

RESEARCH ARTICLE OPEN ACCESS

The Incremental Growth of Data Infrastructure in Ecology (1980–2020)

Karen S. Baker¹  | Florence Millerand² ¹Department of Information Sciences, University of Illinois, Urbana-Champaign, Urbana, Illinois, USA | ²Département de Communication Sociale et Publique Montréal Québec, Université du Québec à Montréal, Latin Quarter, Montreal, Canada**Correspondence:** Karen S. Baker (karensbaker@gmail.com)**Received:** 27 February 2024 | **Revised:** 18 September 2024 | **Accepted:** 28 September 2024**Funding:** The authors received no specific funding for this work.**Keywords:** data infrastructure | data management | incremental growth | information system trajectory | open data | open science

ABSTRACT

After decades of growth, a research community's network information system and data repository were transformed to become a national data management office and a major element of data infrastructure for ecology and the environmental sciences. Developing functional data infrastructures is key to the support of ongoing Open Science and Open Data efforts. This example of data infrastructure growth contrasts with the top-down development typical of many digital initiatives. The trajectory of this network information system evolved within a collaborative, long-term ecological research community. This particular community is funded to conduct ecological research while collective data management is also carried out across its geographically dispersed study sites. From this longitudinal ethnography, we describe an Incremental Growth Model that includes a sequence of six relatively stable phases where each phase is initiated by a rapid response to a major pivotal event. Exploring these phases and the roles of data workers provides insight into major characteristics of digital growth. Further, a transformation in assumptions about data management is reported for each phase. Investigating the growth of a community information system over four decades as it becomes data infrastructure reveals details of its social, technical, and institutional dynamics. In addition to addressing how digital data infrastructure characteristics change, this study also considers when the growth of data infrastructure begins.

1 | Introduction

We investigate the growth of an information system and its role as data infrastructure for a research community responding to continuing change in the digital era. The Long-Term Ecological Research (LTER) program's development of a Network Information System (NIS) for managing long-term data provides an example of supporting scientific research while addressing the increasing demands of data work.

Information systems are central to contemporary Open Science efforts to ensure the transparency of science. Open Science includes Open Data that aims to assemble data and make it accessible. This data effort depends upon the development of

individual and collective data practices as well as data infrastructure. Such efforts benefit from ethnographic studies of research environments with their many significant differences in logistics, methods, and interests as well as in support structures, data arrangements, and community assumptions. These differences give rise to the variety of data infrastructures available today (Pomerantz and Peek 2016; Hampton et al. 2015; Borgman 2015; Kitchen 2014). Care is needed to avoid considering only early technology-centric designs that may preclude community colearning, user participation, and preservation of local knowledge as growth occurs. National and international reports describe Open Science and Open Data as having goals such as facilitating data access and reuse (TRS 2012; EU 2016; OECD 2015; NAP 2018). Descriptions of existing digital

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Ecology and Evolution* published by John Wiley & Sons Ltd.

arrangements contribute to understanding potential responses to Open Data mandates such as those from the U.S. Office of Science and Technology (OSTP 2013, 2022) and the National Science and Technology Council (NSTC 2022).

This case presents the opportunity to address three research questions:

- How does the growth of digital data infrastructure occur over time?
- What characteristics and data workers are associated with the evolution of an information system?
- When does the growth of digital data infrastructure begin?

This study provides some options for planning data management to achieve Open Data and makes a number of additional contributions relating to these research questions. First, using this case, we propose an Incremental Growth Model that describes an information system trajectory as a series of phases initiated by pivotal events, thereby capturing both the stability and change involved. Second, we identify major characteristics of each phase of an information system's growth. Third, rather than selecting a few snapshots in time, analysis of each phase provides key insights relating to the growth of data infrastructure over 40 years. Finally, to facilitate comparative studies, we outline this case's contextual conditions.

2 | Background

To study a scientific research program as it grapples with data issues and data systems, we draw on Infrastructure Studies informed by Science and Technology Studies, an approach that ensures sensitivity to the complexity of a digital environment. In addition, Organizational Change Research contributes to the framework, which was developed to capture the broader context of events and major characteristics associated with an information system's trajectory. We begin by introducing the concepts of data infrastructure, continuing design, and pivotal events. To address data infrastructure growth, we also consider key issues associated with data work and data workers before introducing the LTER program.

2.1 | Data Infrastructure

The term digital infrastructure is associated with computers, platforms, and software. In a National Science Foundation (NSF) report, the scope of infrastructure was broadened dramatically from a technically oriented understanding to include services and organizations as well as the people who design, develop, and facilitate (Atkins 2003). Subsequently, NSF reported on cyberinfrastructure as meeting the need for new kinds of organization and computing that would include planning for data and a data workforce (NSF CI Report 2007, 2019).

Infrastructure as conceptualized and characterized by Star and Ruhleder (1996) is a multifaceted concept referring to interrelated arrangements involving hardware, software, standards,

procedures, practices, and policies together with support for human communication and capabilities. Research has highlighted the need to consider how infrastructure:

- Forms and accumulates over time in collaborative and iterative design (Edwards et al. 2007; Erickson and Sawyer 2019; Karasti, Pipek, and Bowker 2018).
- Ensures flexibility through configurations including an installed base, kernel, and/or gateways (Star and Ruhleder 1996; Hanseth 2010; Ribes 2014).
- Benefits from the development of standards (Edwards 2004; Millerand and Bowker 2009; Hanseth 2010).
- Represents a dynamic digital ecosystem (Mongili and Pellegrino 2014; Harvey, Jensen, and Morita 2017; Millerand and Baker 2020).
- Maintains resiliency via loose couplings and recouplings within and between elements, components, and systems (Winter et al. 2014; Misangyi 2016; Miller 2018; Orton and Weick 1990; Arango-Vasquez and Gentilin 2021).
- Evolves into a variety of forms of knowledge infrastructure (Edwards et al. 2013; Bowker 2017; Borgman, Scharnhorst, and Golshan 2019).

Infrastructures today are considered complex phenomena that involve a collection of interacting digital systems and services (Bowker et al. 2010). Infrastructure from an information systems perspective, shifts from considering the alignment of digital systems to a broader view that takes into account the changing components of systems and subsystems supported by digital workers. Data Infrastructure Studies recognize the agency of humans and technologies separately and together: “infrastructures are shaped by multiple agents with competing interests and capacities, engaged in an indefinite set of distributed interactions over extended periods of time. The characteristics of infrastructure emerge out of these interactions” (Harvey, Jensen, and Morita 2017). To address infrastructural complexity, Hepsø, Monteiro, and Rolland (2009) present the notion of ecologies of e-infrastructure within a business setting where six general categories of important events are identified, and Pollock and Williams (2010) develop the “Biography of Artifacts” concept using multi-sited studies for comparative analysis of enterprise systems in business and health settings.

Infrastructure emerges when interacting systems of systems and people become an established part of a community's work. Many terms are used in referring to such digital arrangements including information infrastructure, research infrastructure, and cyberinfrastructure. We focus on data infrastructure “as one type of information infrastructure that supports science by facilitating individual and collective work with data (...) where it includes more than digital technologies as [it is] also constituted of individuals, organizations, routines, shared norms, and practices” (Millerand and Baker 2020). The interdependence of social, technical, and institutional facets of these systems results in a complexity often difficult to describe fully. In such an environment, it is critical to keep in mind that any single perspective on such phenomena will be partial.

2.2 | Digital Data Infrastructure and Continuing Design

A linear approach was used in building early digital systems with all stages of development planned at project start. In time, such static planning has been found to be limited by a lack of user input and inadequate adaptive capabilities. An iterative design approach incorporating agile and user-centered development avoids many such limitations through the use of dynamic planning including repeating cycles of design-develop-test. This kind of growth provides the time needed for learning between participants. Each cycle spurs the evolution and evaluation of software, applications, and systems. Iterative design -including methodologies such as participatory design (Simonsen and Robertson 2013), participatory infrastructuring (Bødker, Dindler, and Iversen 2017), collaborative design (Pipek, Karasti, and Bowker 2018), and meta-design (Fischer and Ostwald 2002; Fischer and Herrmann 2011, 2015) – focuses on dynamic processes and interactions between designers, users, and their environment. Iterative design circles back to the concept of “continuing design” (Henderson and Kyng 1991) to emphasize developments both within a cycle and cumulatively over all the phases. The term “infrastructuring” emerged to underscore the active process of creating and maintaining infrastructure (Karasti and Baker 2004; Pipek and Wulf 2009; Pollock and Williams 2010; Karasti, Pipek, and Bowker 2018). It signals a move from the idea of infrastructure as a built object or a stable product to imagining an evolving entity (Hanseth 2010) or a process of growth sometimes referred to as “infrastructure in-the-making” (Edwards et al. 2007; Pipek, Karasti, and Bowker 2017).

In cases of complexity and change, the notion of sensitizing concepts is used (e.g., Ribes and Polk 2014). For the case under study here, two facets of change are central: sociotechnical and institutional. A sociotechnical facet draws attention to the design and use of digital data and systems by humans (Aken 2004; Salgado, Morel, and Vêrilhac 2018). It refers to the intertwining of the social and technical arrangements or actions. We define the sociotechnical facet to be inclusive of situations in digital arenas where participants work at individual, project, and community levels, such as with the use of spreadsheets in an individual's lab or submission of data to a local data system. Arrangements and actions in larger-scale arenas are also sociotechnical. They are designated institutional facets, however, to highlight their place in hierarchical structures, such as a national organization's actions or an agency's funding program arrangements.

2.3 | Incremental Growth and Pivotal Events

When an object of study is a process, new approaches are needed. Organization Research has developed “event structure analysis” (Hak, Jaspers, and Dul 2013) while Science and Technology Research recounts how “both rhythms and events can disrupt, shape and reorient plans” (Steinhardt and Jackson 2014). In the development of digital systems subject to ongoing interactions and change, sequential periods can be identified as phases. Whereas a long-term plan provides an overarching general mission, a short-term plan lays out possible advances for the

system and its products given resources at hand for the current phase. A phase initiated by a major pivotal event persists for a period of time, ending when another pivotal event takes place. A pivotal event is defined as a time-bound situation with highly consequential potential (Cutcher-Gershenfeld, Brooks, and Mulloy 2015; Cutcher-Gershenfeld 2020). It is a precipitating event that represents a point at which opportunities occur and choices are made. The term incremental growth is used to describe the manner in which plans for a digital system unfold over time in phases.

Information Systems Research focuses on digital networks and their relations as key to understanding infrastructure and to establishing, maintaining, or increasing alignment of systems. In their study of digital infrastructure evolution from a configurational perspective, Henfridsson and Bygstad (2013) identify generative mechanisms as “causal structures that generate observable events”, thereby creating potential paths for infrastructural evolution. They categorize generative mechanisms as adoptive, innovative, scaling, or some combination of these resulting in new services. Adoption mechanisms lead to more coordinated participation, innovation enables change, and scaling mechanisms lead to modular growth. Pivotal events function as generative mechanisms that spur critical changes in the trajectory of an information system.

2.4 | Data Work and Data Workers

The notion of data, with its tangled relations to information and knowledge, frequently proves elusive to define (e.g., NRC 1997; Beaujardièrre 2016; Leonelli 2015, 2016). Borgman (2015) reports concisely that “data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship.” The increasing diversity in kinds of data in terms of origin, methods, collection, use, and users makes the work associated with data difficult to classify. Indeed, recent observations and cases illustrate the importance of the product manager and the commonplace lack of regard for the role in large-scale digital efforts (e.g., Pahlka 2023). The role of data manager, one type of product manager, may encounter such disregard even in a data-oriented community. The LTER case, with its Open Data and data infrastructure efforts, holds the potential to add to Pahlka's notion of “bottom-up” or support for the “people in the trenches.”

Data work is defined as an effort applied to data that may involve heterogeneous data activities occurring in diverse settings that are carried out by humans and technologies separately and together (Baker 2017). For data workers today, data work roles are frequently force fit into earlier job classification systems that are inadequate for the wide range of data skills, tasks, and responsibilities now associated with research. Over the last decade, Human Resource Offices within universities (e.g., UCSD n.d.) and staff within libraries (e.g., Cox and Corral 2013) have begun to address digital data roles. Whether expanding old job profiles or developing new ones with updated job functions, these efforts often lag behind the rapid diversification of the digital data workforce. Despite emerging titles such as software engineer, data manager, and

visualization specialist, data workers draw attention to the marked lack of career paths today.

