braa et al., 2007; Henfridsson and Bygstad, 2013). Therefore, the evolutionary view should be broadened into a coevolutionary view when studying CI development, according to Kauffman's (1993) complexity theory. As a complex system, CI should be decomposed into multiple components (Atkins et al., 2003). The scientific discipline and the supporting CI are two separate systems on the surface but actually foster complex interactions within and across the systems at multiple levels to constitute the scientific world. Such complexity is in line with Baum's (1999) synergistic coevolutionary system, in which the agents are reinforced by each other. To the best of our knowledge, we are pioneers in trying to disclose the dynamism of the complex coevolution between scientific disciplines and the supporting CI and its impacts.

Specifically, we study the relationship of CI development and scientific progress of biodiversity (both intra- and cross-disciplinary progress), in the LTER (Long-Term Ecological Research) program created by the National Science Foundation in U.S.A. (US-NSF). The US-LTER program includes the scientific discipline of biodiversity, as well as the CI supporting it. We choose to study US-LTER for two reasons. First, the establishment and duration of US-LTER, since its emergence in 1980s, coincides with the evolution of biodiversity as a scientific discipline. Over the following three decades and beyond, US-LTER and its investments in cyberinfrastructure have been one of the longest continuously federally funded biodiversity programs. The US-LTER program now extends to 30 research sites, involving a wide range of types of ecosystems distributed across the continental United States, other territories and globally over other countries. Unlike much longer-standing disciplines such as high-energy physics or astronomy, whose origins and history may be lost in antiquity, a study of the impacts of CI investments on the discipline of biodiversity can be conducted over the life span of the discipline. Second, because biodiversity is a relatively new discipline, documentary evidence as well as access to scientists who have lived through this evolution and remember their experiences, are available for the duration of its evolution. This makes it possible for us to collect a comprehensive set of their first-hand experiential data on the evolution of the US-LTER cyberinfrastructure, and the associated evolution of the scientific discipline of biodiversity.

The rest of this paper is organized as follows. We first review the coevolutionary theories and explain a conceptual lens. Next, we describe our research methodology and briefly depict the three-plus decades of development of the US-LTER program for biodiversity. Then, we build a research framework to show the coevolution of CI and sciences, and discuss four key findings based upon the analysis of the case. Finally, we outline the implications as well as the limitations of this research, and future directions it suggests.

2. Coevolutionary theories

In this research we use a coevolutionary perspective on CI development and scientific progress. Therefore, this section outlines the underlying theories of coevolution and their application to organizational growth.

Conventional wisdom regarding organizational evolution (Aldrich, 1979; McKelvey, 1982) is rooted in Darwinian concepts of "natural selection" (Darwin, 1859; Thompson, 1994). The Darwinian process of natural selection postulates that selection and survival processes drive out organisms that are less fit, i.e. those are less suited to their environment. This leaves "order" as a consequence of the more fit organizations' survival, ultimately leading to the survival of the fittest. Coevolution occurs when the direct or indirect interaction of two or more evolving agents produces an evolutionary response in each other (Van Valen, 1983). In the organizational context, Lewin and Volberda (1999) defined coevolution as "the joint outcomes of managerial intentionality, environment, and institutional effects" (p.526). Change may occur in all interacting populations of organizations, driven by both direct interactions and feedback from the rest of the system (Volberda and Lewin, 2003).

Drawing on complexity theory, Kauffman (1993) proposed the coevolutionary complexity model in which "organisms do not merely evolve, they coevolve both with other organisms and with a changing abiotic environment" (p.237). According to the second law of thermodynamics (Prigogine, 1962, 1980), importation of energy into the system creates evolutionary structures and keeps the emergence and maintenance of structures in states far above entropy. The coevolution of the agents in complex systems describes the agents' co-learning and adaptation. This learning and adaptation predict a particular pathway into the future, rather than converge on to some generic equilibrium structure. In this system, successful agents handle environmental pressures by importing the energy required to sustain the system's viability.

The influx of new energy (e.g., investment, technology, knowledge, etc.) into a system exceeds the limits of the old conventional structure's capacity to contain it. The old rules are challenged, the old structures are broken, so leading to the emergence of new opportunities and the rise of a new generation of structures, processes, and governance. During the coevolution of agents, time exerts unrelenting forward pressures on successful agents to continuously adapt. In this sense, selection and adaptation are not opposing forces but are fundamentally interrelated and coevolving (Volberda and Lewin, 2003). When organizations coevolve with changes in the environments requiring adaptation, organizational forms will mutate and new forms will emerge from the existing population of organizations (Lewin et al., 1999).

Further, the coevolutionary effects take place at multiple levels, such as the community, population, organization, and intra-organizational levels (Baum and Singh, 1994). McKelvey (1997, 1999) distinguished micro-coevolution, which emphasizes the coevolution of intra-organizational resources, dynamic capability, and competences, from macro-coevolution, which emphasizes the coevolution across organizations, population and communities. In light of the whole-part structure, coevolution is nested, and lower levels of coevolution always occur in the context of higher levels of coevolution (Baum, 1999; Baum and Singh, 1994; Lewin and Volberda, 1999). The nested perspective of coevolution has been used in research, but related studies are still sparse. By conducting case studies in the publishing industry, Van den Bosch et al. (1999) described the building blocks within firms that continuously updated knowledge bases and combinative capabilities for absorptive capacity (micro-level coevolution), as well as the coevolutionary paths between firms and their turbulent knowledge environment, all of which has resulted in the transformation from traditional publishing simplicity to a multimedia industrial complex. Dong et al. (2016) illustrated a nested structure of coevolution in a Chinese context; that is, coevolution of micro states within an organization emerges in the coevolution of the organization and its environment.

Coevolution also occurs in different forms or configurations. Depending on the form of coevolution, an increase in fitness by one agent alters the fitness of the other agent in the system in different ways. Based on Kauffman's coevolutionary complexity model, Baum (1999) differentiated three typologies of coevolutionary configurations, comprising (1) the independent configuration, in which an increase in one agent's fitness does not affect the other's fitness; (2) the competitive configuration, in which the increase in one agent's fitness results in a decrease in the other's fitness; (3) the synergistic configuration, in which an increase in one part's fitness results in the other's fitness being increased as well. While competitive coevolutions will result in suboptimizations or local optimization of parts, synergistic coevolutions can increase the likelihood of global optimization. As CI investments are hoped to enhance scientific progress, a scientific discipline and its CI are mutually reinforced rather than being competitive. Therefore, it is important to uncover how the synergistic coevolutions are produced and how these complex coevolutions result in global impacts at the discipline level.

3. A conceptual lens for CI impacts on scientific disciplines

The coevolutionary theories imply the potential for complex interactions among agents or components in a system. Thus, we first identify key concepts of a scientific discipline and CI, based upon the extant literature on scientific progress and disciplines (Kuhn, 1996; Popper, 1959, 1963) and on CI (Hughes, 1987; Star and Ruhleder, 1996). Then, we adopt the coevolution view to initially propose a possible interaction between a specific scientific discipline (e.g., biodiversity) and CI (e.g., LTER). The coevolutionary journey of any complex system requires energy importation (Prigogine, 1962, 1980). Therefore, CI investment is regarded as an initial stimulus, spurring the coevolutionary process of the development of CI with a scientific discipline, in hope of generating evolutionary and revolutionary scientific progress.

When the history of evolution and coevolution of scientific disciplines and the supporting CI is retrospected, the lens of imprinting is also helpful in explaining how the components of science and CI interact to shape scientific progress. Stinchcombe (1965) was the first to advance the concept of imprinting in organizational studies. Based on a systematic literature review, Simsek et al. (2015) suggested that imprinting involves three successive processes: (1) genesis, whereby the characteristics of the imprinters (the sources of imprints) and the imprinted (the target entity that bears an imprint) interact in ways that culminate in the formation of an imprint; (2) metamorphosis, the evolutionary processes or dynamics by which imprints persist, amplify, decay, and/or transform; and (3) manifestations, the influence of the imprint on entity characteristics, and the direct and indirect effects of the imprint on entity survival and performance. CI investment and CI development act as imprinters that interact with the imprinted disciplines. We further postulate that the coevolutionary processes and dynamics between CI and scientific disciplines, instead of the evolution of CI and science independently, embody the imprinting process (i.e., metamorphosis). Their influences are manifested by evolutionary and revolutionary scientific progress.

3.1. Components of scientific disciplines

The aspects of science disciplines describe how science is done (the modes of practice in a science discipline), how science is thought about and discussed (philosophy of science), and the resultant scientific progress. The practice of science disciplines consists of three subcomponents: *community of scientists*, *problems and puzzles* that community members work on, and *methodologies* used by that community to conduct research to solve problems and puzzles (Kuhn, 1996; Popper, 1959, 1963). These three sub-components collectively describe the concepts underlying how science is done, and they belong to an over-arching concept – the *paradigm* of the research discipline.

A community of scientists is a body of scientists that practices research as defined by its current problems and puzzles. It is grounded on the discipline's existing scientific achievements and knowledge (Kuhn, 1996). *Problems and puzzles* define the propositions and questions used to articulate the paradigms from which they were derived. An anomaly is a particular kind of problem that appears as an inconsistency between observations and the discipline's commonly accepted knowledge (or current theories) (Graham and Dayton, 2002). Communities of scientists usually define problems and puzzles within the framework of an adopted paradigm. Problems within the scope of a paradigm are normally the only ones recognized by its scientific community as legitimate scientific problems. 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).

A scientific discipline's methodology either tacitly or explicitly describes a legitimate and accepted way in which science is practiced in a discipline. A methodology includes (1) a criterion for choosing problems that have solutions guided by a paradigm, and (2) a set of agreed upon rules to determine both "what are acceptable solutions" and "how to obtain them" (Kuhn, 1996). The criterion results in a set of problems that the community of scientists accepts as scientific, and which therefore encourages its members to further articulate and refine the paradigm from which they were derived. The rules about how to obtain acceptable solutions influence the design and acceptability of the instrumentation used for aiding observation, collecting data, and solving puzzles. These rules are typically expressed as explicit observation and data collection methods attached to statements of scientific concepts and theories.

Paradigms (philosophy of science) and *paradigm shifts* reflect how science is thought about. 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 (Kuhn, 1996). A period of paradigm shift commences when a paradigm remains in crisis resulting from important 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. During a paradigm shift a science discipline is (a) more 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. Changes to a paradigm modify the accepted problems and puzzles of a discipline, and their solutions.

Scientific progress has been classified into either *evolutionary* or *revolutionary* progress by the U.S. National Science Board (NSB). *Evolutionary scientific progress* is defined as incremental advances in scientific understanding that build upon the results of prior scientific knowledge. Using hypotheses and theories from a currently prevailing scientific paradigm, evolutionary progress attempts to gradually refine the existing hypotheses and theories, thereby extending the life of the prevailing paradigm. By contrast, *revolutionary scientific progress* takes place when scientific understanding advances dramatically, increasing the rate of discovery of new ideas, solutions and systems. Such progress transforms science by overthrowing entrenched paradigms and generating new ones. Typically, revolutionary scientific progress is accompanied by paradigm shifts. With revolutionary progress, there is presumed to be an increased opportunity for rapid innovation and its attendant buoyant economic development and growth.

3.2. Components of cyberinfrastructure

Infrastructure consists of heterogeneous interacting artifacts that are adapted to work together and interoperate to exchange information (Law, 1987). The pure IT infrastructure refers to an arrangement of shared technical components and IT services: platforms, networks and telecommunications, data, and software applications (Duncan, 1995; Roberts and Grover, 2012). Broadbent and Weill (1997) viewed IT infrastructure as a broader concept and decompose IT infrastructure into its technical and human components. Similar conceptualization of cyberinfrastructure is also held for scientific disciplines. Atkins et al. (2003) defined CI as socio-technological infrastructures that integrate IT with human resources and organizations to advance the discipline and practice of sciences. Calls for new cyberinfrastructure to support data sharing and new interdisciplinary approaches have led to significant new funding for network-based services, such as grid computing, data federation and community building (Edwards et al., 2009). Therefore, we conceptualize CI to consist of three aspects, comprising technical infrastructure, data infrastructure, and socio-organizational infrastructure. The interaction of both social and technical components results in CI development (Star and Ruhleder, 1996).

Technical infrastructure. This is composed of *physical objects*, *technologies*, and *instrumentation*. *Physical objects* can either be artificial, or natural. *Artificial physical objects* such as museums, laboratories, field stations, and collections of research specimens can serve as containers of natural and other artificial ecological objects. In the context of biodiversity, *natural objects* are typically preserved or living specimens. *Technologies* are artifacts used by scientists to conduct their research on physical objects. For example, computers in the pre-Internet period were used to computerize scientific instrumentation, such as telescopes in astronomy or microscopes in biology, and resulted in increases in rates of discovery as compared to research using basic non-computerized telescopes or microscopes (Robertson, 2003). Modern arrays of optical computer-augmented telescopes have contributed to revolutionary improvements in astronomical observations. *Instrumentation* represents a special class of technological objects that measure or observe natural phenomena. Instruments, such as a thermometer or a distance measurement-device, provide precise quantitative measurements that can be measured using specific measurement scales. Not all instrumentation provides quantitative data. Instrumentation is built using various technologies, such as audio microphones or video cameras, and gives scientists enhanced capabilities for observing, discovering, and recording natural phenomena.

Data infrastructure. This refers to the resources orchestrated for the generation, collection, archiving and transmission of data. *Data* is typically a codification of measurements or observations, usually in a quantitative form, using one or more symbols, which can be interpreted and given meaning (Churchman, 1971). However, analog or qualitative observations, such as the distress calls of elephants or baboons, or the colors of an ocean fish, may also be digitized for storage and for transmission through digital means. In a codified (object) form, data can be transmitted across some telecommunications media, such as the Internet and the high capacity Data Grid. Data center technologies should be regarded as an essential aspect of IT infrastructure and the associated governance (Tallon et al., 2013; Weill and Ross, 2004).

The evolution of science, including both natural science and social science, demands the acquisition and integration of vast amount of data of many types, at multiple scales in time and in space. The data infrastructure must be connected, sharable, compactable, and reliable to support the actions of an increasingly large community of scientists. In the biological field, GenBank, the NIH (National Institutes of Health) genetic sequence database, is designed to provide and encourage access within the scientific community to the most up-to-date and comprehensive DNA (Deoxyribonucleic acid) sequence information. The evolution of GenBank offers evidence that technological advances and cultural metamorphoses generate paradigm shifts in science. With a tug from software to manage genomic data online and a push from publishers unwilling to continue editing and printing the growing volume of gene sequences, a robust data repository for gene sequences was born. After almost 30 years, registering gene sequences and sharing them broadly is now the norm, and is recognized as fostering one of the greatest scientific revolutions of the past century (Reichman et al., 2011). In the ecology discipline, Reichman et al. (2011) also urged the establishment of well-curated, federated data repositories that will provide a means of preserving data while promoting attribution and acknowledgement of its use.

Socio-organization infrastructure. This consists of *process*, *organization structure* and *governance*. *A process* is a way of organizing work and resources (people, instrumentation, technology, etc.) (Kettinger et al., 1997; Sharp and McDermott, 2008). Process infrastructure supports coordination and exchange of resources between aspects of a science discipline, data, and hard infrastructure categories of cyberinfrastructure. *Organization Structure* is a representation of the organization's components and their relationships; whether of a physical organization, a virtual organization, or a hybrid (DeSanctis and Poole, 1994). Scientific organizations are referred to as scientific communities. In cyberinfrastructure, organizations are constantly changing and emerging. Owing to environmental pressures, scientific communities are undergoing continuous change or adaptation. New organizational components often emerge or come into existence, while other components die out as a result of environmental changes and transformations of existing research domains. *Governance* aims to create structure, and steer and manage the activities that lead to coordination or order between the parts of the organization (Thompson, 1991). MacLean (2004) conceptualized governance as a range on a continuum from "hard" governance structures, such as laws and regulations, to "soft" forms of governance structures, such as standards, policy coordination and voluntary cooperation. As scientific disciplines evolve and change over time, and across geographical and political regions, both hard and soft governance structures are also evolving. Regarding IT infrastructure development, the governance of both physical IT artifacts and nonphysical or information artifacts is required (Tallon et al., 2013).

3.3. Stimuli of the evolving CI for science

According to the complexity science of coevolution, importation of new energy is essential for the activation of coevolution in a system. Prior organizational research implies stimuli can be external and internal and can activate various processes within or across organizations (Zahra and George, 2002). External stimuli such as radical innovations, technological shifts, emergence of a dominant design, and changes in government policy, etc., may influence the future of the industry within which organizations operate (Bower and Christensen, 1995). Internal stimuli can be in the form of organizational crises or important events that redefine an organization's strategy. Kim (1998) demonstrated that a crisis, although a negative event, can intensify an organization's efforts to achieve and learn new skills and to develop new knowledge.

As an external stimulus, CI investments are intended to create and support a shared national cyberinfrastructure, thereby supporting basic Research and Development (R&D) and Applied Research. We conjecture that the majority of stimuli in IT investments are initially for the development of technological infrastructure. We collected data on investments in the United States from reports of the federal agencies involved in the support of computing and communications for science research. These reports show that the US Federal CI investments were continuously growing. While the early CI investments were 638.3 million dollars in 1992, the requested CI investments reached 5277.5 million dollars in 2019. The CI investments were largely in technological components, later, the data infrastructure, its investments in human resources and other social-organizational aspects have been limited. It is noteworthy that the investments in data infrastructure and socio-organizational infrastructure have increased in a faster rate since 2015.

Advances in technology infrastructure have changed the ways in which biodiversity scientists work and conduct research. Given that scientific inquiry is highly dependent on instrumentation, physical materials, knowledge, and human resources, CI investments have played a particularly important role in the production of both biodiversity knowledge and building the discipline's longer-term dynamic capabilities. Prior studies have demonstrated that IT investment and deployment have a positive association with research productivity in the disciplines of oceanographic science (Hesse et al., 1993) and life science (Ding et al., 2010; Winkler et al., 2010).

In addition to CI investments, other types of stimuli may also emerge during CI development and play different roles from such investments. For instance, technological advancement in IT may influence the scientific community in which scientists perform collaborative work, and thus influence the future of their discipline. The problems and puzzles, especially anomalies, that emerge within a traditional paradigm, can precipitate a crisis that activates coevolution between the scientific discipline and its CI as well as coevolution across different scientific disciplines, thus leading to the achievement of evolutionary or revolutionary progress. To the best of our knowledge, these stimuli are seldom discussed in the extant literature, thus requiring more research exploration and validation in the science context.

When we consider the LTER program as a complex system, we postulate that the scientific discipline and its cyberinfrastructure coevolve. Evolution in CI influences how the components of the scientific discipline evolve; and as these components evolve, they necessitate improvements in the cyberinfrastructure that lead the CI components to further adjust and evolve. The influx of new energy (e.g., CI investment stimulus, technological advancements, problems and puzzles, etc.) into the system (e.g., biodiversity community and LTER program) exceeds the limits of the old structure's capacity to contain it. The old structures are then broken, thus leading to the emergence of new opportunities and the rise of a new generation of scientific paradigms. Funding agencies hope that such coevolution of the scientific discipline and its cyberinfrastructure will lead to evolutionary and revolutionary progress in science.

4. Methodology

In order to examine the coevolution of scientific disciplines and their infrastructure as well as the coevolutionary effects of this coevolution on scientific progress, we adopt the methodology of "Interactive Model of Research Design" (Maxwell, 2012). This methodology is an ongoing process that involves tracking back and forth between the different components of a study, assessing the implications of their goals, conceptual framework, research questions, methods, and validity threats to each other. Observing that the US-LTER program is longitudinal and rich, we adopt the process thinking that focuses attention on how and why things emerge, develop, or terminate over time (Langley et al., 2013). Our unit of analysis is the process of coevolution of biodiversity and its cyberinfrastructure. Mixed methods combining interviews, observations, and archival data are used to collect the current and past records of the history of biodiversity and its supporting CI. These records underly this study's in-depth examination of the coevolutionary process. Biodiversity is a relatively young discipline; consequently we can gain access to the first-hand experiences and memories of living and recent biodiversity scientists.

Our methods for data gathering are organized along the principal activities suggested by Walsham (2006) and Yin (2009). First, we decided upon *the style of researcher involvement*. This involvement, in our case was a spectrum having external observers at one end and fully involved participant researchers at the other. With the help of one of the authors as internal participant, we were able to *gain and maintain access* to local, as well as to a geographically diverse set of researchers and documents from a variety of sources.

Second, we *collected field data by a combination of interviews, document collection, and observation*. Qualitative interviews involved an exchange of information in the form of interactive dialogues between the investigators and key informants. There were 14 informants, comprising 10 senior scientists who had been practicing science in their community for the past 30-plus years and 4 LTER program directors who were involved in the program design and management. All were considered to be knowledgeable sources of information about the evolution of their discipline from the aspects of both science and its infrastructure. The interview questions were semi-structured with open-ended questions asked to elicit information about communities of scientists, their problems and puzzles, contemporary (evolutionary) events, and the pressures being faced by the community. Each interview lasted from 40 to 90 min, was recorded using a recording device, and transcribed later by a professional transcriber. Some informants were interviewed multiple times. Sample interview questions are listed in Appendix A.

To round out our knowledge about the coevolutionary process, we collected large amounts of documentation, including *strategic documents* in the forms of reports, bulletins or policy statements from scientific institutions; *programmatic documents* from NSF that describe the social and institutional environment of biodiversity as a science discipline; *peer-reviewed documents* produced by the members of a discipline; and *informational documents* used to record information concerning events, activities, discussions, decisions, outcomes, etc. Langley et al. (2013) postulate that archival data are particularly suitable for tracing event chronologies and discourses over long or very long periods of time. Moreover, the researcher involved directly observed the events arising from the LTER program and the work groups in the situated site, and also organized two workshops in which 38 scientists and funding agencies participated. Observations from workshop discussions and memos and informal interviews provided additional materials about what had happened, and which solutions were designed for the LTER program. We summarize all the data sources in Appendices A and B.

For data analysis in this interpretive inquiry, we adopted a strategy suggested by Eisenhardt (1989) and extended by Walsham (2006). For process data, Langley (1999) further suggested seven strategies that could be used in combination for structuring materials. In particular, the grounded theory that involves data-driven categories and the alternate templates that involves theory-driven constructs, considered as grounding strategies, can be combined to contribute to the construction of narratives (i.e., narrative strategy) and the analysis of phases (i.e., temporal bracketing strategy) (Langley, 1999). Therefore, we approached data analysis as an interplay between the field data and the conceptual lens described above. As we were dealing with a single overall case, biodiversity, we used the temporal bracketing strategy to decompose the longitudinal case into three successive periods and constructed the narratives incorporated with events, practices, and quotes within the overall case. Our findings are grounded in data (Strauss and Corbin, 2008) and are used iteratively to develop a new conceptual framework. Synthesis based upon interviews, direct observations and documents was employed for triangulation and validation of our conclusions.

5. Case description: US-LTER and its development

In this section, we briefly describe the set-up of US-LTER program and its evolution over three plus decades. In 1980, the US-NSF awarded six research grants and launched the US-LTER program at six initial sites. The US-LTER program has now since expanded to 30 research sites. A comprehensive inventory identified 1268 contemporary Biological Field Stations (BFS), located in 120 countries (Tydecks et al., 2016). The goal of the US-LTER program was to stimulate long-term, comparative research on ecological processes in a network of geographically distributed and ecologically diverse field sites (Callahan, 1984). Based on the analysis of qualitative data collected from various sources, we partitioned the US-LTER development period into three stages: (1) closed science and research; (2) data-driven research; and (3) synthesis science.