Data work was once solely in the hands of researchers. Today data workers are specialists in digital environments advancing digital fluency with new vocabulary such as metadata, data package, data flow, database, and data stewardship. The Long-Term Ecological Research (LTER) program (e.g., Baker et al. 2000; Karasti and Baker 2004; Baker and Millerand 2010; Stafford 2021) provides an early example of establishing new data management practices and developing a data workforce at universities and other research organizations supported by federal funds. The role of data manager typically includes familiarity with the research process as well as selecting, implementing, maintaining, and advancing technology, software, applications, and/or services associated with data. This role is distinct from roles such as research assistant, data analyst, programmer, data scientist, software engineer, cyberinfrastructure specialist, or others concerned with computational systems and methods. The role is often situated within an environment that calls for skills in interpersonal communications and data care (Baker and Millerand 2007, 2010; Baker and Karasti 2018). From this position, data managers are able to introduce updates in data practices in addition to facilitating data organization and access. The role is sometimes recognized as a “sociotechnical bridge” for a community of researchers given the “boundary-spanning” work needed across arenas (individual, project, network, and national) as well as across fields (data, informatics, software, computing, technology, and scientific research).

In this paper, several terms are used to describe those engaging in data work. The notion of “data workers” includes data managers and other data-oriented participants. The role of the data worker at the first LTER sites was initially called “data manager” (DM), a designation used throughout this paper for clarity though the role was later renamed “information manager” within the LTER (Baker et al. 2000). The term “NIS-data team” refers to LTER Network Office (LNO) software engineers and LTER site-based data managers working together on a Network Information System with a high degree of awareness of the sociotechnical ramifications of data work. We use the software engineer category broadly to encompass a number of software, technology, and data specialists including developers, designers, and computing and networking professionals.

2.5 | The LTER Program

2.5.1 | LTER History

The LTER program and its data efforts were informed by lessons learned a decade earlier in the International Biological Program (IBP). During the IBP from 1964 to 1974, ecological data was sometimes delivered as paper records to a set of offices distributed across the U.S. (Golley 1993; Aronova, Baker, and Oreskes 2010; Coleman 2010). The IBP goal to aggregate data was hindered by the lack of data practices, technologies, and management resources needed to support the challenges faced by the diversity of biological field projects. A LTER grant proposal to NSF to address this succeeded by focusing on the importance of long-term data. The LTER program and its data management efforts

have been supported by a number of visionary program leaders at NSF who recognized the need for and benefits of change. The NSF Division of Biological Infrastructure (DBI) provided initial guidance. LTER was also fortunate to have a number of insightful Principal Investigators who supported the new kinds of work associated with long-term data.

The LTER program began as a set of six sites established in 1980, a time when databases were a new computational tool available to researchers in a few research arenas. Sites were funded as independent projects though expected to collaborate as a network. Through fieldwork at these ecological sites, researchers aimed to discern changes in each site's biome by collecting long-term data. By 2017, the network had grown to 28 sites, each anchored at a university or other research center and consisting of an interdisciplinary team of scientists focused on a particular location. Unlike the highly centralized data facilities of the IBP, LTER in the beginning was configured to support site-based efforts. An NSF mandate requiring each site to identify and support a local data manager set the stage for the program's work with data. Initially, sites developed local data practices and data systems that assembled data from the many project members at a site. Just as in ecology, where differences in ecosystems are fully appreciated and expected, the LTER sites developed understandings of their local “data ecosystem.” In the 1990s, when an LTER Network Office (LNO) was established, each site contributed to cross-site ecological research as well as to network-wide data work. Data managers had to learn how to balance site and network data activities (Baker and Karasti 2004; Karasti, Baker, and Millerand 2010). Their cross-site communication and coordination of data was enabled by the early formation of an Information Management Committee (IMC) that was established as an LTER Network standing committee. The IMC became an active community of practice with data managers engaged from all the sites.

2.5.2 | LTER Contextual Conditions

Henfridsson and Bygstad (2013) work with the concept of contextual conditions to facilitate comparative analyses of case studies. Six LTER conditions describe the setting and arrangements within which the LTER Network Information System (NIS) evolved. They are summarized in Table 1 and discussed below.

(1) **Program Structure:** Each site studies an ecological biome over time, investigating local themes and five network-wide core themes: “pattern and control of primary production, spatial and temporal distribution of populations selected to represent trophic structure, pattern, and control of organic matter accumulations, patterns of inorganic input and movements, and patterns and frequency of disturbance to the research site” (Callahan 1984). The LTER network carries out interdisciplinary research. It has characteristics associated with successful synthesis centers including active management, computing and informatics capabilities, flexibility, student and fellow support, diversity, and placing value on unstructured time (Baron et al. 2017).

(2) **Long-Term Funding:** Sites were expected to take responsibility for the generation of long-term data rather than the more typical 1- to 3-year collections of data (Magnuson 1990). A new NSF funding model supported six-year LTER funding

TABLE 1 | Contextual conditions describing the LTER case setting and arrangements for science and data.

<p>1. <i>Program structure</i>: A science-drive community established as a set of geographically distributed research sites, each funded independently to collaborate on a multi-investigator project studying a designated biome, where these decentralized sites are joined together as a network with a shared vision, cross-site research, and community governance with a network office providing supporting communication, and coordination.</p>
<p>2. <i>Long-term funding</i>: Provisional, long-term core funding supports the sites individually and the network while participants also pursue resources from other sources to augment existing capabilities and to fund new pursuits and/or equipment.</p>
<p>3. <i>Communication structure</i>: Five major components - site projects, community governance, independently funded services (Network Office and Network Information System/ EDI), partners, and NSF - represent work arenas all interconnected by continuing reciprocal communications.</p>
<p>4. <i>Data management</i>: A data manager works at each site tasked with developing collective data practices, data management, and technical capabilities including juggling local data needs with network-wide data activities associated with cross-site research.</p>
<p>5. <i>Technology engagement</i>: Both site and network data workers take an agile, incremental, and integrative approach using iterative design with digital data work in support of scientific research.</p>
<p>6. <i>Community information system</i>: A network information system grew as a loosely coupled architecture with data-oriented developers at the network office working closely with data managers and researchers at the sites, a collaboration that facilitates system use and data reuse.</p>

cycles for each site with the expectation of a new cycle if their renewal proposal were accepted (Hobbie 2003). This new research model of long-term funding was overseen by individuals in the often underappreciated position of NSF program director who over the years have played a significant role in both building and sometimes damaging LTER. These individuals together with program managers have been important as the network responded to change. For instance, the LTER community originally aimed to discern changes in their designated ecosystems. Studies initially focused on changes due to natural forces but soon included attention to human impacts (LTER Risser Report 1993). This envisioning of change due to human influences led in 1997 to the network expanding to include two urban sites.

(3) **Communication Structure**: The LTER communication structure is portrayed with five major components (Figure 1) that are discussed below.

- *Individually Funded LTER Site Projects*. Together the sites comprise the LTER Network community where interactions

- are actively supported via annual meetings, multi-site working groups, newsletters, and mailing lists;
- *LTER Governance*. The LTER community is self-governing with by-laws that describe the work carried out by an Executive Board, Science Council, and standing committees, one of which is the Information Management Committee;
- *Independently Funded Services*. The LTER Network Office was initially the single independently funded service while network data management and digital systems were overseen by LNO as a standing committee and by NSF, which often led to complications in setting priorities; this changed in 2016 when network data management and NIS merged into the Environmental Data Initiative (EDI) to become an independently funded entity that partnered with LTER;
- *Partners*. Partners have added significantly to the community and have changed through the phases;
- *NSF Funding Source*. NSF is the agency that has funded LTER over the decades with support and input provided in time by multiple directorates.

This graphic is a simplification, reducing a complex web of interactions to a few components in order to portray the LTER communication structure as an ecosystem. It underscores the extent and

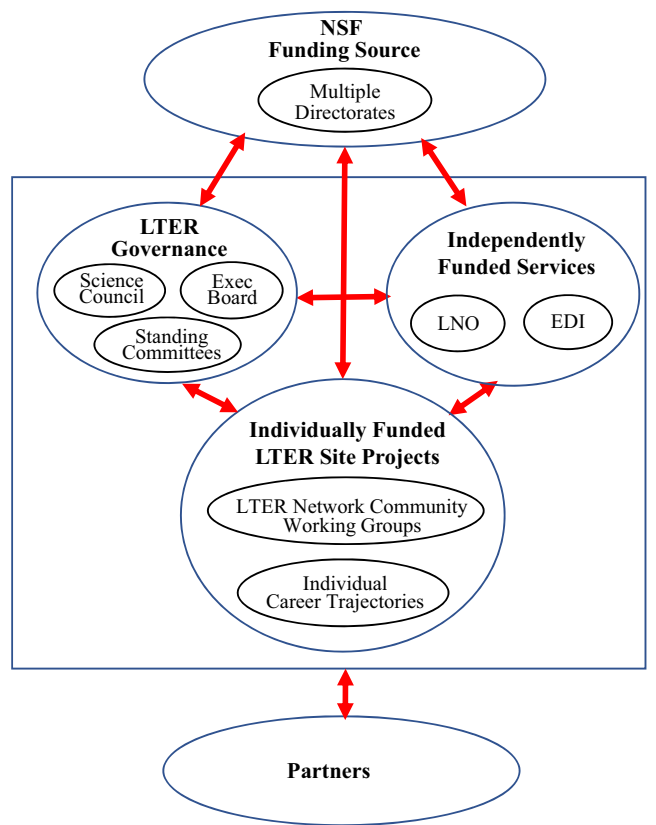


FIGURE 1 | LTER communication structure: An ecosystem of reciprocal interactions (Drawing from Waide and Kingsland 2021, LTER <https://lternet.edu/network-organization/>). This version shows arrangements after EDI was funded in 2016. Prior to this, the LTER Network Office was the only independently funded service.

dynamics of communications that tie together LTER participants. The figure does not intend to show decision-making status or the respective influence of actors. Instead, it suggests that actions are negotiated, considered not just from one perspective but reconsidered, and informed by input from across the various components.

(4) **Data Management:** The designation of site-based data managers raised and sustained awareness of data practices and data collection. Local site data work was continuously informed and influenced by data activities of researchers, data managers at other sites, and in time by development of the Network Information System. Annual funding for IMC meetings advanced understandings of data diversity, sharing of site-based strategies, and approaches to reaching network goals. The shared experiences and assembly of information from all the sites spurred development of data sharing and engagement with the Network Information System. The close relationships of data workers with researchers at each site meant data activities remained aligned with everyday research practices and were tailored to facilitate data flow. Data workers associated with sites, LNO, and partnerships were recognized, resisted, and eventually trusted as change agents. Working within the research community, data workers in partnership with researchers became sustaining agents for Open Data efforts that enhanced scientific research and transformed the community culture.

(5) **Technology Engagement:** For a long-term approach with minimal resources for technologies associated with data, LTER participants adopted practical approaches. The resulting development of new practices and procedures created time and opportunities for collective consideration and stepwise change while avoiding the burden of technology-oriented grand promises and partnerships often accompanied by great expectations. The minimal funding approach prompted incorporation of short-term goals and products within long-term plans (Karasti, Baker, and Millerand 2010; Baker and Millerand 2010), creating a pace that enabled community participation in development of data standards (Millerand and Bowker 2009; Millerand et al. 2013), and giving rise to data strategies tailored to local interests (Millerand and Baker 2010).

(6) **Community Information System:** Imagining and enacting the community's network information system was done collaboratively by participants at the network office and the sites. Data management's sheltered position within a long-term scientific research community shaped the growth of the data-oriented digital realm. The IMC and a Network Information System Advisory Committee (NISAC) brought together data and technical specialists with researchers and LNO participants (Stafford 2021). The science-driven community prompted and reviewed plans for data and computing via the IMC and NISAC. Negotiated understandings within the IMC sometimes were presented in the form of best practices (e.g., LTER EML Best Practices for LTER Sites 2004).

3 | Methods

We use qualitative methods for this ethnography of infrastructure growth for a large community composed of individuals with differing viewpoints. Closely related to this case are studies at multiple locations over many years from which Pollock and Williams (2009, 2010) developed the notion of “strategic

ethnography” for studying infrastructural technologies. This approach to ethnography incorporates theoretically informed, multi-site, longitudinal investigations necessary to gain insights into the how, what, and when of infrastructure growth.