5.1. Closed science and research in decade I (1980–1990)

The philosophy in Decade I was mainly kept as closed science, in which scientists in different communities worked independently and mutual influences across communities were rare. The closed science fits to Baum's (1999) independent configuration for coevolution. Prior to the 1980s, the lack of adequate data and analytical tools severely limited the research that could be conducted involving spatial and temporal scales of biodiversity. Biodiversity-ecological research was normally conducted on ground plots of 1 m squared or smaller, over short periods of time, usually one to three years (Gosz et al., 2010). This severely limited the scientists' ability to conduct long-term scientific research and develop long-term conclusions, especially when the effects of the phenomena of interest typically accumulate over long distances and long periods of time (Michener et al., 1997).

By the 1980s, availability of new data (e.g., remote sensing data), new technology (e.g., GIS, Geographical Information Systems) and tools for analysis and modeling enabled scientists to increase the scale of biodiversity-ecological research (Stafford et al., 1994). By the early 1980s, the community of scientists involved with long-term ecological research was arguing for the capability to expand the scale of inquiry from the small area of a sample plots, habitat patches and small watersheds of traditional ecology to the much larger scale of landscape, geographic region, continent, ocean and even the entire Earth (Callahan, 1984).

Stimuli in Decade I. Stimuli for changing the LTER Network, its infrastructure, and the ecology discipline came from two sources: government and technology. Government was concerned with global change issues, biodiversity and sustainability (Callahan, 1984). Technology created an enormous supply of primary data, and tools for organizing and analyzing them. Availability of GIS technology and data from remote sensing were important external CI stimuli to LTER.

Technical and Data Infrastructure Development in Decade I. While investments in GIS technology, by transforming ecological data into geo-referenced data, led to dramatic increases in the quantity of data, investments in data infrastructure escalated the development of tools to analyze this primary geo-referenced data and transform it into ready-for-science useable data (Stafford et al., 1994). By 1985, owing to new data collection and management capabilities, remote sensing data available from external sources, and additional financial and intellectual stimulus from the NSF and other sponsors, the quantity of geo-referenced data was growing at a fast rate. With the availability of data, and the enhanced capabilities for organizing and analyzing large data sets, the role of data management in LTER and ecology began to change. Data were becoming central to the research process. This was also a result of a publication of a workshop proceedings volume in 1986. This volume described methods for data management and development of metadata.

Socio-organizational Infrastructure Development in Decade I. Demand for access to geo-referenced data sets increased dramatically within the LTER network, and from other communities working with long-term data. However, the sharing of long-term data sets between sites in the LTER Network and other disciplines was not progressing well. Porter (2010) attributed the notion of sharing data as foreign to most researchers in the LTER network in the 1980s and 1990s. This was a socio-organizational, not a technical issue, and data sharing was typically not customary in the biodiversity discipline. Typically, scientists would hoard their data so they could be the first to report on a data-based discovery. For example, the North Temperate Lakes (NTL) site in Wisconsin declined to share its data with the Konza Prairie (KNZ) site. Robbins (2011) observes, "Most researchers are loath to part with their data. After all, data are the raw material out of which scientific discovery, and thus scientific careers are made." The ecologists also described a similar situation:

> *"… the same thing happened with the crystallography community. The bio-chemical pharma guys pounced all over the crystallography databases, because it fed into their drug discovery and development plans. … The crystallographers push back saying, 'hey, this is not your data, it's our data!'" (Interview with an Ecologist)*

Another reason for reluctance to share data was that, along with sharing came responsibility for active participation in the cleanup, certification (of reliability and validity), and preparation of documentation (metadata) for the data to be useable by others (Michener et al., 1997; Porter, 2010). These two reasons inhibited data-sharing on any significant scale during the initial years of LTER programs.

Fortunately, by the 1990s a technological imperative began to influence the data sharing goals of the LTER community. Stimuli from IT investments had resulted in changes to data in two ways. First, from the supply-side, by adding geo-referencing attributes, consequent on their integration with remote sensing data. As well, the increased ability to manage the increased size and temporal and spatial complexity of data sets made it technically feasible to store, access, and analyze large amounts of data. Second, on the demand side, practical puzzles and real-life ecological events, ecological crises and changes were necessitating the collection and use of large data sets. 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 arising from this condition.

Scientific Progress in Decade I. Scientific discovery in Decade I primarily occurred at local LTER sites (Johnson et al., 2010). These sites focused their efforts on developing long-term research plans and sampling protocols, with an emphasis on site-specific research. Cross-site research activities and network-level activities were not much in the picture until Decade II. However, the scientific community gradually recognized that a longer-term perspective was necessary to understand the ecological response to gradual changes or rare events. The technology from LTER enhanced the instrumentational capability of the discipline. The scientists involved had similar opinions on the technological changes.

> *"The vision that Matt and his original crew of people – this goes back to the early '90s – the vision that they had was to do what now exists, and that is to say a very network-intensive, remotely operable, high speed communication, video connections between sites, and that was 15 years*

ago that they were dreaming those dreams and wrote them down on paper …" (Interview with an Astronomer)

"… the convention when people do have money, and this has revolutionized our field, is to have digital microscopes and cameras where we can take pictures of all the different taxa. Make the pictures available somehow in a database, first and then link that database on to the Web. Some places have done that really well, and it has transformed the way we can do things remotely." (Interview with a LTER Ecologist)

Findings during Decade I increased the understanding of the accumulated large effects arising from gradual changes. For example, researchers at the North Temperate Lakes LTER site reconstructed an extensive record of water clarity by obtaining the original Secchi disks and inter-calibrating them (Lathrop et al., 1996). LTER researchers found that long-term changes in water clarity were related to nutrient discharges into lake and food web structures, which increased the effect of grazing on water clarity (Hobbie et al., 2003).

5.2. Data-driven research in decade II (1991–2000)

We regard Decade II as a period of data-driven science, in which scientists and communities tended to share data for comparative research, however, such sharing was limited within a specific discipline. Different disciplines were racing to build the CI for supporting their own scientific discovery. The data-driven science developed in a local view rather than a global view, formulating Baum's (1999) competitive configuration for coevolution among scientific communities, although such competition was weak. As the data for science grew with velocity and variety, the request for cross-site comparative research and synthesis emerged in this decade.

Prior to 1991, research methods used at LTER sites were primarily driven by traditional hypotheses, typically based on situations in which very specific questions can be addressed within tightly bounded spheres of inquiry (O'Malley et al., 2009). In 1991, the marine science community joined LTER and established the Palmer research site in Antarctica. In contrast to a traditional hypothesis-driven approach, marine scientists used a data-driven research approach. This approach is based on open-ended questions and the exploration of phenomena through the use of new technologies (O'Malley et al., 2009), to develop ecological models in response to the aforementioned five core areas of research. Scientists conducted observations, including gathering remote sensing data, and constructed empirical data-based models before developing hypotheses. Integration with the marine ecology community was transforming LTER into a more diverse and complex network. This was a true paradigm shift from hypothesis-driven science to data-driven science. In 1993, the International LTER (ILTER) was also initiated, which further expanded the network of communities.

Stimuli in Decade II. The development of the Internet provided further technical stimulation for LTER. At the start of the 1990s, various technology advances affecting LTER and other science communities occurred. Use of the Internet began to accelerate with the launch of Internet-based online systems (FTP, Gopher, and subsequently the World Wide Web). LTER sites were connecting to the Internet as well, though still some scientists continued to be reluctant to allow online access to their data sets, fearing that their data could be appropriated by others (Porter, 2010).

The second stimulus to progress was internal and came from the changes to the scientific community itself. By 1993, the LTER Network had grown to 18 sites nationwide. By 1997, it had expanded to include sites in urban areas and land margins. This meant that increasingly diverse ecological systems were generating the network's local and cross-site data. The LTER community appealed for technological solutions to bridge the incompatibilities with different data schema and communication platforms (e.g., browsers) at different sites. Also, the trickle of online data sets that started in 1994 had turned into a deluge (Porter, 2010). The lack of multi-site data integration and a metadata standard for ecological data increased the costs of producing "ready-for-science" data sets.

"You must have someone who has the time to work with all those data, has knowledge of them, and knows how to digitize them. It's a huge demand. That's why it's not only costly, but it's really tough to go out and find somebody who can support that kind of activity to catch up." (Interview with a LTER Ecologist)

The third stimulus came from a special competition announced by the NSF to increase cross-site comparative research and synthesis, by applying new Internet-based technologies and tools.

Technical and Data Infrastructure Development in Decade II. The appearance of the XML markup language in 1996 helped settle issues of meta-standardization and integration of data. The adoption of XML markup language for the exchange of data in communities of science was a key event for LTER and the National Center for Ecological Analysis and Synthesis (NCEAS). It helped lead to the creation of the Partnership for Biodiversity Informatics (PBI). This partnership enabled the community to adopt machine and human parsable XML schema content standards for LTER metadata (Michener et al., 2011).

In 2000, NSF provided new funding for "The Knowledge Network for Biocomplexity". The objective of this network was to develop the work and usability tools for Future of Long-Term Ecological Data (FLED). FLED was a working group of the Ecological Society of America; its goal was to develop the metadata standard into an XML schema. Meanwhile, the release of Windows 98 with imbedded browser capability in 1998 enabled the sites from the Land Margin competition to be added to the LTER program (Robbins, 2011). Congress, at the same time, passed legislation for the Next Generation Internet Research Act to promote investment in the development of technologies to advance Internet capacity and capabilities.

Socio-organizational Infrastructure Development in Decade II. By 1993, the demand for access to data sets was increasing rapidly. 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 selected representatives from the community of scientists wrote a sample policy for future use. This sample policy recognized the need to reward scientists who were open to sharing their data, either by receiving credit within the science community through acknowledgement, citation, or co-authorship, or by receiving financial remuneration (e.g., royalties, potential future grant funding).

"We recognize the importance of ecology is long-term data sets and we have a lot of long-term data sets, but those data sets were created in the way you just described: the researcher would have taken them and done what they wanted to do with them, then put them on a shelf. This long-term data is still around … …Archibold has data from the 1930s, monthly data on water quality and air temperature, thermal, etc., all these things that we care about now. So, somebody would have to take all those data and put them into a format that we can work with digitally, possibly available on the Web and quality control them." (Interview with a LTER Ecologist)

Based on metadata conventions and common elements surveyed from each site, an initial metadata content standard was developed in 1994. This standard was later adopted by the LTER Information Management Committee. Michener et al. (2011) and Porter (2010) both emphasized the importance of allowing the data policy to gradually 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)."

Recognizing these Internet-based technology advancements and the growth of the LTER community, the National Science Foundation (NSF) announced a special competition for cross-site data comparisons and synthesis. As a part of their funding renewal proposals, sites began to compete with one another to build capabilities for producing shareable data sets. This impetus provided increased motivation for LTER investigators to publish their data sets and make them available online, thereby increasing data sharing (Porter, 2010). The LTER Governance Committee further mandated that each site should have at least one data set available online by the end of the calendar year.

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 and Foster, 2008). Traditionally, ecological and social scientists had worked in isolation from each other, remaining within the boundaries of their respective disciplines (Law, 1987). Recognition of the need to integrate data brought together social, earth, and life sciences researchers. This further increased the understanding of the human dimensions of ecological change for the LTER network.

Scientific Progress in Decade II. Unexpected events (that is, ecological surprises) driven by new phenomena, or different combinations of variables that occur on an infrequent basis, generate new questions and new discoveries (Gosz et al., 2010). The significant epidemiological discoveries in this period include the identification of Hantavirus in 1993 (Nichol et al., 1993) and its association with "El Niño" in 1998–2000 (Yates et al., 2002). These discoveries arising from the use of long-term data were later recognized by NSF as one of the 50 most significant discoveries made with NSF funding that have had the biggest impact on the lives of Americans.

Also, network-level studies using cross-site comparative methods began to appear in decade II. The findings could not have been reached with only site-level data. One such study was the analysis of variability in North American ecosystems. This study used climatic, chemical, plant, and animal measures from each of 12 LTER sites to develop general conclusions about spatial (position in the landscape) and temporal (inter-annual) variability within lake, stream, forest, grassland, alpine tundra and desert ecosystems. General and simple principles of variations evolved from the analyses of this cross-site data collected over larger periods of ecology-evolution, and these were found to be robust across ecosystem types (Gosz et al., 2010). This discovery was initially a surprise to scientists.

5.3. Synthesis science in decade III and onward (2001 to present)

In 2001, the NSF commissioned a 20-year review of the LTER program. The report contained findings and recommendations that guided the LTER Network into its third decade – "the decade of synthesis science." Synthesis refers to a process in which the data and knowledge accumulated over the past twenty years of the LTER Network program, and from current studies, are brought together to reach new levels of understanding of long-term ecological patterns and processes. The committee recommended that "LTER science" should be encouraged to become "multidisciplinary, multidimensional, scalable, information driven, predictive and model based, education oriented, and increasingly virtual and global (Robbins, 2011)." A recommendation was made to also implement a "systemic information infrastructure." This recommendation was funded in 2009. Reviewers of the LTER program recognized the potential emergence of an information infrastructure to automate parts of the process of data reuse. Hence, the synthesis science with a global view in Decade III and onward leads the coevolution of science and its CI to a synergistic configuration.

Stimuli in Decade III. In the "Synthesis" decade, the LTER scientific community developed its own internal stimuli. These stimuli encouraged scientific inquiry to become more multi-disciplinary and concerned with global issues. The increased stimuli for increased cross-site comparative research, on both the supply and demand sides, improved the scientific community's ability to link the findings at one site to findings at another. This linkage allowed for the exploration of questions at broader geographic and longer time-scales (Peters, 2008; Robertson et al., 2012). As a result, demand further increased for the network to make its primary data more directly useable by other communities (Brunt and Michener, 2009). 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 would include the capacity to discover, access, interpret and process data for facilitating the integration and synthesis of primary data.

Technical and Data Infrastructure Development in Decade III. The LTER formally adopted the Ecological Metadata Language (EML) version 2.0 as the LTER metadata standard in 2003. Available database systems were adapted to this standard for management of metadata (Michener et al., 2011). This further increased the demand for management of EML documents. Specialized database systems for the management of XML documents were adapted and developed to manage EML metadata. In 2008, in response to the recommendation of developing a "systemic information infrastructure," the LTER program continued to enhance online data exploration and discovery capabilities using new Web services technologies, such as portal technology. Using these technologies, scientists and information managers were 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, scientists demanded tools for conducting cross-site comparative research and for developing a better understanding of the long-term evolutionary processes in Earth's ecosystems. In 2009, NSF approved funds for enhancing the cyberinfrastructure capability of LTER by improving the automation of metadata generation and creation of sharable data sets.

> *"... That's what is happening now. A lot of the museum data is being repurposed or reused in new ways. New ways that had not been possible before, because of the volume, and the ease of access." (Interview with an Ecologist)*

Socio-organizational Infrastructure Development in Decade III. LTER Network data policy was modified in 2005. The new policy defines the responsibilities of a "Data-Collector." The policy dictates how long data access can be restricted (two years after collection); identifies special conditions that may allow additional restrictions for a more extended period (e.g., locations of endangered species, human confidentiality); and outlines the properties of the required metadata. With a metadata standard established, LTER focused its effort on facilitating discovery and development of comparable data sets across sites. In 2007, the LTER Cyberinfrastructure Strategic Plan (Robertson, 2007) was published. This plan articulated a vision of how LTER science could be transformed through the use of cyberinfrastructure; it provided a blue-print for stimulating further investments in cyberinfrastructure technologies and methods in the Network.

The LTER program organized a number of mini symposiums on integration of geosciences and social sciences with LTER. These symposiums included initiatives such as 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); and Ecosystem Services in a Changing World (2010). Integration of knowledge and data with other disciplines had begun.

Strengthened Synthesis Science after Decade III. Since the end of Decade III, the LTER Network Office funded the provision of critical services for cross-disciplinary research and education as well as extended communities. NSF-US also established the Dimensions of Biodiversity program by focusing on the integration of genetic, taxonomic/phylogenetic, and functional dimensions of biodiversity. This program extended to global partnerships with NASA, NSF-China, and the São Paulo Research Foundation of Brazil to support collaborative research at the national level. The CI development aimed to: (1) provide delivery of high-quality field-based and derived data products; (2) build computational environments that integrate large amounts of multi-site, multidisciplinary data in conjunction with theory, modeling, and experimentation; and (3) develop collaborative work environments that house comprehensive tools and algorithms for concept sharing, data mining, and knowledge discovery. Advanced environmental CI were developed to collect, store, retrieve, visualize, and integrate complex data streams, all of which facilitated the much requested, highly coordinated, research networks that allowed knowledge exchange among key user groups. Cross-disciplinary collaborative partnerships among the geological, ecological, and social sciences were also supported by the CI development.

Scientific Progress in Decade III and onwards. In Decade III, the number of cross-site comparative research studies increased dramatically (Johnson et al., 2010; Robertson et al., 2012). These included the long-term inter-site decomposition experiment, the Lotic Inter-site Nitrogen eXperiment (LINX), and cross-site comparative studies of plant productivity and diversity, etc. The cross-site comparative research studies added substantial value to site-based research, and provided the opportunity for developing and testing theories that led to accelerated scientific progress and broadened knowledge of ecology. Over the past three plus decades, the research effort of the LTER community has expanded from site-specific studies to the production of numerous inter-site studies and publications (Johnson et al., 2010). Progress since decade III has shown a dramatic improvement in cross-site comparative research methods, and has resulted in increased understanding of the processes and diversity of life in Earth's ecosystems, and is helping meet current and emerging environmental research challenges (from the archived documents). The changes made since Decade III confirm the emerging central role of data found in the past decades.

6. Discussion

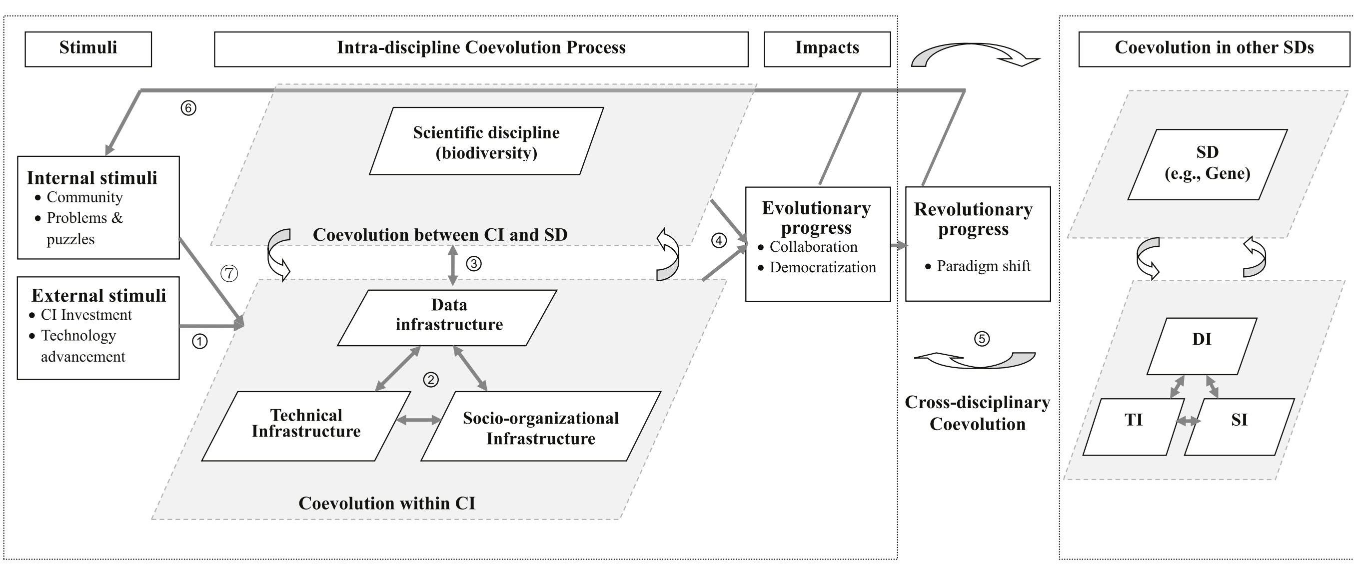
The retrospective longitudinal study of the US-LTER program demonstrates that the stimulus through investment in CI does not have direct impacts on a discipline. Instead, these impacts are mediated through a complex series of mechanisms, including the coevolution among various components of the CI and their overall coevolution with the components of science. CI development not only exerts impacts internal to the discipline; it also transforms the cross-disciplinary relationships. Based upon our case study and the analysis by aspects, we create a research framework (Fig. 1), that depicts the coevolution of complex CI development and the biodiversity discipline, and its impacts on scientific progress. We arrive at four key findings which will be elaborated in the following sub-sections.

6.1. Synergistic coevolution

Baum (1999) illustrated a variety of coevolutionary systems, in which the relationships among agents can be independent, competitive, or synergistic. In a synergistic coevolutionary system, an increase in one agent's fitness results in an increase in other agents' fitness. LTER development in the past three plus decades demonstrates the synergistic coevolution between CI and biodiversity. As coevolution progresses, CI investments and the advancement of technologies are the initial and external stimuli stimulating CI development (Arrow , Fig. 1)., or can be regarded as the sources of imprinting. Such coevolution is synergistic and occurs at three levels.

First, within the LTER infrastructure, the technical, data and social-organizational components of CI synergistically coevolve with iterative effects (Arrows②, Fig. 1). As investment in the technological infrastructure increases, the physical and computing instrumentation collects increased amounts of data. The social and governance structures evolved with the requirement that the increased amounts of data generated and collected by physical and computing infrastructures could be coordinated and linked to each other, to create an integrated database over time and geographies. The governance for data standardization, collection, and sharing among scientists are gradually established.

Funding agencies reacted to the changes and the increasing requirements of the scientific community by providing additional grants as stimuli for increased physical and computing instrumentation, which in turn created additional data. This increased the need for socio-technical structures to coordinate, govern, and integrate the increased availability of data, thereby creating a virtuous or positive feedback or coevolution cycle. While initially, at the beginning of the LTER program, the CI investment stimulus was applied directly to the physical infrastructure, it soon started creating the beginnings of the data infrastructure, and CI investments shifted to include investments in data and data-processing tools.



Note: CI (Cyberinfrastructure); DI (Data infrastructure);
SD (Scientific Discipline) sub-components include community of scientists, problems and puzzles, methodologies;
TI (Technical Infrastructure)sub-components include physical objects, technologies, and instrumentation;
SI (Socio-organizational Infrastructure)sub-components include process, organization, and governance.

Fig. 1. Coevolution of cyberinfrastructure and scientific disciplines.

"In fact, I was looking at modern analysis of that earlier and it talks about identifying the missing links for bringing the NSF vision to completion and preparing it for an operational mode, all through the network level of communication and exchange of data and information, ushering in a new generation of science education. That scale is simply not possible in today's LTER program, so it's really–I view it as really the one point in time, if you will, where we should really expect to see a transformative change in LTER functionality." (interview with a NSF LTER program director)

Second, CI coevolves with a scientific discipline, its problems and puzzles, its methodology, and its scientific community, in which data act as a "linking pin" (Arrows③, Fig. 1). The increased amounts of this data led to increased greater knowledge in the scientific discipline. The access to computer-readable data, in turn could change the nature of problems and puzzles in the discipline, thereby changing the nature of the scientific community and its practices – that is, its methodology. Over time, we have observed that the research methodology is shifting from one of physical data collection and physical experimentation to that of using computers for analyzing and experimenting with digital data. Biotechnology has been characterized as a rational, science-driven approach to discovery, which can be ascribed to innovative bioinformatics techniques and vastly increased analytical power (Gittelman, 2016).

Also, we observed that not only did the use of technical infrastructure increase the widespread access to data; the available computing power also made it possible for the researchers to analyze large amounts of data, over longer time spans and a greater range of locations. This further accelerated the generation of new knowledge from these data; which, in turn, helped transform scientific communities. This positive coevolution cycle between CI and scientific disciplines shapes the generativity of digital infrastructure (Henfridsson and Bygstad, 2013).