While ethnographic methods are diverse, they make use of tools such as case studies, participant-observer roles, and quotes to capture participant voices. We worked alongside community members and also joined as genuine participants in the design process. From this “in house” position, we found interactions with LTER participants highly informative. Though providing detailed knowledge of data activities, we were aware an insider role could introduce potential biases from a site-based, data-oriented view. Collaboration, colearning, and entanglements were part of this long-term study. These are notions explored by Hahn et al. (2018) who focus on the specifics of infrastructure events and activities that prompt collective reflection.

The first author of the present study spent a number of years (1990–2011) as data manager at two LTER sites and was active in the all-site IMC as well as in working with the LTER NIS-data team and NISAC. After the first decade, social scientists were invited to join in studying data management, data practices, and design of community information systems. This included the second author who studied LTER activities starting in 2004. Study continued thereafter with both more intense and less focused periods with the information management community, thereby blending over time both insider and outsider perspectives. From the second author's position within the LTER, a deep insider understanding of the data work and design activities developed. She witnessed and experienced many changes over time in collaboration with the first author.

Working with the LTER community included participating in meetings, conference calls, and workshops attended by a mix of researchers and data workers. We engaged in discussions sometimes at their request. Community member perspectives are presented in the following section using references and excerpts from their writings. This study has also drawn on digital documents maintained in personal files of the authors, the LTER community archive, and notes from sidebars at LTER data management meetings. The generation of field notes and field memos provided a record of immediate impressions and insights.

Our special relationship with LTER resulted in publications about data management as a community of practice (Karasti, Baker, and Halkola 2006; Karasti and Baker 2008; Baker and Millerand 2010), community metadata standard making as a process (Millerand and Bowker 2009), and the power and tensions involved in collaborations infused with “data troubles” (Millerand et al. 2013; Baker and Karasti 2018) including termination of a site (Kaplan, Baker, and Karasti 2021). Multi-community, comparative studies that include LTER have provided further insights (Karasti and Syrjänen 2004; Ribes and Finholt 2007, 2009; Mayernik, Batcheller, and Borgman 2011; Ribes 2014).

In addition to information drawn from document analysis and previous studies, we also rely on recent interviews and informal meetings. In the last 6 years, our study conducted in-person, semi-structured interviews typically of 1.5 h within the LTER community. These included: researchers (4), data managers (18),

information technologists (4) and program managers (2) from LTER sites, the LTER network, and NSF. These interviews focused on questions of design as well as data practices and infrastructure growth. One annual meeting of data managers in 2017 was filled with discussions about change by participants from new sites, mature sites, and terminated sites. Several in-depth interviews, with one information technologist spurred our approach to seeing the development of the LTER NIS in phases.

Analysis began with open coding of transcriptions of audio-recorded interviews and meetings. A number of synthetic timelines, overviews, and diagrams were created. Data triangulation was used to bring these materials together. We focused on the dual themes of data management and the Network Information System. In aiming to capture the 40-year NIS trajectory, we developed, reworked, and reconfigured tables of category codes to identify major characteristics of significant events in the development of the LTER NIS.

Sampling across stakeholders is uneven over the period of this study, a limitation ameliorated by the review of historical reports and other documents. Additional interviews with information technologists and researchers would have improved the study. Another shortcoming is the loss of complexity and richness in order to convey the material within the length limitations of a journal article. Many important activities and events have been eclipsed.

4 | Growth of the LTER Network Information System

The Long-Term Ecological Research (LTER) Network program provides a unique example illustrating the evolution of

an information system. The conceptual framing for our analysis draws on the notion of incremental growth. We show the growth of the LTER NIS as a sequence of six relatively stable phases within a multi-level environment (Figure 2). For our research questions, these phases address the “how” of infrastructure growth.

Although many activities occur during a phase, our focus is on pivots that bring about a major change in data arrangements. Table 2 lists six major pivots and provides some context for each phase in the form of a brief timeline with dates of activities and digital products. The phases identified for the LTER NIS are described below. Each pivot initiates a phase featuring particular data work arenas where joint experiences and discussions occur. Below each phase is discussed together with changing community assumptions and the digital context with other efforts.

4.1 | Phase 1: Embedding Data Management at LTER Sites (1980–1990)

The role of data manager was expected to appear as a line item in each site's budget. Naming of the data management position provided significant visibility to data-related work. This role was a major change for research projects in 1980. With the prospect of continuing but limited funding for LTER from NSF, resources were stretched in supporting both site-based collaborative research and the long-term collection of field data. Tensions developed relating to supporting a data manager. Ecologists' exposure to technology in the field or lab often made advances easier to accept when they related to collecting the data rather than managing the data collectively for future data reuse. Traditionally researchers' primary accountabilities were related to their own

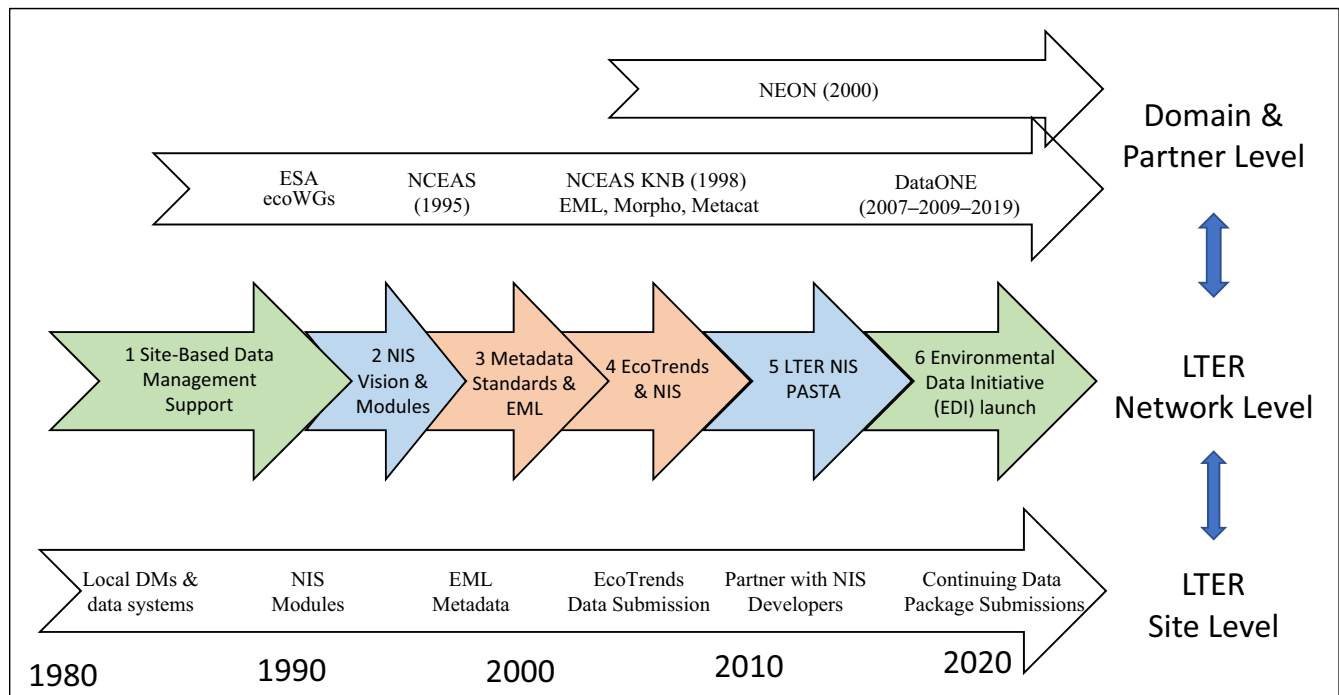


FIGURE 2 | The six phases of the LTER Network Information System (NIS) trajectory are labeled “LTER Network Level.” Below this level are data activities at the “LTER Site Level” and above are related activities at the “Domain & Partner Level.” The phase colors highlight: Green, scaling prompts from external sources (see Section 5.1.6); Blue, internal partnering (see Section 5.1.4); and Red, external partnering (see Section 5.1.5).

TABLE 2 | Each pivot in the NIS trajectory is identified by a phase name and its focus. Dated events are listed to provide context followed by the approximate duration of each phase.

Name pivot & phase	Focus of the phase	Data activity timeframes	Duration
1. Site-Based Data Management Support	Embedded data manager for local data assembly, management, and system development	1980 site DM role required 1982 first DM workshop 1983 annual DM meetings begin 1989 first Databits newsletter 1990 all-site Core Dataset catalog printed	10 years
2. NIS Vision & Modules	LTER Network Office computational capabilities developed together with a Network information System Framework while site online capabilities expanded together with design of NIS modules for data sharing	1990 Data policy committee formed to address information policy and data sharing 1993 10yr Risser Report on LTER 1994 Each site shares one dataset online 1996 Developing NIS framework and vision 1997 LTER data sharing policy 1997 All-site bibliography	10 years
3. Metadata Standards & EML	Ecological Metadata Language (EML) explored, prototyped, and adopted as LTER metadata standard with metadata submitted to centralized Metacat catalog	2000 Site websites continue developing 2001 Partnership for Biodiversity begins at NCEAS 2002 20yr Harris-Krishtalka Report on LTER 2002 LTER NIS Advisory Committee formed 2002 EML adopted by IMC & LTER 2003 NCEAS Metacat development 2003 LNO hires staff for NIS	4 years
4. EcoTrends & NIS	EcoTrends, a project that became a NIS module synthesizing key long-term variables, aggregates data from sites to a centralized EcoTrends system using the Metacat catalog and is designated a NIS module	2004 EML Best Practices v1.0 2004 LTER Decadal Planning Process (2004-2007) 2006 EcoTrends project funded (2006-2013) 2007 Network Cyberinfrastructure Strategic Plan	6 years
5. LTER NIS PASTA	Centralized NIS PASTA SOA architecture, data package quality report & data portal development by LNO in partnership with the Information Management Committee (IMC)	2010 LTER NIS funded via ARRA (2010-2014) 2011 NSF 30yr Michaels-Powers Report on LTER 2011 NSF 30yr minority DM Robbins Report 2011 EML Best Practices v2.0 2011 Data Quality Report development begins 2013 LTER NIS and Data Portal launch 2013 Publication of EcoTrends book	4 years
6. Environmental Data Initiative (EDI) Launch	LTER Network Office split into LNO & EDI; LTER NIS rescoped as Environmental Data Initiative (EDI) system with a centralized repository for multiple communities	2014 NISAC report on NIS 2015 LTER NCO UNM to UCSB/NCEAS 2015 McKnight Task Force re LNO 2016 EDI (re)funded 2016, 2019, 2022 2017 NIS with repository launch 2019 & 2022 EDI refunded	6 years

research, not to external data standards, metadata creation, and data sharing (Mayernik 2017; Baker and Mayernik 2020). Researchers were trained in collecting and using their own data. Data managers themselves were grappling with identifying the responsibilities of data work. Annual meetings and monthly video conferencing of the all-site IMC strengthened cross-site ties with other data managers. Early work included physical artifact collections such as specimens, formatted data sheets, and print data reports. The creation of digital databases fueled growth of site computational capabilities. Assembly of data often went hand in hand with development of site-based data systems. This experience positioned data managers as data champions able to identify and articulate data management

issues at the opening of the 21st century with developments in digital technologies.

Collective data management was received as an unfunded expectation by LTER researchers though small additional funds for data management were made available intermittently. This required data workers to collaborate closely with site leaders to design and document data needs in small proposals to NSF that generally avoided the quick fixes often promised by high-cost technologies. In working closely with researchers who generate the data, data managers emerged as a data workforce in sync with local research. Eventually data managers were able to look past the researchers' traditional horizon of data use for grant

requirements and individual career advancement, to see new responsibilities of planning for data *reuse* by others.

During this NIS phase, the community assumption that “Data is best managed by individual researchers” shifts to “Data for multi-investigator research efforts benefits from continuing development of project data management and data services.” This assumption shows recognition of the need for new ways of thinking about “data sets” by researchers (Franklin, Bledsoe, and Callahan 1990) and of the responsibilities that emerged for the role of data manager (e.g., Ingersoll, Seastedt, and Hartman 1997). At a time when the majority of researchers managed their own data, the LTER NIS trajectory Phase 1 raised awareness of the diversity of data and of managing data collectively. As researchers began asking longer-term and larger-scale research questions (Magnuson 1990; Swanson and Sparks 1990), the data managers were taking steps that would enable addressing their own longer-term and larger-scale questions about data issues including Open Data.