At the cross-disciplinary coevolutionary level, the infrastructures of interacting scientific disciplines such as biodiversity, genomics, ethnobiology, or geography; and the disciplines themselves coevolve with each other as well. Combining data from multiple disciplines is expected to lead to increasing discovery. The biodiversity and genomics are a perfect pair. While genomics can be classed as a micro-level discipline, biodiversity is a macro-level discipline. As investments in DNA sequencing technology have risen, and these sequencing systems adopted, biologists are able to generate new data for organism classification and discovery at dramatic rates and volume. The DNA barcoding technique that originated in genomics provides biodiversity scientists with a new methodology, enabling them to accelerate the process of species identification and classification. Therefore, as the CI in both disciplines evolves, niches of data are being created in both environments of genomics and biodiversity disciplines. Biodiversity scientists can use data from genomics and combine it with their data, which could lead to increasing discovery. Increases in multidisciplinary research have raised demand for inter-data network exchanges. Data providers explore "federation" between data networks to support inter-data network exchanges (Reichman et al., 2011). As such, the coevolution at the cross-disciplinary level exerts important impact on the revolution of each discipline.

6.2. *Centrality of data sharing*

A synergistic coevolutionary system needs "linking pins" to bridge gaps among different infrastructures and agents in scientific disciplines, enabling them to reinforce each other and achieve scientific progress. Our analysis further promotes data infrastructure as a "linking pin" between technical infrastructure, socio-organizational infrastructure, and science. We characterize this relationship as a "linking pin" to emphasize that changes in technology infrastructure affect data infrastructure, which in turn affects socio-organizational infrastructure, and vice versa. The coevolutionary changes between physical infrastructure and socio-organizational infrastructure were mediated primarily through changes to the data component of CI. This component includes data, meta-data models, and data-processing and analysis tools.

"Individual sites developed meta data. Sharing meta data between sites became more important. LTER Network Office is in process of building a network information system – a data management system that will ingest data from sites along with meta data. Associated metadata with standards, quality assurance are concerns of the Network Office and the individual sites. NSF has given mixed message of cross site activity …. New synthesis centers perform synthesis on existing data sets. They hired computer scientists to solve data synthesis issues in synthesizing existing data sets." (interview with a NSF LTER program director)

Our findings show that data have been central to CI development and to the goals of the LTER program, one of which has been to increase cross-site comparative research. A similar mediating pattern for data is observed when CI interventions create a need for socio-technological interventions in data governance and policies. For example, social governance interventions were introduced into the LTER community to ease issues involving data standardization and sharing, so as to stimulate cross-site comparative research and science synthesis. The LTER Information Management Committee (IMC) funded a project to develop standard ways of exchanging metadata that would be both human- and machine-readable, and the LTER Network Office created policy for rewarding scientists who were open to sharing their data. Further, in LTER, curation activities play an essential role in data discovery and retrieval, quality control, and reuse over time (Ahn et al., 2016, 2015). The newly-appreciated central role of data fits into the emerging epistemology and paradigm shift driven by big data and data analytics in a number of disciplines. Another NSF LTER program director indicated that,

"… Where it [LTER] becomes sort of viewed, if you will, as an observatory network, of course, is that if you take the same measures which you might be taking data to test hypotheses in the experiment, but if you're doing this over 30 years, yeah it becomes observational data defacto …"

"… It has only really been in say the past five or six years that we at NSF and the review community had started saying in reviews of renewal proposals that hey this has to change or NSF will have to rethink its funding of this site because where's all the data? And now I can say that it's really common that even the PIs at the sites now are all involved with information management because they want the data. They need the data. They want to have their data out there. They recognize the format. They know what the issues are and they brought out their information managers to and recognize how critically important they are and the whole game has moved up …"

Velu et al. (2013) have attempted to explore the contexts for centralization or decentralization of data management. Their findings, though in organizational contexts, offer insights into the management of data infrastructure for scientific disciplines. The centralization and standardization of data infrastructure is appropriate and beneficial to all stakeholders in LTER, as scientific discovery in biodiversity faces high levels of uncertainty and varieties. A senior scientist interviewee indicated that:

"… And so people who are there associated with that institution and continue using the institutional governance and institutional policies or whatever for accessing that instrument that owned and controlled by the institution … but if they wanted to work with an instrument that was somewhere else in the continent they can still do that …"

6.3. *Synergistic coevolution of CI leading to evolutionary progress*

Our analysis allows us to build a chain of evidence linking investments in IT infrastructure development to changes in the practice of science and the scope of scientific inquiry, resulting in evolutionary progress within the scientific discipline. The synergistic coevolutions within CI and within the particular biodiversity mainly contribute to the evolutionary progress of this scientific discipline. CI development enabled changes of practices in biodiversity, shaped by the enhanced collaboration and democratization of the whole biodiversity community (Arrows④, Fig. 1). Many groups of researchers in the LTER progr am, beyond the original data collection sites, continue to benefit from CI development and data sharing at a global scale.

Enhanced Collaboration. The mission of the LTER network was to address long-term ecological phenomena through research at individual sites, as well as by comparative research and synthesis among sites. Johnson et al. (2010) reported on how the LTER achieved its mission by using inter-site publications as a measure of collaboration. They explicitly recognized that key components of the LTER mission are to promote: inter-site co-authorship; cross-site measurements and comparisons (Hobbie et al., 2003; Redman and Foster, 2008); information technology transfer (Brunt and Michener, 2009; Porter, 2010); documentation of methodologies (Fahey and Knapp, 2007); and synthesis of ecological concepts (Peters, 2008). Over recent decades of the LTER program, the research collaboration efforts of LTER scientists have expanded from site-specific studies to the production of numerous inter-site publications (Johnson et al., 2010).

Democratization. The ILTER network formed in 1993 was largely extended in Decade III to facilitate communication and information exchange between international sites conducting comparative research. Not all researchers and scientists in various regions of the world are equally endowed with technological and data-collection resources. The ILTER program has facilitated the development of LTER-type programs where they did not exist, providing scientists with the opportunity to collaborate and have access to data and knowledge resources. The ILTER network provides all member sites with the opportunity to participate, regardless of economic status or ranking in the global science community, and thereby to pursue a diversity of approaches to interdisciplinary science (Hobbie et al., 2003). Moreover, by increasing the scope of data-collection sites, the network also directly benefits the biodiversity community by collecting data across a much wider geographical region. ILTER sites provide a rich suite of long-term interdisciplinary data sets that are beneficial for new studies (Haase et al., 2016). One ecologist interviewee indicated that:

> *"ILTER Network fosters collaborations among member scientists to extend the scope of their research across disciplinary boundaries and across more of the ILTER's 600+ research sites. The ILTER Network also has many long-term data sets that are freely available for use by students, scientists, and policymakers all over the world." (This quotation was also published by this ecologist)*

6.4. Cross-disciplinary connections leading to revolutionary progress

As data becomes central to the coevolution of CI and a discipline, its effects are manifested through both effects within the discipline, and in the relationships between the discipline and other related disciplines, such as genomics, sociology and ethno-biology, and geography (Arrows⑤, Fig. 1). For instance, biodiversity scientists intend to use genomics data by adopting DNA sequencing and bar-coding technologies to generate new data for organism classification and discovery at dramatically increased rates and volume. Further, the use of CI to integrate data from multiple disciplines has the potential to lead to paradigm shifts, or even revolutionary scientific discovery. Being cross-disciplinary and cross-paradigmatic, scientists need to go beyond the comfort zones of their established disciplinary paradigms and try to connect to another discipline's paradigm. This leads to paradigm shifts and revolutionary progress in disciplines (Kuhn, 1996).

In our study of the LTER Network, we found that scientific disciplines, not only in related or adjacent fields (such as, genomics, oceanography, climatology, etc.) but also in relatively distant disciplines (such as sociology and economics), were finding the LTER network data useful in solving research puzzles and addressing anomalies in their discipline. As data sharing increases between sites, it leads to an increase in cross-site comparative research and knowledge synthesis which can lead to revolutionary progress in related scientific disciplines (Johnson et al., 2010). However, it is noteworthy that cross-disciplinary connections facilitate but do not warrant revolutionary progress. Gittelman (2016) challenged the productivity paradox in the bio-medical research landscape, where breakthroughs in genomics did not lead to the drug discovery in medical research.

In turn, both evolutionary progress within biodiversity and revolutionary progress across disciplines will create new problems and puzzles for the scientific communities (Arrow⑥, Fig. 1), These stimuli generated within components of scientific disciplines (Arrow⑦), together with new CI investments from funding agencies (Arrow), require further CI development. As such, the scientific discipline and its supporting CI coevolve for enhancing scientific progress.

7. Implications

Our longitudinal case study of CI development for scientific progress allows us to build up a new theoretical model to depict the complex relationship between IT infrastructure and science. Our research entails significant contributions to coevolutionary theories and IT impacts, as well as to digital infrastructure development research and practices.

7.1. Implications for coevolutionary theories

Our case study enriches coevolution research by disclosing coevolution took place over a period of decades. Aligning with the complexity theory of coevolution, which posits the importance of energy importation in triggering and maintaining coevolution (Kauffman, 1993; McKelvey, 1999; Tushman and O'Reilly, 1996), our study demonstrates that various stimuli supply the energy needed to further the coevolution of CI and scientific disciplines. We distinguish external and exogenous stimuli, such as technical advancement and increased scientific funding, from the internal stimuli endogenous to the CI-supported discipline, such as problems and puzzles developed in the scientific community. While external stimuli trigger coevolution at an early stage, the endogenous stimuli emerging within the scientific community boost continuous external CI investment and provide the energy necessary for sustaining coevolution. Thus, the internal and external stimuli can shift or interweave to stimulate complex coevolution processes between CI and scientific disciplines.

Further, given the non-competition (or, at most, weak competition) among agents in the context of science, coevolution takes place at multiple levels in a synergistic system. Baum (1999) depicts a typology of four coevolution systems. This research further proposes that one particular synergistic coevolution has a nested nature, and that the agents coevolve. The micro-coevolution of three components of CI emerges in the context of macro-coevolution of CI and its supporting scientific discipline. The micro-coevolution shaping CI development enhances collaboration and democratization in the scientific community, resulting in the macro-coevolution between CI and scientific disciplines. The micro-macro synergistic co-evolutions within a specific discipline do, indeed, generate evolutionary progress of that discipline; however, revolutionary progress relies more on even broader cross-disciplinary coevolution, where CI originally applied to different scientific disciplines becomes sharable, so accelerating the scientific discovery of new problems and puzzles, especially anomalies. Therefore, the synergistic coevolutionary system in a nested nature, with continuous stimuli, can sustain scientific progress. The complex coevolution process in the LTER program sheds light on the features of embeddedness, multidirectional causalities, nonlinearity and positive feedback

(Lewin and Volberda, 1999). Also, multilevel coevolutions embody a new way of imprinting process and dynamism, enriching the imprinting framework (Simsek et al., 2015).