During this period, the U.S. Joint Global Ocean Flux Study (JGOFS, 1989–2005) was another early field research program that developed program-oriented data management. Though the JGOFS centralized development differed from that of the LTER NIS trajectory, a JGOFS data office and repository was transformed in 2006 into a national Biological-Carbon Oceanography-Data Management Office (BCO-DMO, <https://www.bco-dmo.org>) for ocean sciences. It was funded by the Division of Ocean Sciences in the NSF Geosciences Directorate (Glover et al. 2006; Baker and Chandler 2008; BCO-DMO 2013). BCO-DMO continues to work today with oceanographic investigators via a central Data Management Office.

4.2 | Phase 2: Envisioning a Distributed LTER Network Information System (1990–1999)

By 1990, internet connectivity brought dramatic change in terms of information access and data expectations. LTER published a Catalog of Core Data Sets in print form (Michener, Miller, and Nottrott 1990) with information about one dataset per page. Data accessibility was frequently noted as “digital, tape.” Due to rapidly evolving capabilities, however, electronic versions of a personnel directory and a Core Data Set Catalog were available at LNO by 1991 (Brunt 1998). Soon after, a NIS Interoperability Framework for a distributed data system was developed that depended on LNO computational technology as well as on specifications developed for harvest, exchange, metadata, indexing, and interfacing. All sites lacking digital capabilities were given critical support by LNO and IMC.

Making the data itself publicly available began network-wide with an NSF prompt to LTER. In 1994, LTER leaders agreed that each site would make one dataset available on their site's webpages as a demonstration. Slowly the number of shared datasets online grew. As network planning and an increase in cross-site research activities added to the interest in compiling data from all the sites (Johnson et al. 2010; Collins 2021), the vision for a Network Information System advanced. Development began with data managers at different sites and combinations of sites designing and hosting data modules including a climate database (Henshaw

et al. 1998), a site description directory (Baker et al. 2000), and then a digital Data Table of Contents or DTOC (Porter 2010). DTOC was a precursor to updated catalog efforts in Phases 3 and 5. Leveraging the LTER Network Office technical resources, some modules were migrated to LNO. From prototyping work on NIS modules, data managers learned critical lessons about the little-recognized work of “local enactment,” that is, about the socio-technical activities required to align existing local data practices and systems with network applications (Millerand et al. 2005).

During this NIS phase, the community assumption that “Data can be stored using individual and laboratory storage devices” shifts to “Data from multiple sources can be brought together for sharing and reuse via submission to coordinated information system modules.” While a few researchers in Phase 2 envisioned some standardized approaches to network-wide data handling (e.g., Franklin, Bledsoe, and Callahan 1990), others gained experience from the development of local data practices and data systems. In contrast to top-down designs, the data managers' first approach to cross-site data aggregation was to create data modules themselves since no money was budgeted for this (e.g., Henshaw et al. 1998; Baker et al. 2000). Participants from all the sites agreed to populate these modules, thus launching network information sharing and eventually prompting the development of data-sharing policies, first at individual sites and eventually unified at the network level (Porter 2010). This phase demonstrates the LTER sites moving forward together given their collective involvement in the “how” of assembling centralized information.

During this period, access to the World Wide Web led many data managers in the 1990s to become their site's webmaster who organized online site content. This enabled sites to post project information, thereby contributing to site self-awareness and eventually to making data available. In reviewing the web pages of other sites over time, researchers were exposed to a variety of materials, visualizations, and categories of information. These websites also enhanced site identity by prompting lively discussions about presentation of a site's research priorities and data priorities.

4.3 | Phase 3: Partnering on Metadata and a Metadata Catalog (2000–2004)

From 1995, advances in computing, technology, and communication had increasing impacts on scientific work. Those managing data were contending with an array of instruments, procedures, documents, formats, and analyses. In assembling data from multiple sources, the need to annotate datasets with descriptive information called metadata became evident. Each site developed metadata that provided highly structured information including basics such as the dataset title. As the number of datasets increased, the metadata enabled sorting, discovering, and access to data.

LTER's work on metadata began in conjunction with the Ecological Society of America's Future of Long-Term Ecological Data Committee (Michener et al. 1997), a committee with LTER members. NCEAS was active in supporting working groups of ecologists including LTER scientists, Software engineers at NCEAS were important partners for LNO and LTER data managers in development of data tools and data storage services as

well as the Ecological Metadata Language (EML). After extensive work with NCEAS on metadata file formats, LTER adopted EML as the LTER metadata standard in 2002. Although aggregating data in a standard form had created difficulties earlier with the first LTER all-site bibliography at LNO (Chinn and Bledsoe 1997), it took years to recognize that the movement of data from site systems to a community repository in a standardized form was a major issue rather than simply a local trouble (Millerand et al. 2013).

During this NIS phase, the community assumption that “Data can be described by metadata developed at a local level” shifts to “Data can be shared with larger audiences by mapping local metadata to metadata standards.” As LTER data managers aggregated data prior to Phase 3, they developed site-specific metadata for use in local databases (Porter, Henshaw, and Stafford 1997; Millerand and Bowker 2009). Sites typically did not convert to EML locally but maintained their rich local metadata tailored to capture descriptive elements of a site’s datasets for findability by local data system users. Sites eventually developed automated applications to map local metadata to EML prior to submission to NCEAS’s metadata catalog (called Metacat) (Berkley et al. 2001). This centralized catalog and discovery system held metadata from all the LTER sites, initially with links to data stored at the sites. Local data systems served as site-based gateway systems, eventually enabling automated machine-to-machine data submissions as well as the application of filters for data delivery. Questions that arose relating to the meaning of metadata entries for different communities led to LTER participants writing a customized LTER guide to the interpretation of descriptive tags for their community (LTER EML Best Practices for LTER Sites 2004). Thus the “how” of infrastructure for the LTER case involved partnering that expanded capabilities while documenting the particular needs of the LTER community.

During this period, partnerships contributed to formulating plans to advance data access. A Partnership for Biodiversity Informatics (PBI) working group was hosted by NCEAS in 2001 to pursue joint data interests, bringing together partners, including NCEAS, LNO, the San Diego Supercomputer Center (SDSC), and the Natural History Museum and Biodiversity Research Center at the University of Kansas, with the aim of giving access to information needed to sustain the earth. NCEAS led the multi-institutional Knowledge Network in BioComplexity (KNB) project (Jones et al. 2006; Stafford 2021). Ecology as a whole benefited from the continuing efforts to raise awareness about data being both hard to find and hard to reuse. Although there were tensions, particularly associated with technical development, the role NCEAS played as an active center promoting data access was critical (Fegraus et al. 2005; Jones et al. 2006; Reichman, Jones, and Schildhauer 2011; Hampton et al. 2015).

4.4 | Phase 4: EcoTrends and LTER NIS (2005–2013)

The LTER Network Office was moved to the University of New Mexico in Albuquerque in 1996 with the promise of increasing both university space and LTER resources including those for information system efforts. The slow progress in access to data resulted in impatience for some researchers. Researchers who recognized the opportunity afforded by EML, proposed

EcoTrends as a project to develop an ecological information system. The specific intention was to provide data access for data synthesis (Peters et al. 2013). Funded in 2005, EcoTrends leaders were LTER investigators who also worked for U.S. federal agencies. Fifty biome study sites distributed largely across the United States were funded to participate by NSF funding for LTER and by two U.S. Department of Agricultural agencies—the Agricultural Research Service (USDA-ARS) and the Forest Service (USDA-FS). Ties were maintained with NCEAS for technical support of remote data submission and ingestion into an EcoTrends data collection. The focus of EcoTrends was on producing data products for ongoing scientific analyses. LTER data managers collaborated with NCEAS and USDA partners by submitting and checking data.

The LTER Coordinating Committee approved EcoTrends as an LTER Network Information System (NIS) module. The LNO NIS-data team was responsible for the use of the information system and of the Metacat metadata catalog in addition to the development of the project website. With LNO and NCEAS seen as a digital hub, the EcoTrends website became a familiar gateway for data submissions. LNO developers were familiar with individual data managers, the LTER IMC, and the dynamics of the LTER research community as a whole. Their attendance at LTER data manager meetings meant they were readily available and in constant contact with data managers. Together, these data groups served as a bridge between “researchers as data generators” and “researchers as data users.” An EcoTrends project coordinator proved critical to coordination and assessment of submitted data. Maintaining bilateral communications with site participants throughout the project, ensured ingestion of high-quality data. Difficulties discovered with uploaded data were logged into a spreadsheet reporting system that was a hands-on approach to data checking. Site-based and cross-site LTER scientific groups used this data to generate derived datasets that supported scientific publications. Previously such data was rarely available for discerning larger-scale and longer-term trends.

During this NIS phase, the community assumption that “A community data system to support scientific research is typically developed by computing professionals working for one community with funding from a single source” shifts to “A community data system to support scientific research can be led by researchers together with data workers and software engineers from one or more communities with support from one or more sources.” EcoTrends resulted in the development of an ecological information system for a number of networks of sites (Servilla et al. 2006; Peters et al. 2013). Unlike in Phase 3, where the primary leaders funded for NIS development were software engineers, the EcoTrends funded leaders were researchers who collaborated closely with software engineers and data workers. A number of new databases and communication processes were established. EcoTrends created opportunities for shared data work by researchers and data workers (Laney, Peters, and Baker 2013). EcoTrends also represented an additional workload and new partnerships at a time when LTER researchers and the NIS-data team were considering a more comprehensive NIS architecture. NIS developers explain: “Our frustration with this [EcoTrends] labor intensive approach to providing useful ecological data gave us motivation and guidance in our development of PASTA [the

Provenance Aware Synthesis Tracking Architecture]” (Servilla et al. 2016). As data managers worked on data procedures for providing quality data to EcoTrends, researchers were gaining experience with Open Data as EcoTrends made centralized data available for their immediate use.

During this period, EcoTrends datasets were being made available and NCEAS continued development of its Metacat repository. By 2009, a number of other groups were investigating and prototyping how to make data available. For instance, the Dryad effort (<http://datadryad.org>) was begun by library and publishing communities to provide an open data publishing platform (Vision 2010). For researchers with no repository available for their domain, it aimed to create a generalist repository that would simplify the process of linking research papers to related datasets. Dryad continues today as one of a number of generalist repository options.

4.5 | Phase 5: An Updated LTER NIS Framework (2006–2013)

A three-year NSF-funded planning process by LTER to improve network-level science began in 2004 in response to LTER reviews (Collins 2021). A model was created to address research needs that included restructuring governance and increasing cyberinfrastructure (LTER Decadal Plan 2007). Interactions with NSF failed as the agency requested a strategic plan rather than a proposed conceptual framework for social-ecological research. An LTER Network Cyberinfrastructure Strategic Plan (2007) was not funded but was revised as an LTER NIS proposal (Servilla et al. 2016; Stafford 2021). This redesign was funded due to ARRA “stimulus funds” made available to NSF as part of government efforts to stimulate the U.S. economy. Collins (2021) noted, “Because NSF forced the LTER Network to develop a very detailed SIP [Strategic Implementation Plan], including plans for an advanced information management system to support synthesis, the LNO was poised to receive ARRA funding through NSF.” The LTER NIS ARRA proposal (2010–2014) provided a budget to support development staff and a web-based data portal (Servilla et al. 2006, 2008).

NIS plans explained that “individual and/or collections of PASTA framework services may provide utility to sites by offering a centralized and off-site data repository and distribution point, metadata and data congruency checks for data quality assurance and correctness, automated metadata generation and management, data access and use audits, and a source of workflow and data transformation algorithms” (Michener et al. 2011a). The architecture was designed for the upload and management of site-generated data packages. The concept of a data package highlights the need for data submissions to include data and metadata together. Each package would be assigned a unique Digital Object Identifier (DOI). As the LTER data management community grew due to an increasing number of sites with an increasing number of data packages, the expectations and responsibilities for site data management increased as well. Bylaws were developed by the IMC as it matured as a community of practice to become a consortium with codified rules of governance (Baker, Kaplan, and Melendez-Colom 2010). A range of viewpoints on and assumptions about data management

contributed to discussions of whether minimally funded local data efforts were effective (e.g., LTER Michaels-Powers Report 2011; Robbins 2011).