7.2. Implications for IT impacts on science

Our longitudinal study illustrates that the evolutionary progress of a discipline is shaped by socio-organizational changes with enhanced collaboration and democratization, whereas the revolutionary progress of a discipline is more likely to be achieved via cross-disciplinary connection. By revealing how CI development impacts the ways in which biodiversity achieves scientific progress, this research helps to close the gap between IT deployment and research productivity (Ding et al., 2010; Hesse et al., 1993; Winkler et al., 2010), while also giving valuable explanations for the IT productivity paradox (Brynjolfsson, 1993). Our case study demonstrates that CI can enhance collaboration and connectivity among scientists. Enhanced collaboration and connectivity, combined with socio-organizational complementary innovations, can lead to better scientific progress. Also, IT can produce a democratizing effect, which benefits underrepresented groups (e.g., researchers having less funding, or less access to field data) by providing them with online access to shareable data. As the ILTER network has transitioned from a satellite of the U.S. LTER to an autonomous international organization (Shibata et al., 2015), it makes the data and knowledge available to a broader set of potential researchers, thereby democratizing a discipline, and dramatically increasing the effort that can be directed into research. Sharable data also generates an equalizing force, and multiplies the pool of researchers who conduct research with the data, thereby providing a greater boost to productivity and more collaborative opportunities for researchers who are less well positioned in academe. Furthermore, the evolving technical and data infrastructures of LTER enable cross-site and even cross-disciplinary research, which makes the paradigm shifts more likely. These findings are in line with Reichman et al. (2011), who call for open data in ecology, regarding it as an essential request for interactions with other disciplines. Recently, Schlagwein et al. (2017) also highlighted the need for openness characterized by transparency, free access, inclusive participation and democracy, for various domains, e.g., open science, and open source development. Digital technologies set up a foundation for openness, and open resources on a digital platform will lead to a democratizing effect through open processes such as data sharing.

7.3. Implications for CI development for science

Our research highlights the importance of digital infrastructure development for innovations. Our longitudinal case study reveals the central role of data in linking the components within CI, as well as linking CI to scientific disciplines, thus highlighting data infrastructure as a foundation for digital infrastructure development. Yoo et al. (2010) argued that digital innovation relies on reprogrammable digital devices, widely accessible homogenized data, and the self-referential nature of digital technologies. When digitization reaches a tipping point (Hughes, 1987), the reshaping of socio-organizational infrastructure follows (Tilson et al., 2010). We find that CI development is shaped by the interplay of technical infrastructure, data infrastructure, and social-organizational infrastructure; and that standardization and shareability of data is the key. High-throughput, instrument-based data collection, fine-grained multiple modality, large-scale records, and scientific discipline-specific data formats constitute the characteristics of data infrastructure in CI. Harmonizing data for scientists' use is a significant problem. Although data infrastructure emerges playing a central role in CIdevelopment, it is essentially entangled with, and constituted through, scientific practices. Ure et al.'s (2009) study of HealthGrid in UK in the bio-medicine domain also emphasized that the challenge of data infrastructure development is to achieve the effective alignment of coupled technical and human information infrastructures. Venters et al. (2014) further argued for the mingling of practices within the computing grid infrastructure to support scientific work in the CERN particle physics community over time.

The "entanglement of practices" view (Orlikowski, 2009) and the "long now" view (Ribes and Finholt, 2009) should be acknowledged when considering scientific CI development. Kim et al. (2012) adopted the entanglement of practices perspective to conceptualize IT capability as the synergistic interactions of IT infrastructure capability, IT personnel capability, and IT management capability through continuous imbrications of human and material agencies. In a "long now" CI for science, participants experience an expanded and multi-dimensional time horizon, so that immediate problems and tactical maneuvers need to be addressed simultaneously with strategic goals and potential future alignments. Tuertscher et al.'s (2014) longitudinal study in ATLAS-CERN also demonstrates the importance of justification and interlaced knowledge sharing for architectural design and sub-system integration in a complex technological system. The engagement of domain scientists, computer scientists, ecoinformaticists and other stakeholders is crucial to create a CI enabled ecosystem with open and cutting-edge scientific services (Beaulieu et al., 2017; Farley et al., 2018). Thus, CI development should bring multiple concerns from different stakeholders, and matters of organization and technology, into a strategic frame of action, such as CI development in Decade III with a strategic plan. Furthermore, it is noteworthy that while external stimuli of investment in technologies for data collection have created a data deluge (Hey and Trefethen, 2005), maintaining data takes much organization (Lynch, 2008).

This research also suggests that CI developers have gone beyond a support role and stepped into the role of technology specifier as well as technology creator. IT has evolved from a tool supporting individuals' and organizations' work to a tool being fused into products and services, finally to become a platform nurturing innovations (Tanriverdi et al., 2010). It seems that as a result of funding stimuli, IT has progressed and developed at a faster rate than the target discipline of biodiversity (Wilson, 2013). In such a context, the digital infrastructure development often takes on a life of its own, and users' needs may be treated as secondary, especially in the earlier stages of infrastructure development. However, the principle of coevolution of CI and biodiversity discovered in this research suggests that the CI builders should be sensitive to the two-way influence between users and technologists. By studying the Worm Community System for geographically distributed geneticists, Star and Ruhleder (1996) postulated three levels of infrastructural complexity, including the specificity level, the context level and the meta-context level, that challenged both users and developers. Therefore, CI developers should recognize that infrastructure is context for both communication and learning within the web of computing, rather than a substrate that merely carries data on it. Otherwise, CI will not help the progress of scientific disciplines, owing to the transcontextual syndrome, or the contextual divide, identified by Star and Ruhleder (1996).

Our study implies that a hybrid governance arrangement for different layers of infrastructures may be an appropriate solution. On the one hand, CI acts as a generative platform supporting diversified demands from communities of scientists. CI builders need to understand how identifying, anticipating, and meeting user-requirements is important for the development of technology. This helps to address the first level of the infrastructural complexity challenge (Star and Ruhleder, 1996). On the other hand, CI builders, considering the emerging central role of data, may take into account the need for data standardization and many other aspects of data access, sharing and utilization policies. Lindgren et al.'s (2015) longitudinal study of mobile ecosystem evolution highlights the fundamental role of standardization, which makes ecosystems generative to create identity-breaking opportunities for service innovation. Data standardization and appropriate governance for data sharing help address the challenges resulting from the contextual and meta-contextual complexity of infrastructure (Star and

Ruhleder, 1996). The necessary higher level of learning and adaption have long-term implications, as it may mitigate cross-disciplinary disputes and result in revolutionary progress of the disciplines.

7.4. Recommendations to funding agencies

Our research leads to important recommendations to funding agencies and stimulus providers. In the past, funding agencies and sponsors, such as NSF, have provided only minimal support for the development of social and managerial infrastructures. Typically, the investment in CI has mostly been in technological infrastructure development, while issues of socio-organizational infrastructure have mainly been ignored. However, changes in data collection, data ownership, and data storage and organization issues are creating social issues both at the level of the discipline (Community of Science – deliberations) as well as in the socio-organizational infrastructure of CI (Process, Organization, and Governance). So far, these issues are been addressed and solved on an ad hoc basis. However, with an intensifying focus on Big Data, these issues need to be explicitly addressed in a planned manner. It is worth noticing the question of "does policy matter?" raised by Blume-Kohout and Adhikari (2016). The stage-gate process of innovation investment in life sciences (Soenksen and Yazdi, 2017) is also worthy of reference.

In the LTER program, funding agencies responded to the emerging centrality of data by providing additional funding and stimulus for data infrastructure. We have found that increasing IT investment and technological advancement resulted in dramatic increases in data size and complexity. Changes to data, eventually led to changes in processes, organization, and governance. Furthermore, the then existing social infrastructure was not up to the demands of this fledgling data infrastructure. This increased the evolutionary pressures on social infrastructures to change and adapt to the changing demands of the data infrastructures. Therefore, appropriate technology and science policies should take both the technical and social aspects of infrastructure development into account.

8. Conclusion, limitations, and future work

In conclusion, our longitudinal case study shows that the scientific discipline of biodiversity and its CI coevolve synergistically at multiple levels. Synergistic coevolution within CI happens through increased data availability and appropriate data management, showing the linking-pin of data infrastructure. The effect of technical infrastructure evolution on social-organizational infrastructure is mediated through the data infrastructure. The evolution of social-infrastructure lags behind that of technical infrastructure, as the stimulus to CI is typically applied mainly to physical items such as technical and data infrastructures. Standard metadata and sharable data facilitate collaboration and democratization among scientists across scientific communities, showing the linking-pin of data infrastructure between CI and science. Finally, "data" emerges to play a central role in linking the components of CI, and linking CI to the scientific discipline, helping bring about both evolutionary and revolutionary progress.

We are aware of the limitations of this study regarding the *generalizability* of the observed patterns of coevolution of biodiversity and CI to other disciplines, such as particle physics and astronomy. Owing to path dependency, longer established disciplines such physics and astronomy may have different coevolutionary patterns than the relatively new "green-field" discipline of biodiversity. The path dependency embedded in each discipline makes it difficult to generalize from the progress of one particular discipline to other disciplines. Testing the theoretical framework in other disciplines, such as physics or astronomy, is necessary before our results can be generalized. Furthermore, biodiversity as a teleological discipline is concerned with the issues of human and species survival. This is different from the presumed "value-neutral" nature of physics and astronomy. Thus, patterns of coevolution are likely to be different in biodiversity from those of other disciplines.

CI developments, especially data infrastructure development, are contemporary national initiatives, providing a fertile ground for future research opportunities. The conceptual framework and theory of this study should be tested against another pair of adjacent disciplines. A suggested research approach would be to undertake case studies for each discipline, with cross case analysis against the biodiversity-genomics disciplines, testing the theory and the macro-micro level relationships. Astronomy and particle physics are the two adjacent disciplines that would be worthy subjects of a longitudinal study. Astronomy-particle physics would represent physical sciences, and biodiversity-genomics would represent life sciences. Comparisons of the different sciences would provide more robust insights into the relationship between CI development and scientific progress.

Acknowledgement

Dr. Yu would like to acknowledge that her work was partially supported by the National Natural Science Foundation of China (Grant No. 71571184, 91846204).

Appendix A. Data Sources Summary

Data sources	Descriptions	Criteria/Content
Observation by researcher involvement	• One author, as a CI developer, was involved in one site of LTER • Two workshops organized in 2006 and 2013 by one author	Observe the deployment of cyberinfrastructure for biodiversity, examples are as below: 2014 - Data indexing form Yahoo, Google, Bing search engines 2014 -Drupal Ecological Information Management System (DEIMS) 2009 -ARRA funding for Cyberinfrastructure development 2003 - EML 2.0 LTER metadata standard adopted 1999 - LNO established ties with UCSD supercomputing facility 1994 - LTER WWW server created 1993 - LTER core data set catalog online 1984 - Network office created at Oregon
Interviews	14 knowledgeable informants were individually interviewed for multiple times, and each interview lasted 40 minutes to 1 hour 30 minutes. The interviewees include: • 10 senior scientists in a particular community of science, who have been practicing science in their community for the past 30 - plus years and are knowledgeable about the evolution of their discipline from the aspects of both the science and the technology; • 4 LTER program directors from government funding agencies (e.g., NSF), who have both strategic and operational view of CI development for science; Two workshops were organized and in total 38 scientists and LTER program directors and officers were involved.	**For senior scientists, sample questions were asked:** • What are the problems and puzzles in the communities? • How did technology early on and cyberinfrastructure now impact scientific progress? • Is there resistance in the community moving to the new paradigm? • What are the challenges and gaps between the old paradigm and new paradigm? • What are the contemporary events and pressures currently being faced? **For LTER program directors, sample questions were asked:** • What stimulated research and advancements in the LTER? • How all of these investments and different forms of stimulus have been generated for LTER over all these years? • What you've witnessed over the years with how data has evolved and the role of technology? • How do you see the impact of datasets availability? • How has cyberinfrastructure impacted the progress in science over the years? • What kind of scientific progress has resulted?
Archival documentation	• Strategic documents that provided information that articulates the vision or strategic goals of a scientific discipline.	• CI investment data were collected from reports archived by the Networking and Information Technology Research and Development (NITRD) Program (1992-2019) • 73 strategic reports and memos from LTER Organization • 13 reports and memos from LTER Network Communications Office
	• Informational documents that record information concerning events, activities, discussions, decisions, outcomes, etc.	• 288 reports from active committees • 129 reports from inactive committees • 250 presentations (e.g., symposia, scientists' meetings, talks, etc.); • 198 archived LTER publications (e.g, site brochures, newsletters, annual reports, etc.)
	• Programmatic documents that solicited proposals from the community of scientists to address problems and puzzles with solutions.	• 143 site proposals • 156 work groups
	• Peer-reviewed documents that shapes the body of knowledge in a discipline, including journal papers and conference proceedings.	• 121 critical LTER-related journal papers and conference proceedings were accessed and reviewed

Appendix B. Summary publication documents (1980–2019)

	Total		Total		Total
Active Committees	288	**Inactive Committees**	129	**LTER Organization**	73
Science Council	16	Executive Board (Old)	25	Decadal Reviews	18
Executive Board	107	Coordination Committee	31	Planning Documents	19
Information Management	63	Scientific Initiatives Committee	1	Bylaws	7
Education Outreach	29	Synthesis Data	1	Guidelines and Policies	13
Diversity Committee	4	Networks Coordination	1	Surveys	6
International LTER	24	Climate Committee	12	All-Scientists Meeting	10
				LTER Network	Total
Communication Committee	10	Social Science Committee	12	**Communications Office**	13
Publications Committee	22	Spatial Data and Analysis Committee	2	LTER Network Branding	8
Graduate Student Committee	13	Technology Committee	9	NCO Proposal	1
	Total				
Presentations	250	LNO Visioning Committee (2013)	1	Annual Reports to Executive Board	2
NSF LTER Symposia	130	National Advisory Board	11	Reviews and responses	2
LTER Science Council Presentations	62	Network Information System Advisory Committee	23		
			Total		Total
LTER Scientists' Meetings Plenaries	15	**LTER Publications**	198	**Programmatic documents**	299
Science Council Lightning Talks	29	Site Brochures	36	Working Groups (2000-2017)	156
Network Office Presentations	22	Annual Reports	4	Site proposals	143
Reusable Slides	1	Newsletters	58	**Peer-reviewed LTER related**	Total
		Archive	100	**documents**	121

References

Ahn, S., Huh, T., Hwang, S., Jang, J., Lee, S., 2016. Data curation for LTER: the case of K-ecohub. Int. J. Database Theor. Appl. 9 (10), 317–326.