During this NIS phase, the community assumption that “A community information system can evolve into a cyberinfrastructure that provides basic access to data and metadata for a single community for discovery, access, and integration while partnering with other environmental observatory networks” shifts to an expanded and updated vision for LTER NIS where “A community information system provides advanced data capabilities that support research by establishing automated checks on data quality in addition to making data queriable, downloadable, and uniquely identified for tracking updates.” In Phase 5, the NIS approach that emerged after EcoTrends was an updated version providing advanced computing services using a service-oriented architecture (SOA) (Servilla, O’Brien, and Costa 2013). This increased technical flexibility via improvements in the web interface, an LTER data repository, and provenance metadata for tracking data packages (Servilla and Brunt 2011; Servilla et al. 2006, 2016; Michener et al. 2011a; Waide and Kingsland 2021). The availability of ARRA funds and the readiness of the NIS-data team explain “how” the NIS advanced during this period to become an updated, operational community data infrastructure able to support Open Science with LTER Open Data.

During this period, NSF began funding a new cyberinfrastructure program aiming to manage geoscience data. EarthCube (2011–2022, <http://earthcube.org>) was a 10-year joint initiative between the NSF Directorate for Geosciences (GEO) and the Division of Advanced Cyberinfrastructure (ACI). This effort differs from the LTER and BCO-DMO programs with their incremental growth of data infrastructure. Visioning and development for EarthCube were carried out by a number of groups of high-performance computing specialists partnered with geoscientists each with existing data collections from within the various fields comprising the geosciences. Many individual projects were funded to provide access to the data collection. The intention was to have these collections integrated into a national technical platform when its design was complete. This program was paused in 2021 to review and reformulate thoughts about domain-wide services that included the diverse geoscience fields (EarthCube Office and EarthCube Leadership Council 2022).

4.6 | Phase 6: Environmental Data Initiative: A Data Management Office and Repository (2014–Ongoing)

While the LTER NIS-data team worked on expanding NIS into a proposed LTER cyberinfrastructure, a major change for LTER was initiated by NSF with the announcement of a re-competition for the LTER Network Office. An open call for proposals focused on support for the LTER scientific communication and education efforts of the LTER without mention of support for data efforts. The award for this Office went to NCEAS at the University of California, Santa Barbara in 2015 for an LTER Network Communications Office, referred to as LNO today. Subsequently, an NSF call was issued for proposals to open a new Data Management Office

(DMO) to support not only centralized data submissions from LTER participants but also from individuals and communities in both ecology and the environmental sciences.

A number of data managers were interested in leading an expanded DMO effort that would then be anchored at their site. Discussions within the LTER information management community and at the sites led to recognition of the need for an IMC vote to select one of these efforts to send forward. A proposal led by the University of Wisconsin-Madison was selected by the LTER data management community and submitted. A proposal by the LNO NIS development team was also submitted to expand the LTER data repository into a multi-community framework (sometimes referred to as updating PASTA to PASTA+) to support multiple environmental communities and individuals. At NSF's suggestion, these two efforts were merged into a single successful proposal. This grant established the two-component Environmental Data Initiative (EDI): (1) technology services associated with the updated data repository infrastructure at the University of New Mexico at Albuquerque and (2) a Data Management Office at the University of Wisconsin at Madison, collocated with the North Temperate Lakes LTER site (Stafford 2021).

During this NIS phase, the community assumption that "A community information system can be developed to provide access to structured data and metadata for the LTER community" shifts to "An advanced community information system developed for a single community can be scaled up to include new services and a number of audiences." After more than three decades of growth for LTER NIS and local data systems, Phase 6 began with the splitting of the LTER Network Office into two offices that reoriented LNO and established a multi-community data effort. EDI became an independent program with funding to support ecology and environmental sciences data (Gries et al. 2021, 2023). The LTER data repository was merged into the EDI data archive with new features added. EDI today is exploring the archive of analysis-ready data such as observations from monitoring networks. Further, EDI is a member node of the DataOne federation (Michener et al. 2011b) of data repositories. This phase makes evident how scaling up and reconfiguration of the data infrastructure was accomplished to support Open Data from ecology and the environmental sciences.

During this period, agile concepts were becoming recognized for digital infrastructure development. NSF had begun a program in the mid-1990s for funding large-scale science projects through Major Research Equipment and Facilities Construction (MREFC) rather than through a particular directorate, as was the case for LTER, BCO-DMO, and EarthCube. Initially, MREFC was for major physical structures using a design approach following a "waterfall" model. This model called for design plans to be carried out over a 15- to 25-year period before building of a facility would begin. The notion of "implementation" prior to "operation" was introduced in 2019 for what MREFC now was calling "large facilities" (NSF Major Facilities Guide 2019). Steinhardt (2016) observed use of spiral prototyping during work on the National Ecological Observatory Network (NEON, <http://neonscience.org>). NEON is a centrally administered observing network for digital data generation and distribution that was envisioned beginning in 2000, funded by MREFC in 2011 with an operational launch in 2016, and full operation in 2019 that continues today. In the international sphere, there also has been work with prototyping cycles (Chabbi and Loescher 2017; Kaltenbrunner 2017) that contributes

to the loosening of digital infrastructure planning processes to allow for unforeseen pivots and actions sometimes referred to as "infrastructure growth" rather than "built infrastructure."

5 | Discussion

Having described the phases of incremental growth for the LTER NIS, in this section major characteristics of change are followed by overviews of the shifts in community data management assumptions and of the ongoing support for community interactions.

5.1 | Incremental Growth Phases: Characteristics of Change and Key Findings

Establishing site-based data management (Phase 1) and collectively envisioning network data coordination (Phase 2) are community-defining achievements often overshadowed by advances such as the development of metadata and metadata standards (Phase 3) that support the flow of data to centralized data systems (Phase 4). As the LTER understanding of data work matured, a new data system was designed and launched from within the community (Phase 5), a system that was expanded upon subsequently in response to a call for a data management office to serve ecology and the environmental sciences (Phase 6). The complexity of growth is reflected in the number of characteristics identified below to describe each phase. These characteristics identify and name aspects that elucidate the "what" of data infrastructure, one of our research questions.

Major characteristics of the six phases in the NIS trajectory are presented in Table 3. Characteristics are grouped into three categories: scope, change element(s), and lead actor(s). There are two scope characteristics: column 2 shows the dominant arena of NIS activity and column 3 indicates the intended users of the information system. Change elements (columns 4–7) are defined earlier in the Background Section. The final three columns indicate leadership leaders in terms of responsible PIs for funded proposals (column 8), DM/Tech efforts (column 9), and major partnering contributors (column 10).

Assessing change in each characteristic over time reveals patterns, anomalies, and notable relations. Six key findings related to the incremental growth of data infrastructure emerge from such analysis across the phases of LTER NIS.

5.1.1 | Embedded Data Management

Establishing a data management role at each site was a pivotal event with long-term ramifications. Table 3 (columns 8 and 9) shows that the data managers and the IMC play an ongoing role in NIS design in collaboration with LNO. These data-related roles seeded the development of a workforce to support immediate site data needs and to shape longer-term, Open Data efforts. A nuanced understanding of and respect for differences in the data and digital circumstances was stimulated by interactions with other sites. Eventually data workers developed a voice in research arenas (e.g., Michener 1986; Ingersoll, Seastedt, and

TABLE 3 | Major characteristics of each phase identified during the growth of the LTER Network Information System.

	Scope	Scope	Change element	Change element	Change element	Change element	Lead actor(s)	Lead actor(s)	Lead actor(s)
1 Name pivot & Phase	2 Level/arena of activity	3 Reach/audience(s) served	4 Change facet	5 Major change impetus	6 Kind of change	7 Generative mechanisms	8 Principle investigator	9 DM/tech	10 Major partners
1. Site-Based Data Management Support	Site	LTER sites	Institutional Sociotech	Funding grants	Planned Emergent	Scaling	LTER site scientists	DMs & LNO	NSF DBI gives LTER program support (ongoing)
2. NIS Vision & Modules	Network	LTER Network & sites	Sociotech	Site-network partnering	Emergent	Innovation & Adoption	LTER site scientists	LNO & IMC	NSF
3. Metadata Standards & EML	Multi-Community	LTER Network & sites & ecology	Sociotech	Strategic partnering	Planned	Innovation & Adoption	NCEAS & LNO	NCEAS & LNO & IMC	NSF, NCEAS & KNB
4. EcoTrends & NIS	Dual-Community	LTER & USDA, ecological & agricultural sites	Sociotech	Strategic partnering	Planned	Innovation & Adoption	USDA scientists & LTER	LNO & partners & IMC	NSF, USDA & NCEAS
5. LTER NIS PASTA	Network	LTER Network & sites	Sociotech	Site-network partnering with LNO project funding	Planned Emergent	Innovation & Adoption	LNO & LTER site scientists	LNO & IMC	NSF
6. Environmental Data Initiative (EDI) Launch	Multi-Community	Ecology & environmental sciences including LTER network & sites	Institutional Sociotech	Funding grants	Planned Emergent	Scaling	EDI	LNO & IMC	NSF DBI & DEB support EDI (ongoing)

Hartman 1997; Michener, Porter, and Stafford 1998; Benson et al. 2006; Michener et al. 2011a; Gries et al. 2016; Kaplan, Baker, and Karasti 2021). Embedded data managers at sites and LNOs led to researchers becoming familiar with data issues on an everyday basis. Data managers became facilitators as well as advocates for change, gaining respect as they supported research and served as both data and technology consultants at their site.

5.1.2 | Changing Scopes

Another key finding is that the information system scopes—both work arena level and audience shift in sync over time whether increasing or decreasing, rather than evolving in a linear fashion. The scopes of NIS during development change (columns 2 and 3), sometimes increased over the phases (Phases 1–2) while at other times they shifted to produce an interim product (Phases 3–4). The scope also contracts while regrouping in response to a community-centric infusion of funds (Phase 5) before then expanding to serve a broader community (Phase 6).

5.1.3 | Awareness of Social and Technical Interdependence

A key finding for this case is that a continuing awareness of sociotechnical issues during the growth of an information system ensured community understanding and use of the information system as it changed over time. Collaborative data work calls for sensitivity when addressing changes in existing data practices

and data systems. Awareness of the concept of “sociotechnical” was crucial in all six phases (column 4) ensuring high priority was given to data work and data sharing by LTER participants. We note that though EcoTrends scientists succeeded in producing a variety of published outcomes, the project's data system was not sustainable. Subsequently, in Phase 5, lead actors were from the NIS-data team with experience with the sociotechnical aspects of LTER data practices and system design.

5.1.4 | Internal Partnering for Emergent Growth

Emergent growth of data management during development of a network information system was fostered by the data manager's close collaborative relationships with a combination of other data managers, researchers, and developers within the community. There are phases of emergent growth of the LTER NIS (column 6) during phases 1, 2, 5, and 6 that underscore the NIS-data team work. In these phases, the leaders were situated within the LTER itself rather than external to the community (column 9). Internal relations within the community take a variety of forms such as working groups, workshops, and a small group of data managers working with NIS developers on a targeted topic, referred to as “tiger teams.”

5.1.5 | External Partnering for Planned Growth

Another key finding is that intermittent work with external partners (column 10) was critical to the growth of the community

and the community's information system. Partners contributed to developing approaches, procedures, and services for work with data and technology. Their work is detailed in Section 4 by phase. In addition, the NSF funding agency played a significant role as an institutional partner in communicating with LTER participants and guiding research arrangements.

5.1.6 | Scaling Prompts From External Sources

Major increases in audience for data management and computing efforts were spurred by external interventions in the form of institutional mandates associated with funding. Working with community members, NSF with great vision and remarkable insight established LTER. Dramatic change in research arrangements was introduced in Phase 1 via a mandate for each site to designate a data management role tied to their long-term funding. The role was made explicit in NSF proposals, yet loosely defined, leaving the flexibility for it to evolve at the project level. Another example of a scaling prompt was NSF's call for proposals in Phase 6 to create a national data initiative that served LTER and multiple other communities. NSF funding supported expansion of the LTER NIS from a single-community information system to its current state with EDI. Scaling prompts are reflected in the audience served (column 3), the institutional change (column 4), and the impetus enabling change (column 5). Program and project funding represents a significant generative mechanism for scaling data management and data system activities (column 7).

These six key findings from assessing change during the NIS incremental growth over time appear to sidestep some of the “path dependence” or “lock-in” associated with traditional major digital endeavors with more hierarchical structures (Edwards et al. 2007; Hirsch, Ribes, and Inman 2022). For example, in Phase 2 sites were successful in making local data accessible and prototyping network modules within the community. In contrast, during Phase 3 sites shifted focus to work with external technology-oriented partners on an ecological metadata standard. Yet, lock-in did not occur when these partners presented metadata generation tools. The tools were explored but typically not adopted at LTER sites since they did not align with the diverse local data workflows.

5.2 | Changing Assumptions Relating to Data Management

Changes in long-held perceptions or operating assumptions about data management in the LTER community over the time period of NIS development (see Section 4) are summarized in Table 4. They show the assumptions becoming more complex over time as data management itself becomes more complex.

Research has shown that unnoticed community and cultural assumptions create barriers to change (e.g., Gitelman 2013; Edwards et al. 2011). A literature survey of large-scale architectures examined the need for more widespread understanding of changing data practices across communities (Uludag

et al. 2018). This includes providing time for the development of what is described as agileness of “culture and mindset.”

Stepwise change in community understanding is evident in the LTER case. The community's uptake and conceptual shifts over the extended period suggest that the LTER NIS incremental growth provided the time required for wide-spread discussion and understanding of data practices across the network that contributed to changing assumptions about data management. There is ongoing socialization and a form of response that has been called “assumptions wrangling” (Cutcher-Gershenfeld 2018). The many year time period of each phase provided ample opportunities for community members to witness and experience digital prototypes and products as outcomes or indicators of success.

5.3 | Supporting LTER Community Interactions

The LTER program communication structure is key to making use of the time provided by the evolution of NIS. This structure (Figure 1) is identified as a contextual condition of the LTER Network (see Section 2.5.2). Its support for continuing interactions is crucial to collaborative processes involving LTER data-oriented, technology-oriented, research-oriented, and policy-oriented participants within and across work arenas. Collaboration is integral to the interdisciplinarity of ecology and the value of long-term ecological projects in terms of their potential to help participants learn to collaborate has been recognized (e.g., Mauz et al. 2012). In support of the LTER community's collaborative activities, each component in the communication structure actively processes and shares information while making decisions in both planned and improvisational ways. The non-hierarchical nature of Figure 1 suggests that actions are negotiated, considered not just from one perspective but reconsidered and informed by many participants across the components. Bringing these components together distributes power and ties together a diversity of perspectives that stimulates discussion and prompts reflection. This visual shows the extent of potential interactions and agency as well as the messiness inherent to data and infrastructure work in the LTER. Such extensive interactions are time-consuming and require patience, yet they can have positive impacts on research coordination together with management of knowledge, data, and the program. These are features identified as contributing to development and sustainability of large-scale enterprises (Uludag et al. 2018). The structure creates many potential points of learning, adaptation, and trust-building (Gunderson and Holling 2002). Such an environment requires continuing attention to balancing the predictable and unpredictable. Co-learning, or learning as a group, is fostered by such an environment. Creativity, together with learning, is fostered by such an environment. Engeström (2001) reminds us that creativity does not only happen inside people's heads.

The LTER approach exhibits many “bottom-up” insights into how technology can support effective and large-scale digital efforts. This case with its multi-level communications and work arenas, however, may be recognized more fully as encompassing multiple approaches to growth given its development in phases with differing

TABLE 4 | Transformation in LTER assumptions about digital data management by phase.

Phase	A. Original assumption	B. Transformed assumption
1	Data is best managed by individual researchers.	Data for multi-investigator research benefits from continuing development of project data management and data services.
2	Data can be stored using individual and laboratory storage devices.	Data from multiple sources can be brought together for sharing via submission to coordinated information system modules.
3	Data can be described by metadata developed at a local level.	Data can be shared for reuse by larger audiences through mapping local metadata to network metadata standards.
4	A community data system to support scientific research is typically developed by computing professionals working for one community with funding from a single source.	A community data system to support scientific research can be led by researchers together with data professionals from one or more communities with support from one or more sources.
5	A community information system can evolve into a cyberinfrastructure that provides basic access to data and metadata for a single community for discovery, access, and integration while partnering with other environmental observatory networks.	A community information system provides advanced data capabilities that support research by establishing automated checks on data quality in addition to making data queriable, downloadable, and uniquely identified for tracking updates.
6	A community information system can be developed to provide access to high quality structured data and metadata for a single community.	An advanced community information system developed for a single community can be scaled up to include new services and a number of new communities as well as individuals.

characteristics and many data work arenas (Table 3). Figure 1 suggests that change in the LTER NIS occurs (a) bottom-up with a local narrative that is maintained and promoted informally at sites, (b) middle-out via lateral or partnering to address change with a network narrative that promotes updates and revisions, and (c) top-down with a meta narrative that may be community promoted or formally mandated. LTER NIS Phases 1 and 6 in particular reflect top-down interventions that do not disturb scientific research plans. These top-down interventions do, however, seed what occurs with NIS prototyping (Phase 2), with NIS redesign (Phase 5), and with NIS evolution into a multi-community system (Phase 6).

The Phase 6 pivot that resulted in reorganization of LNO effectively separated direct oversight of the LTER NIS growth from the governance of LNO and LTER researchers (Figure 1). This represents a major shift from its earlier launch and growth environment for data management, which was nested within the LTER research community. The independent EDI organizational structure is now more similar to other large-scale digital infrastructures, many of which are led by computing, technology, and/or software engineers unlike the early NIS efforts with a number of participants having a long-term view of data practices. There are issues to consider with these new arrangements: (a) the separation of the network information system from direct governance by the

community generating the data; (b) the potential distancing of local data managers as contributors to the development of network data infrastructure which in turn impacts local data and design capacities; and (c) the impact on identity and leadership for LTER IMC from which emerged a number of EDI leaders. Continuing study of EDI as a national data management center and as an environmental sciences platform, including its data archive, would be informative. Will the LTER community assumptions continue to be transformed? What pivots and opportunities will influence the EDI trajectory as it becomes one of a growing number of national data offices?

6 | Conclusion

The LTER case illustrates a research-driven ecology program that fostered the development of site-based data management and local data systems prior to designing, launching, sustaining, and scaling up a centralized network information system. This case provides an approach to Open Data that has proven effective in its ability to implement data management in addition to supporting the evolution of data infrastructure. The three research questions posed earlier are discussed below with the LTER case in mind and are followed by final remarks.

6.1 | The How, What and When of Data Infrastructure

6.1.1 | How Does the Growth of Digital Data Infrastructure Occur Over Time?

The Incremental Growth Model operationalizes the notion of “infrastructuring,” creating an analytical tool offering specificity useful for analysis. Considering the LTER NIS as a multi-phase trajectory foregrounds the concept of infrastructure growth as an ongoing process from an initial vision through to continuing use. Drawing on Science and Technology Studies and Infrastructure Studies concepts of “continuing design” and “infrastructure growth,” this long-term ethnographic story of a community information system could be framed as a biography of data infrastructure (Pollock and Williams 2010). Instead, the Incremental Growth Model incorporates some empirical contributions from Organizational Change Research, including the concepts of contextual conditions, pivotal events, and major phase characteristics. By merging the concepts from these fields, the model integrates the variety of often taken-for-granted activities into a story that highlights emergent opportunities and adaptive responses to change in the trajectory of an information system.

Given the incremental growth of LTER NIS, time was available for the interactions that resulted in the building of personal relationships and trust. The pace of NIS development allowed time for understandings to grow and diffuse across the community, thereby shifting assumptions about digital data management. The start and end of each phase served as points of community discussion where experiential learning and shared decision-making occurred. Recognition of individual phases and their outcomes stimulated periodic reflections on successes among participants while reaffirming data-related work as a long-term mission that required “staying the course” with nudges for those within the community who were resisting or lagging in their data practices. Stepwise development also enabled data workers to grow into key actors in the digital environment. Indeed, a number of site-based data managers and LNO NIS developers who participated in earlier phases of NIS growth were active in the scaling up to EDI.

6.1.2 | What Characteristics and Data Workers Are Associated With the Evolution of an Information System?

Major characteristics summarized in Table 3 and discussed in Sections 4 and 5 describe the data work associated with each phase of the LTER NIS trajectory. Insights into infrastructure-making will increase as major characteristics and categories of infrastructure in a variety of contexts are identified in long-term studies. The many processes used and paths occurring in data infrastructure growth are leading to collections of empirical studies (Harvey, Jensen, and Morita 2017) and to calls for more comparative analysis (Henfridsson and Bygstad 2013; Miller 2018). In seeking explanations that take into account both the dynamics of digital environments and the various circumstances shaping their evolution, Henfridsson and Bygstad (2013) conclude “more knowledge about what drives digital infrastructures would be

highly valuable for managers and IT professionals confronted by the complexity of managing them.”

Many roles are emerging to carry out data work today. A notable aspect of the LTER NIS evolution is the development of a data workforce within the community. Data management is identified as a contextual condition of the LTER network (see Section 2.5.2). Embedded data management is described in the first phase of the LTER NIS trajectory and highlighted as one of six key findings in analysis of the NIS evolution (see Section 5.1.1). LTER data workers keep data management and local research needs in sync, thereby minimizing disruptions to research often caused by new technologies and data requirements. Although the phases in this case highlight a variety of actors, the role of data managers for LTER remains key as they serve as liaisons, designers, implementers, and sustainers of data-related work (Millerand and Baker 2010, 2020).

6.1.3 | When Does the Growth of Digital Data Infrastructure Begin?

The LTER information system trajectory is incremental and cumulative, making it difficult to pinpoint a single time at which infrastructure begins. Each pivot signals the beginning of change and demarks a new phase. Each phase sets the stage for subsequent phases. In the 1980s (Phase 1) when the internet was not widely available, the NSF funding agency mandate establishing data management operationalized a role that facilitated change in community data practices and in data system developments. When funding became available in 2008 Phase 5 for an LTER NIS-data team to revise the NIS Framework, the LTER community had matured to recognize the LTER NIS as sustainable infrastructure providing access to LTER data. And Phase 5 was key to scaling up in Phase 6 when the support for the LTER NIS shifted from direct reporting within the LTER to become EDI interacting with LTER as an independently funded data entity.

6.2 | Final Thoughts

The LTER case illustrates a way of working with data and data systems on a time scale that accommodates collaborative activities and the incremental growth of data infrastructure. The evolution of data-related arrangements was facilitated by the continuity afforded to the LTER program by long-term support from NSF. Researchers traditionally trained and experienced in using their own data were exposed, over their years with LTER, to the work associated with data sharing that led to Open Data. Study of the six phases of the LTER NIS trajectory has revealed the many ways participants contribute to the growth of data infrastructures. The Incremental Growth Model of data infrastructure captures the NIS trajectory by focusing on three distinct features: (1) contextual conditions, (2) sequential phases initiated by pivotal events, and (3) major characteristics for each phase.

Open Science depends upon Open Data, that is, high-quality data widely available for reuse. The LTER case provides specifics regarding data arrangements and struggles associated with data infrastructure growth. This longitudinal study brings

together insights into the incremental growth of a network information system together with the cultural growth of a research community steadfast in addressing changes in both ecological systems and digital systems. The LTER NIS trajectory culminated in a Data Management Office and a digital hub with a data repository. Open Data requires data repositories available to ingest, store, and package data for access, discovery, and reuse yet critical, long-term studies of how to develop data infrastructure are rare.

The LTER case is an example of successful data infrastructure growth. As experience with digital systems grows alongside understandings of how “long-term matters in collaborative development” (Karasti, Baker, and Millerand 2010), we can reflect on the four decades of continuing design leading to the opening of the EDI national data management office and data repository for the environmental sciences. Early programs did not consider sociotechnical development nor did they prioritize local data managers. The scaling up of LTER NIS to EDI was configured to ingest data not only from gateway systems but also from individual investigators. Further, unlike early design approaches, the LTER NIS illustrates continuing design as it contributes to shifts in community assumptions, draws on effective communication structures, and facilitates growth of a data workforce aware of sociotechnical complexity. EDI will continue to address sustainability challenges and new challenges that may benefit from further collaboration with social scientists, historians, and organizational change researchers as LTER enters its fifth decade.