Ahn, S., Hwang, S., Jang, J.H., Lee, S.H., Kim, S.K., 2015. Lter platform: requirement, technology and trend. Int. J. Software Eng. Appl. 9 (5), 173–182.

Aldrich, H., 1979. Organizations and Environments. Prentice-Hall, Englewood Cliffs, NJ.

Atkins, D., Droegemeier, K.K., Feldman, S.I., Hector, G.-M., Klein, M.L., Messerschmitt, D.G., Messina, P., Ostriker, J.P., Wright, M.H., 2003. Revolutionizing Science and Engineering through Cyberinfrastructure. Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, pp. 1–84.

Avila-Robinson, A., Sengoku, S., 2017. Multilevel exploration of the realities of interdisciplinary research centers for the management of knowledge integration. Technovation 62–63, 22–41.

Baum, J.A.C., 1999. Whole-part coevolutionary competition in organizations. In: Baum, J.A.C., McKelvey, B. (Eds.), Variations in Organization Science: in Honor of Donald T. Campbell, vol. 7. Sage, London, pp. 113–135.

Baum, J.A.C., Singh, J.V., 1994. Evolutionary Dynamics of Organizations. Oxford University Press, New York.

Beaulieu, S.E., Fox, P.A., Stefano, M.D., Maffei, A., West, P., Hare, J.A., Fogarty, M., 2017. Toward cyberinfrastructure to facilitate collaboration and reproducibility for marine integrated ecosystem assessments. Earth Sci. Inf. 10 (1), 85–97.

Blume-Kohout, M.E., Adhikari, D., 2016. Training the scientific workforce: does funding mechanism matter? Res. Pol. 45, 1291–1303.

Bower, J., Christensen, C., 1995. Disruptive technologies: catching the wave. Harv. Bus. Rev. 73 (1), 43–53.

Braa, J., Hanseth, O., Heywood, A., Mohammed, W., Shaw, V., 2007. Developing health information systems in developing countries: the flexible standards strategy. MIS Q. 31 (2), 381–402.

Broadbent, M., Weill, P., 1997. Management by maxim: how business and IT managers can create it infrastructures. Sloan Manag. Rev. Spring 77–92.
Broadbent, M., Weill, P., St Clair, D., 1999. The implications of information technology infrastructure for business process redesign. MIS Q. 23 (2), 159–182.
Brunt, J.W., Michener, W.K., 2009. The resource discovery initiative for field stations: enhancing data management at north American biological field stations. Bioscience 59 (6), 482–487.
Brynjolfsson, E., 1993. The productivity paradox of information technology. Commun. ACM 36 (12), 66–77.
Callahan, J.T., 1984. Long-term ecological research. Bioscience 34 (6), 363–367.
Chen, H., Chiang, R.H., Storey, V.C., 2012. Business intelligence and analytics: from big data to big impact. MIS Q. 36 (4), 1165–1188.
Churchman, C.W., 1971. The Design of Inquiry Systems: Basic Concepts of Systems and Organization. Basic Books, New York, US.
Darwin, C., 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. Murray, London, UK.
DeSanctis, G., Poole, M.S., 1994. Capturing the complexity in advanced technology use: adaptive structuration theory. Organ. Sci. 5 (2), 121–147.
Ding, W.W., Levin, S.G., Stephan, P.E., Winkler, A.E., 2010. The impact of information technology on academic scientists' productivity and collaboration patterns. Manag. Sci. 56 (9), 1439–1461.
Dong, X.Y., Yu, Y., Zhang, N., 2016. Evolution and coevolution: dynamic knowledge capability building for catching-up in emerging economies. Manag. Organ. Rev. 12 (4), 717–745.
Duncan, N.B., 1995. Capturing flexibility of information technology infrastructure: a study of resource characteristics and their measure. J. Manag. Inf. Syst. 12 (2), 37–57.
Ec, 2013. Research & innovation: European research area progress report 2013. EUR 26136 EN. http://ec.europa.eu/research/era/pdf/era_progress_report2013/era_progress_report2013.pdf. European Commission.
Edwards, P.N., Jackson, S.J., Bowker, G.C., Williams, R., 2009. Introduction: an agenda for infrastructure studies. J. Assoc. Inf. Syst. Online 10 (5), 364–374.
Eisenhardt, K.M., 1989. Building theories from case study research. Acad. Manag. Rev. 14 (4), 532–550.
Fahey, T.J., Knapp, A.K., 2007. Principles and Standards for Measuring Primary Production. Oxford University Press, New York, New York, USA.
Farley, S.S., Dawson, A., Goring, S.J., Williams, J.W., 2018. Situating ecology as a big-data science: current advances, challenges, and solutions. Bioscience 68 (8), 563–576.
Gittelman, M., 2016. The revolution re-visited: clinical and genetics research paradigms and the productivity paradox in drug discovery. Res. Pol. 45 (8), 1570–1585.
Gosz, J.R., Waide, R.B., Magnuson, J.J., 2010. Twenty-eight Years of the US-LTER Program: Experience, Results, and Research Questions, Long-Term Ecological Research. Springer Netherlands, pp. 59–74.
Graham, M.H., Dayton, P.K., 2002. On the evolution of ecological ideas: paradigms and scientific progress. Ecology 83 (6), 1481–1489.
Haase, P., Frenzel, M., Klotz, S., Musche, M., Stoll, S., 2016. The long-term ecological research (LTER) network: relevance, current status, future perspective and examples from marine, freshwater and terrestrial long-term observation. Ecol. Indicat. 100 (65), 1–3.
Henfridsson, O., Bygstad, B., 2013. The generative mechanisms of digital infrastructure evolution. MIS Q. 37 (3), 907–931.
Hesse, B.W., Sproull, L.S., Kiesler, S.B., Walsh, J.P., 1993. Returns to science: computer networks in oceanography. Commun. ACM 36 (8), 90–101.
Hey, T., Trefethen, A.E., 2005. Cyberinfrastructure for e-science. Science 308 (5723), 817–821.
Hobbie, J.E., Carpenter, S.R., Grimm, N.B., Gosz, J.R., Seastedt, T.R., 2003. The US long term ecological research program. Bioscience 53 (1), 21–32.
Hughes, T.P., 1987. The evolution of large technological systems. In: Bijker, W.B., Hughes, T.P., Pinch, T. (Eds.), The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology: 51-82. MIT Press, Cambridge, MA.
Johnson, J.C., Christian, R.R., Brunt, J.W., Hickman, C.R., Waide, R.B., 2010. Evolution of collaboration within the US long term ecological research network. Bioscience 60 (11), 931–940.
Kabo, F.W., Cotton-Nessler, N., Hwang, Y., Levenstein, M.C., Owen-Smith, J., 2014. Proximity effects on the dynamics and outcomes of scientific collaborations. Res. Pol. 43, 1469–1485.
Kauffman, S.A., 1993. The Origins of Order: Self-Organization and Selection in Evolution. Oxford University Press, New York.
Kettinger, W.J., Teng, J.T., Guha, S.G., 1997. Business process change: a study of methodologies, techniques, and tools. MIS Q. 21 (1), 55–80.
Kim, G., Shin, B., Kwon, O., 2012. Investigating the value of socio-materialism in conceptualizing it capability of a firm. J. Manag. Inf. Syst. 29 (3), 327–362.
Kim, L., 1998. Crisis construction and organizational learning: capability building in catching-up at Hyundai motor. Organ. Sci. 9 (4), 506-402.
Kuhn, T.S., 1996. The Structure of Scientific Revolutions. University of Chicago Press, Chicago, IL, US.
Lane, J.I., Owen-Smith, J., Rosen, R.F., Weinberg, B.A., 2015. New linked data on research investments: scientific workforce, productivity, and public value. Res. Pol. 44, 1659–1671.
Langley, A., 1999. Strategies for theorizing from process data. Acad. Manag. Rev. 24 (4), 691–710.
Langley, A., Smallman, C., Tsoukas, H., Van de Ven, A.H., 2013. Process studies of change in organization and management: unveiling temporality, activity, and flow. Acad. Manag. J. 56 (1), 1–13.
Lathrop, R.C., Carpenter, S.R., Rudstam, L.G., 1996. Water clarity in lake Mendota since 1900: responses to differing levels of nutrients and herbivory. Can. J. Fish. Aquat. Sci. 53 (10), 2250–2261.
Law, J., 1987. Technology and heterogeneous engineering: the case of Portuguese expansion. In: Bijker, W.E., Hughes, T.P., Pinch, T. (Eds.), The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology: 105-128. MIT Press, Cambridge, MA.
Lewin, A.Y., Long, C.P., Carroll, T.N., 1999. The coevolution of new organizational forms. Organ. Sci. 10 (5), 535–550.
Lewin, A.Y., Volberda, H.W., 1999. Prolegomena on coevolution: a framework for research on strategy and new organizational forms. Organ. Sci. 10 (5), 519–534.
Lindgren, R., Eriksson, O., Lyytinen, K., 2015. Managing identity tensions during mobile ecosystem evolution. J. Inf. Technol. 30 (3), 229–244.
Lynch, C., 2008. Big data: how do your data grow? Nature 455 (7209), 28–29.
MacLean, D.F., 2004. Internet Governance: A Grand Collaboration Herding Schrödinger's Cats": Some Conceptual Tools for Thinking about Internet Governance. United Nations Publications, New York.
Maxwell, J.A., 2012. Qualitative Research Design: an Interactive Approach, third ed. Sage, Thousand Oaks, CA.
McKelvey, B., 1982. Organizational Systematics-Ttaxonomy, Evolution, Classification. University of California Press, Berkley.
McKelvey, B., 1997. Quasi-natural organization science. Organ. Sci. 8 (4), 352–380.
McKelvey, B., 1999. Avoiding complexity catastrophe in coevolutionary pockets: strategies for rugged landscapes. Organ. Sci. 10 (3), 294–321.
Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B., Stafford, S.G., 1997. Nongeospatial metadata for the ecological sciences. Ecol. Appl. 7 (1), 330–342.
Michener, W.K., Porter, J., Servilla, M., Vanderbilt, K., 2011. Long term ecological research and information management. Ecol. Inf. 6 (1), 13–24.
Nichol, S.T., Spiropoulou, C.F., Morzunov, S., Rollin, P.E., Ksiazek, T.G., Feldmann, H., Sanchez, A., Childs, J., Zaki, S., Peters, C.J., 1993. Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. Science 262 (5), 914–917.
NSF-US, 2006. Cyberinfrastructure in FY 2006 Budget Request. National State Foundation, United States, pp. 395–397. http://www.nsf.gov/about/budget/fy2006/pdf/2009-NSF-WideInvestments/2032-FY2006.pdf.
O'Malley, M.A., Elliott, K.C., Haufe, C., Burian, R.M., 2009. Philosophies of funding. Cell 138 (4), 611–615.
Orlikowski, W., 2009. The sociomateriality of organizational life: considering technology in management research. Camb. J. Econ. 34 (1), 125–141.
Peters, D.P., 2008. Ecology in a connected world: a vision for a "network of networks". Front. Ecol. Environ. 6 (5), 227-227.
Popper, K.R., 1959. The Logic of Scientific Discovery. Hutchinson, London.
Popper, K.R., 1963. Conjectures and Refutations. Routledge & Kegan Paul, London.
Porter, J.H., 2010. A brief history of data sharing in the us long term ecological research network. Bull. Ecol. Soc. Am. 91 (1), 14–20.
Prigogine, I., 1962. Non-equilibrium Statistical Mechanics. Wiley Interscience, New York.
Prigogine, I., 1980. From Being to Becoming - Time and Complexity in the Physical Science. Freeman, San Fransisco.
Redman, C.L., Foster, D.R., 2008. Agrarian Landscapes in Transition: Comparisons of Long-Term Ecological and Cultural Change. Oxford University Press, USA.
Reichman, O., Jones, M.B., Schildhauer, M.P., 2011. Challenges and opportunities of open data in ecology. Science 331 (6018), 703–705.
Ribes, D., Finholt, T.A., 2009. The long now of technology infrastructure: articulating tensions in development. J. Assoc. Inf. Syst. Online 10 (5), 375–398.
Ribes, D., Polk, J.B., 2014. Flexibility relative to what? Change to research infrastructure. J. Assoc. Inf. Syst. Online 15, 287–305. Special Issue.
Robbins, R.J., 2011. Data Management for the Lter: 1980 - 2010. National Science Foundation.
Roberts, N., Grover, V., 2012. Leveraging information technology infrastructure to facilitate a firm's customer agility and competitive activity: an empirical investigation. J. Manag. Inf. Syst. 28 (4), 231–269.
Robertson, D.S., 2003. Phase Change: the Computer Revolution in Science and Mathematics. Oxford University Press.
Robertson, G.P., 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. LTER Network Office Publications, Albuquerque, New Mexico.
Robertson, G.P., Collins, S.L., Foster, D.R., Brokaw, N., Ducklow, H.W., Gragson, T.L., Gries, C., Hamilton, S.K., McGuire, A.D., Moore, J.C., 2012. Long-term ecological research in a human-dominated world. Bioscience 62 (4), 342–353.
Sambamurthy, V., Bharadwaj, A., Grover, V., 2003. Shaping agility through digital options: reconceptualizing the role of information technology in contemporary firms. MIS Q. 27 (2), 237–263.
Schlagwein, D., Conboy, K., Feller, J., Leimeister, J.M., Morgan, L., 2017. "Openness" with and without information technology: a framework and a brief history. J. Inf. Technol. 32 (4), 297–305.
Sharp, A., McDermott, P., 2008. Workflow Modeling: Tools for Process Improvement and Applications Development. Artech House, Norwood, MA.
Shibata, H., Branquinho, C., McDowell, W.H., Mitchell, M.J., Monteith, D.T., Tang, J., Kopáček, J., 2015. Consequence of altered nitrogen cycles in the coupled human and ecological system under changing climate: the need for long-term and site-based research. Ambio 44 (3), 178–193.
Simsek, Z., Fox, B.C., Heavey, C., 2015. "What's past is prologue": a framework, review, and future directions for organizational research on imprinting. J. Manag. 41 (1), 288–317.