In contrast to the technological focus of many studies, this LTER study highlights the importance of taking a close look at the variety of characteristics that change and define the phases of a data infrastructure trajectory. The NIS growth can be viewed as fragmented and inefficient especially by those whose experience in practice and/or in theory is derived from cases following largely linear, technical plans rather than from cases of complex data ecosystems with interdependent social, technical and institutional facets at multiple levels. We argue that the LTER data work fits within a broader definition of success than is typical for data infrastructure, one that strengthens a community's data literacy, data management, and its local data workforce by continuously attending to communication, collective experience, and change. Supported by development of local data capacities, an ecological community's operational network data infrastructure was transformed into a national data management office and repository for the ecological and environmental sciences. The LTER case of data infrastructure can be described as an organizational achievement attained through long-term incremental growth.

Author Contributions

Karen S. Baker: conceptualization (lead), formal analysis (lead), investigation (lead), methodology (lead), visualization (lead), writing – original draft (lead), writing – review and editing (equal). **Florence Millerand:** conceptualization (supporting), methodology (supporting), visualization (supporting), writing – review and editing (supporting).

Acknowledgements

We thank the LTER community members who provided time and insight into bringing together this story. We give special thanks to Corrina Gries, Nicole Kaplan, Mark Servilla, Charleen Johnson, and Jena Webb

for review of and suggestions on the draft manuscript. We also appreciate the insight and contributions of our reviewers and editors.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The authors have nothing to report.

References

- Aken, J. E. 2004. “Management Research on the Basis of the Design Paradigm: The Quest for Field-Tested and Grounded Technological Rules.” *Journal of Management Studies* 41, no. 2: 219–246.
- Arango-Vasquez, L., and M. Gentilin. 2021. “Organizational Couplings: A Literature Review.” *Innovar* 31, no. 79: 151–168.
- Aronova, E., K. S. Baker, and N. Oreskes. 2010. “Big Science and Big Data in Biology: From the International Geophysical Year Through the International Biological Program to the Long Term Ecological Research (LTER) Network, 1957–Present.” *Historical Studies in the Natural Sciences* 40, no. 2: 183–224.
- Atkins, D. 2003. “NSF-AP Report: Revolutionizing Science and Engineering Through Cyberinfrastructure.” <https://www.nsf.gov/od/oci/reports/atkins.pdf>.
- Baker, K. S. 2017. *Data Work Configurations in the Field-Based Natural Sciences: Mesoscale Infrastructures, Project Collectives, and Data Gateways*. PhD. Urbana, IL: School of Information Sciences, University of Illinois Urbana-Champaign.
- Baker, K. S., B. J. Benson, D. L. Henshaw, D. Blodgett, J. H. Porter, and S. G. Stafford. 2000. “Evolution of a Multisite Network Information System: The LTER Information Management Paradigm.” *Bioscience* 50, no. 11: 963–978.
- Baker, K. S., and C. L. Chandler. 2008. “Enabling Long-Term Oceanographic Research: Changing Data Practices, Information Management Strategies and Informatics.” *Deep Sea Research Part II: Topical Studies in Oceanography* 55, no. 18–19: 2132–2142. <https://doi.org/10.1016/j.dsr2.2008.05.009>.
- Baker, K. S., N. Kaplan, and E. Melendez-Colom. 2010. “IMC Governance Working Group: Developing a Terms of Reference.” *LTER Databits Newsletter*, Fall 2010. <https://lternet.edu/wp-content/uploads/2017/12/2010-fall-lter-databits.pdf>.
- Baker, K. S., and H. Karasti. 2004. *The Long-Term Information Management Trajectory: Working to Support Data, Science and Technology*. San Diego, IL: UC San Diego.
- Baker, K. S., and H. Karasti. 2018. “Data Care and Its Politics: Designing for a Neglected Thing.” Paper Presented at the Proceedings of the 15th Participatory Design Conference (PDC'18): Full Papers—Volume 1, 20–24 August 2018, Hasselt and Genk, Belgium. <https://dl.acm.org/citation.cfm?doid=3210586.3210587>.
- Baker, K. S., and M. Mayernik. 2020. “Disentangling Knowledge Production and Data Production.” *Ecosphere* 11, no. 7: 1–12. <https://doi.org/10.1002/ecs2.3191>.
- Baker, K. S., and F. Millerand. 2007. “Articulation Work Supporting Information Infrastructure Design: Coordination, Categorization, and Assessment in Practice.” Paper Presented at the Proceedings of the 40th Hawaii International Conference on System Sciences. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4076872>.
- Baker, K. S., and F. Millerand. 2010. “Infrastructuring Ecology: Challenges in Achieving Data Sharing.” In *Collaboration in the New Life Sciences*, edited by J. N. Parker, N. Vermeulen, and B. Penders, 111–138. Burlington, VT, USA: Ashgate.

- Baron, J. S., A. Specht, E. Garnier, et al. 2017. "Synthesis Centers as Critical Research Infrastructure." *Bioscience* 67, no. 8: 750–759. <https://doi.org/10.1093/biosci/bix053>.
- BCO-DMO. 2013. "Description of the Biological and Chemical Oceanography Data Management Office." WHOI Report MS #36.
- Beaujardière, J. D. L. 2016. "NOAA Environmental Data Management." *Journal of Map & Geography Libraries* 12, no. 1: 5–27.
- Benson, B. J., P. C. Hanson, J. W. Chipman, and C. J. Bowser. 2006. "Breaking the Data Barrier: Research Facilitation Through Information Management." In *Long-Term Dynamics of Lakes in the Landscape: Long-Term Ecological Research on North Temperate Lakes*, edited by J. J. Magnuson, T. K. Kratz, and B. J. Benson, 258–279. Oxford, UK: Oxford University Press.
- Berkley, C., M. B. Jones, J. Bojilova, and D. Higgins. 2001. "Metacat: A Schema-Independent XML Database System." In *Proceedings of the 13th Int. Conference of Scientific Statistical Database Management*. Washington, DC: IEEE Computer Society.
- Bødker, S., C. Dindler, and O. S. Iversen. 2017. "Tying Knots: Participatory Infrastructuring at Work." *Computer Supported Cooperative Work (CSCW)* 26, no. 1–2: 245–273.
- Borgman, C. L. 2015. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press.
- Borgman, C. L., A. Scharnhorst, and M. S. Golshan. 2019. "Digital Data Archives as Knowledge Infrastructures: Mediating Data Sharing and Reuse." *Journal of the Association for Information Science and Technology* 70, no. 8: 888–904.
- Bowker, G. C. 2017. "How Knowledge Infrastructures Learn." In *Infrastructures and Social Complexity: A Companion*, edited by P. Harvey, C. B. Jensen, and A. Morita, 391–403. London, UK: Routledge.
- Bowker, G. C., K. S. Baker, F. Millerand, and D. Ribes. 2010. "Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment." In *International Handbook of Internet Research*, edited by J. Hunsinger, L. Klastrup, and M. Allen, 97–117. Dordrecht, the Netherlands: Springer.
- Brunt, J. W. 1998. "The LTER Network Information System: A Framework for Ecological Information Management." Paper Presented at the North American Science Symposium: Toward a Unified Framework for Inventorying and Monitoring Forest Ecosystem Resources 2–6 Nov 1998, Guadalajara, Mexico.
- Callahan, J. T. 1984. "Long-Term Ecological Research." *Bioscience* 34, no. 6: 363–367.
- Chabbi, A., and H. W. Loesch, eds. 2017. *Terrestrial Ecosystem Research Infrastructures: Challenges, New Developments and Perspectives*. New York, NY: CRC Press Taylor & Francis Group.
- Chinn, H., and C. Bledsoe. 1997. "Internet Access to Ecological Information: The US LTER All-Site Bibliography Project." *Bioscience* 47: 50–57.
- Coleman, D. C. 2010. *Big Ecology: The Emergence of Ecosystem Science*. Oakland, CA: University of California Press.
- Collins, S. L. 2021. "Network Level Science, Social-Ecological Research and the LTER Planning Process, Chapter 14." In *The Challenges and Accomplishments of Long-Term Ecological Research: New Perspectives on the Past, Present, and Future of Ecological Science*, edited by R. B. Waide and S. E. Kingsland. New York, NY: Springer.
- Cox, A. M., and S. Corral. 2013. "Evolving Academic Library Specialties." *Journal of the American Society for Information Science and Technology* 64, no. 8: 1526–1542.
- Cutcher-Gershenfeld, J. 2018. "Assumptions Wrangling: An Experiment in Culture Change." *Heller Magazine* (Winter). <http://heller.brandeis.edu/news/items/releases/2018/winter-magazine-cutcher-gershenfeld.html>.
- Cutcher-Gershenfeld, J. 2020. "Pivotal Events Driving Organizational and Institutional Transformation." *Negotiation Journal* 36, no. 2: 127–139.
- Cutcher-Gershenfeld, J., D. Brooks, and M. Mulloy. 2015. *Inside the Ford-UAW Transformation, Pivotal Events in Valuing Work and Delivering Results*. Cambridge, MA: MIT Press.
- EarthCube Office and EarthCube Leadership Council. 2022. "EarthCube: A retrospective 2011–2022." In *EarthCube Organization Materials*. San Diego, CA: UC San Diego Library Digital Collections.
- Edwards, P. N. 2004. "A Vast Machine: Standards as Social Technology." *Science* 304, no. 5672: 827–828.
- Edwards, P. N., S. Jackson, G. Bowker, and C. P. Knobel. 2007. "Understanding Infrastructure: Dynamics, Tensions, and Design." Report of a Workshop: History and Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures, Ann Arbor, MI. <http://deepblue.lib.umich.edu/handle/2027.42/49353>.
- Edwards, P. N., S. J. Jackson, M. K. Chalmers, et al. 2013. "Knowledge Infrastructures: Intellectual Frameworks and Research Challenges." Ann Arbor, MI. <http://hdl.handle.net/2027.42/97552>.
- Edwards, P. N., M. S. Mayernik, A. L. Batcheller, G. C. Bowker, and C. L. Borgman. 2011. "Science Friction: Data, Metadata, and Collaboration." *Social Studies of Science* 41, no. 5: 667–690.
- Engeström, Y. 2001. "Expansive Learning at Work: Toward an Activity Theoretical Reconceptualization." *Journal of Education and Work* 14, no. 1: 133–156.
- Erickson, I., and S. Sawyer. 2019. "Infrastructuring as Bricolage: Thinking Like a Contemporary Knowledge Worker." In *Thinking Infrastructures (Research in the Sociology of Organizations)*, edited by M. Kornberger, G. Bowker, J. Elyachar, A. Mennicken, P. Miller, J. Nucho, and N. Pollock, vol. 62, 321–334. Cambridge, UK: Emerald Publishing Limited. <https://doi.org/10.1108/S0733-558X2019000062020>.
- EU. 2016. "Open Innovation, Open Science, Open to the World—A Vision for Europe." *European Commission, Brussels* 10: 1652. <https://doi.org/10.2777/061652>.
- Fegraus, E. H., S. Andelman, M. B. Jones, and M. Schildhauer. 2005. "Maximizing the Value of Ecological Data With Structured Metadata: An Introduction to Ecological Metadata Language (EML) and Principles for Metadata Creation." *Bulletin of the Ecological Society of America* 86: 158–168.
- Fischer, G., and T. Herrmann. 2011. "Socio-Technical Systems: A Meta-Design Perspective." *International Journal of Sociotechnology and Knowledge Development* 3: 1–33.
- Fischer, G., and T. Herrmann. 2015. "Meta-Design: Transforming and Enriching the Design and Use of Socio-Technical Systems." In *Designing Socially Embedded Technologies*, edited by V. Wulf, K. Schmidt, and D. Randall, 79–109. New York, NY: Springer.
- Fischer, G., and J. Ostwald. 2002. "Seeding, Evolutionary Growth, and Reseeding: Enriching Participatory Design with Informed Participation." In *Proceedings of the Participatory Design Conference (PDC'02)*, 135–143. Malmö, Sweden: Malmö University.
- Franklin, J. F., C. S. Bledsoe, and J. T. Callahan. 1990. "Contributions of the Long-Term Ecological Research Program." *Bioscience* 40, no. 7: 509–523.
- Gitelman, L., ed. 2013. *'Raw Data' Is an Oxymoron*. Cambridge, MA: MIT Press.
- Glover, D. M., C. L. Chandler, S. C. Doney, et al. 2006. "The US JGOFS Data Management Experience." *Deep-Sea Research Part II* 53: 793–802.
- Golley, F. B. 1993. *A History of the Ecosystem Concept in Ecology: More Than the Sum of the Parts*. New Haven, CT: Yale University Press.