Soenksen, L.R., Yazdi, Y., 2017. Stage-gate process for life sciences and medical innovation investment. Technovation 62–63, 14–21.
Stafford, S.G., Brunt, J.W., Michener, W.K., 1994. Integration of scientific information management and environmental research. In: Michener, W.K., Brunt, J.W., Stafford, S.G. (Eds.), Environmental Information Management and Analysis: Ecosystem to Global Scales: 3-19. Taylor & Francis, London.
Star, S.L., Ruhleder, K., 1996. Steps toward an ecology of infrastructure: design and access for large information spaces. Inf. Syst. Res. 7 (1), 111–134.
Stinchcombe, A.L., 1965. Social structure and organizations. In: March, J.G. (Ed.), Handbook of Organization. Rand-McNally, Chicago, pp. 142–193.
Strauss, A., Corbin, J., 2008. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory, third ed. Sage, Thousand Oaks, CA.
Tallon, P.P., Ramirez, R.V., Short, J.E., 2013. The information artifact in IT governance: toward a theory of information governance. J. Manag. Inf. Syst. 30 (3), 141–178.
Tanriverdi, H., Rai, A., Venkatraman, N., 2010. Reframing the dominant quests of information systems strategy research for complex adaptive business systems. Inf. Syst. Res. 21 (4), 822–834.
Thompson, G., 1991. Markets, Hierarchies and Networks: the Coordination of Social Life. Sage, London.
Thompson, J.N., 1994. The Coevolutionary Process. University of Chicago Press, Chicago and London.
Tilson, D., Lyytinen, K., Sørensen, C., 2010. Digital infrastructure: the missing is research agenda. Inf. Syst. Res. 21 (4), 748–759.
Tripathi, A.K., Agrawal, S., Gupta, R.D., 2020. Cloud enabled SDI architecture: a review. Earth Sci. Inf. 13 (2), 211–231.
Tuertscher, P., Garud, R., Kumaraswamy, A., 2014. Justification and interlaced knowledge at ATLAS, CERN. Organ. Sci. 25 (6), 1579–1608.
Tushman, M.L., O'Reilly, I., 1996. Ambidextrous organization: managing evolutionary and revolutionary change. Calif. Manag. Rev. 38 (4), 8–30.
Tydecks, L., Bremerich, V., Jentschke, I., Likens, G.E., Tockner, K., 2016. Biological field stations: a global infrastructure for research, education, and public engagement. Bioscience 66 (2), 164–171.
Urbinati, A., Bogers, M., Chiesa, V., Frattini, F., 2019. Creating and capturing value from big data: a multiple-case study analysis of provider companies. Technovation 84–85, 21–36.
Ure, J., Hartswood, M., Wardlaw, J., Procter, R., Anderson, S., Gonzalez-Velez, H., Lin, Y., Lloyd, S., Ho, K., 2009. The development of data infrastructures for eHealth: a socio-technical perspective. J. Assoc. Inf. Syst. Online 10 (5), 415–429.
Van den Bosch, F.A.J., Volberda, H.W., de Boer, M., 1999. Coevolution of firm absorptive capacity and knowledge environment: organizational forms and combinative capabilities. Organ. Sci. 10 (5), 551–568.
Van Valen, L., 1983. How pervasive is coevolution? In: Nitecki, M. (Ed.), Coevolution. University of Chicago Press, IL.
Vasileiadou, E., Vliegenthart, R., 2009. Research productivity in the era of the internet revisited. Res. Pol. 38, 1260–1268.
Velu, C.K., Madnick, S., van Alstyne, M.W., 2013. Centralizing data management with considerations of uncertainty and information-based flexibility. J. Manag. Inf. Syst. 30 (3), 179–212.
Venters, W., Oborn, E., Barrett, M., 2014. A trichordal temporal approach to digital coordination: the sociomaterial mangling of the CERN grid. MIS Q. 38 (3), 927–949.
Volberda, H.W., Lewin, A.Y., 2003. Co-evolutionary dynamics within and between firms: from evolution to co-evolution. J. Manag. Stud. 40 (8), 2111–2136.
Walsham, G., 2006. Doing interpretive research. Eur. J. Inf. Syst. 15 (3), 320–330.
Weill, P., Ross, J.W., 2004. IT Governance. Harvard Business School Press, Boston.
Wilson, E.O., 2013. Beware the age of loneliness. The Economist.
Winkler, A.E., Levin, S.G., Stephan, P.E., 2010. The diffusion of IT in higher education: publishing productivity of academic life scientists. Econ. Innovat. N. Technol. 19 (5), 481–503.
Yang, C., Huang, Q., Li, Z., Liu, K., Hu, F., 2017. Big data and cloud computing: innovation opportunities and challenges. Int. J. Digit. Earth 10 (1), 13–53.
Yang, C., Wu, H., Huang, Q., Li, Z., J, L., 2011. Using spatial principles to optimize distributed computing for enabling the physical science discoveries. Proc. Natl. Acad. Sci. Unit. States Am. 108 (14), 5498–5503.
Yang, Z., Li, W., Chen, Q., Wu, S., Liu, S., Gong, J., 2019. A scalable cyberinfrastructure and cloud computing platform for forest aboveground biomass estimation based on the Google earth engine. Int. J. Digit. Earth 12 (9), 995–1012.
Yates, T.L., Mills, J.N., Parmenter, C.A., Ksiazek, T.G., Parmenter, R.R., Vande Castle, J. R., Calisher, C.H., Nichol, S.T., Abbott, K.D., Young, J.C., 2002. The ecology and evolutionary history of an emergent disease: hantavirus pulmonary syndrome. AIBS (Am. Inst. Biol. Sci.) Bull. 52 (11), 989–998.
Yin, R.K., 2009. Case Study Research: Design and Methods, fourth ed. Sage, Thousand Oaks, CA.
Yoo, Y., Henfridsson, O., Lyytinen, K., 2010. The new organizing logic of digital innovation: an agenda for information systems research. Inf. Syst. Res. 21 (4), 724–735.
Zahra, S.A., George, G., 2002. Absorptive capacity: a review, reconceptualization, and extension. Acad. Manag. Rev. 27 (2), 185–203.

Dr. Yan Yu is an Associate Professor in School of Information, Renmin University of China. She received her B.A., B. Econ. and M.S. degrees from Peking University, and received her Ph.D in Information Systems from City University of Hong Kong in 2009. Her research interests include Digital innovation, Knowledge management, and Organization evolution. Dr. Yu has published papers in scholarly journals, including *Journal of Business Research, Management and Organization Review, Journal of Knowledge Management, International Journal of Information Management, Information & Management, Decision Support Systems,* and others.

Dr. Julio Ibarra is the Assistant Vice President for Technology Augmented Research at Florida International University (FIU). He holds B.S. and M.S. in Computer Science from FIU, and Ph.D. in Telematics and Information Technology from Twente University. He is responsible for furthering the mission of the Center for Internet Augmented Research and Assessment (CIARA) – to contribute to the pace and the quality of research at FIU and its international collaborators through the application of advanced Cyberinfrastructure. He is also responsible for the development and management of the AMPATH International Exchange Point for Research and Education networks, in Miami, Florida.

Prof. Dr. Kuldeep Kumar is an Emeritus Professor in College of Business Administration at Florida International University. He implemented innovations in executive programs and doctoral programs in a global wide. He also served as academic directors of these programs across US, Europe and Asia. His research areas include Management and design of global distributed work, Cyberinfrastructure, Knowledge and innovation management, Offshore outsourcing. Prof. Dr. Kumar has published a number of papers in top journals such as *MIS Quarterly, Journal of International Business Studies, Communications of ACM, Journal of Information Technology, IEEE Transactions on Software Engineering,* etc.

Vasilka Chergarova is a Research Coordinator at Center for Internet Augmented Research and Assessment at Florida International University. She is a Ph.D. student in Information Systems and Security at NOVA Southeastern University. She received her Master in Management of Information Systems and Bachelor of Science in Information Technology from FIU. As an Open Science Data Cloud Fellow (2011), she researched open-source systems for distributed cloud-based infrastructure at Sao Paulo State University (UNESP). She is interested in data mining, data analytics, and data visualization.