- Gries, C., S. Beaulieu, R. F. Brown, et al. 2021. "Data Package Design for Special Cases, Version 1." <https://doi.org/10.6073/pasta/9d4c803578c3fbc45fc23f13124d052>. Environmental Data Initiative.
- Gries, C., M. R. Gahler, P. C. Hanson, T. K. Kratz, and E. H. Stanley. 2016. "Information Management at the North Temperate Lakes Long-Term Ecological Research Site—Successful Support of Research in a Large, Diverse, and Long Running Project." *Ecological Informatics* 36: 201–208.
- Gries, C., P. Hanson, M. C. O'Brien, M. Servilla, K. Vanderbilt, and R. Waide. 2023. "The Environmental Initiative: Connecting the Past to the Future Through Data Reuse." *Ecology and Evolution* 13: 1.
- Gunderson, L. H., and C. S. Holling, eds. 2002. *Panarchy, Understanding Transformations in Human and Natural Systems*. Washington, DC: Island Press.
- Hahn, C., A. S. Hoffman, S. C. Slota, S. Inman, and D. Ribes. 2018. "Entangled Inversions: Actor/Analyst Symmetry in the Ethnography of Infrastructure." *Interaction Design and Architecture(s) Journal* 38: 124–139.
- Hak, T., F. Jaspers, and J. Dul. 2013. "The Analysis of Temporally Ordered Configurations: Challenges and Solutions, Chapter 5." In *Configurational Theory and Methods in Organizational Research in the Sociology of Organizations*, 107–127. Cambridge, UK: Emerald Group Publishing.
- Hampton, S. E., S. S. Anderson, S. C. Bagby, et al. 2015. "The Tao of Open Science for Ecology." *Ecosphere* 6, no. 7: 1–13.
- Hanseth, O. 2010. "From Systems and Tools to Networks and Infrastructures—From Design to Cultivation: Towards a Design Theory of Information Infrastructures." In *Industrial Informatics Design, Use and Innovation*, edited by J. Holmstrom, 122–156. Harshey, PA: IGI Global.
- Harvey, P., C. B. Jensen, and A. Morita, eds. 2017. *Infrastructures and Social Complexity: A Companion*. London, UK: Routledge.
- Henderson, A., and M. Kyng. 1991. "There Is no Place Like Home: Continuing Design in Use." In *Design at Work: Cooperative Design of Computer Systems*, edited by J. Greenbaum and M. Kyng, 219–240. Hillsdale, NJ: Lawrence Erlbaum.
- Henfridsson, O., and B. Bygstad. 2013. "The Generative Mechanisms of Digital Infrastructure Evolution." *MIS Quarterly* 37, no. 3: 907–931.
- Henshaw, D. L., K. Baker, D. Blodgett, and J. H. Porter. 1998. "Climate Database Project: A Strategy for Improving Information Access Across Research Sites." In *Data and Information Management in the Ecological Sciences: A Resource Guide*, edited by W. K. Michener, J. H. Porter, and S. G. Stafford. Albuquerque, NM: LTER Network Office, University of New Mexico.
- Hepso, V., E. Monteiro, and K. H. Rolland. 2009. "Ecologies of e-Infrastructures." *Journal of the Association for Information Systems* 10, no. 5: 2–446.
- Hirsch, S. L., D. Ribes, and S. Inman. 2022. "Sedimentary Legacy and the Disturbing Recurrence of the Human in Long-Term Ecological Research." *Social Studies of Science* 52, no. 4: 561–580.
- Hobbie, J. E. 2003. "Scientific Accomplishment of the Long Term Ecological Research Program: An Introduction." *Bioscience* 53, no. 1: 17–20.
- Ingersoll, R. C., T. R. Seastedt, and M. Hartman. 1997. "A Model Information Management System for Ecological Research." *Bioscience* 47: 310–316.
- Johnson, J. C., R. R. Christian, J. W. Brunt, C. R. Hickman, and R. B. Waide. 2010. "Evolution of Collaboration Within the US Long Term Ecological Research Network." *Bioscience* 6, no. 11: 931–940.
- Jones, M. B., M. P. Schildhauer, O. J. Reichman, and S. Bowers. 2006. "The New Bioinformatics: Integrating Ecological Data From the Gene to the Biosphere." *Annual Review of Ecology, Evolution, and Systematics* 37: 519–544.
- Kaltenbrunner, W. 2017. "Digital Infrastructure for the Humanities in Europe and the US: Governing Scholarship Through Coordinated Tool Development." *Computer Supported Cooperative Work (CSCW)* 26, no. 3: 275–308.
- Kaplan, N. E., K. S. Baker, and H. Karasti. 2021. "Long Live the Data! Embedded Data Management at a Long-Term Ecological Research Site." *Ecosphere* 12, no. 5: 1–25. <https://doi.org/10.1002/ecs2.3493>.
- Karasti, H., and K. S. Baker. 2004. "Infrastructuring for the Long-Term: Ecological Information Management." <https://doi.org/10.1109/HICSS.2004.1265029>. Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 5–8 Jan 2004, Big Island, HI, USA, 1–10.
- Karasti, H., and K. S. Baker. 2008. "Digital Data Practices and the Long Term Ecological Research Program Growing Global." *International Journal of Digital Curation* 3, no. 2: 42–58.
- Karasti, H., K. S. Baker, and E. Halkola. 2006. "Enriching the Notion of Data Curation in e-Science: Data Managing and Information Infrastructuring in the Long Term Ecological Research (LTER) Network." *Computer Supported Cooperative Work (CSCW)* 15, no. 4: 321–358.
- Karasti, H., K. S. Baker, and F. Millerand. 2010. "Infrastructure Time: Long-Term Matters in Collaborative Development." *Computer Supported Cooperative Work (CSCW)* 19, no. 3–4: 377–415.
- Karasti, H., V. Pipek, and G. C. Bowker. 2018. "An Afterword to 'Infrastructuring and Collaborative Design'." *Computer Supported Cooperative Work (CSCW)* 27: 227–289. <https://doi.org/10.1007/s10606-017-9305-x>.
- Karasti, H., and A.-L. Syrjänen. 2004. "Artful Infrastructuring in Two Cases of Community PD." In *Paper Presented at the Proceedings of the Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices, Volume 1*, Toronto, Ontario, Canada, July 27 to 31, 2004, 20–30. New York, NY: ACM Press.
- Kitchen, R. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. Los Angeles, CA: Sage.
- Laney, C. M., D. P. C. Peters, and K. S. Baker. 2013. "History and Organization of the EcoTrends Project, Chapter 2." In *Long-Term Trends in Ecological Systems: A Basis for Understanding Responses to Global Change*, edited by D. P. C. Peters, C. M. Laney, A. E. Lugo, S. L. Collins, C. T. Driscoll, P. M. Goffman, J. M. Grove, A. K. Knapp, T. K. Kratz, M. D. Ohman, R. B. Waide, and J. Yao, 21–27. Washington, DC: United States Department of Agriculture, Agricultural Research Service.
- Leonelli, S. 2015. "What Counts as Scientific Data? A Relational Framework." *Philosophy of Science* 82, no. 5: 810–821.
- Leonelli, S. 2016. *Data-Centric Biology: A Philosophical Study*. Chicago, IL: University of Chicago Press.
- LTER. 2004. "EML Best Practices Guide for LTER Sites." https://lternet.edu/wp-content/uploads/2010/12/emlbestpractices_oct2004_final.pdf.
- LTER. 2007. "Decadal plan." https://lternet.edu/wp-content/uploads/2010/12/LTER_CI_Strategic_Plan_4.2.pdf.
- LTER. 2007. "Network Cyberinfrastructure Strategic Plan." https://lternet.edu/wp-content/uploads/2010/12/LTER_CI_Strategic_Plan.pdf.
- LTER Michaels-Powers Report. 2011. "Long-Term Ecological Research Program, A Report of the 30 Year Review Committee." <https://lternet.edu/wp-content/uploads/2011/12/biol2001.pdf>.
- LTER Risser Report. 1993. "Ten-Year Review of the National Science Foundation's Long-Term Ecological Research Program. Long-Term Ecological Research Office." <https://lternet.edu/wp-content/uploads/2013/04/ten-year-review-of-LTER.pdf>.

- Magnuson, J. 1990. "Long-Term Ecological Research and the Invisible Present." *Bioscience* 40, no. 7: 495–601.
- Mauz, I., T. Peltola, C. Granjou, S. Van Bommel, and A. Buijs. 2012. "How Scientific Visions Matter: Insights From Three Long-Term Socio-Ecological Research (LTSE) Platforms Under Construction in Europe." *Environmental Science & Policy* 19: 90–99.
- Mayernik, M. S. 2017. "Open Data: Accountability and Transparency." *Big Data & Society* 4, no. 2: 1–5.
- Mayernik, M. S., A. L. Batcheller, and C. L. Borgman. 2011. "How Institutional Factors Influence the Creation of Scientific Metadata." Paper Presented at the Proceedings of the 2011 iConference.
- Michener, W., D. Vieglaiss, T. Vision, J. Kunze, P. Cruse, and G. Janée. 2011b. "Dataone: Data Observation Network for Earth-Preserving Data and Enabling Innovation in the Biological and Environmental Sciences." *D-Lib Magazine* 17, no. 1: 3.
- Michener, W. K. 1986. *Research Data Management in the Ecological Sciences*. Columbia, SC: Belle W. Baruch Library in Marine Science, Report No. 16, University of South Carolina Press.
- Michener, W. K., J. W. Brunt, J. J. Helly, R. B. Kirchner, and S. G. Stafford. 1997. "Nongeospatial Metadata for the Ecological Sciences." *Ecological Applications* 7: 330–342.
- Michener, W. K., A. Miller, and R. Nottrott. 1990. *Long-Term Ecological Research Network Core Data Set Catalog*. Columbia, SC: University of South Carolina: Belle W. Baruch Institute for Marine Biology and Coastal Research, for The Long-Term Ecological Research Network.
- Michener, W. K., J. Porter, M. Servilla, and K. Vanderbilt. 2011a. "Long Term Ecological Research and Information Management." *Ecological Informatics* 6, no. 1: 13–24.
- Michener, W. K., J. H. Porter, and S. G. Stafford, eds. 1998. *Data and Information Management in the Ecological Sciences: A Resource Guide*. Albuquerque, NM: Long Term Ecological Research Network Office.
- Miller, D. 2018. "Challenging Trends in Configuration Research: Where Are the Configurations?" *Strategic Organization* 16, no. 4: 453–469.
- Millerand, F., K. Baker, B. Benson, and M. Jones. 2005. "Lessons Learned From EML About the Community Process of Standard Implementation." LTER Databits, Information Management Newsletter of the Long Term Ecological Research Network, Fall 2005. <http://databits.lternet.edu/fall-2005/lessons-learned-eml-about-community-process-standard-implementation>.
- Millerand, F., and K. S. Baker. 2010. "Who Are the Users? Who Are the Developers? Webs of Users and Developers in the Development Process of a Technical Standard." *Information Systems Journal* 20, no. 2: 137–161.
- Millerand, F., and K. S. Baker. 2020. "Data Infrastructures in Ecology: An Infrastructure Studies Perspective." In *Oxford Research Encyclopedia of Environmental Science*, edited by H. Shugart. Oxford, UK: Oxford University Press.
- Millerand, F., and G. C. Bowker. 2009. "Metadata Standards: Trajectories and Enactment in the Life of an Ontology." In *Standards and Their Stories. How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, edited by M. Lampland and S. L. Star, 149–165. Ithaca, NY: Cornell University Press.
- Millerand, F., D. Ribes, K. Baker, and G. C. Bowker. 2013. "Making an Issue out of a Standard: Storytelling Practices in a Scientific Community." *Science, Technology & Human Values* 38, no. 1: 7–43.
- Misangyi, V. F. 2016. "Institutional Complexity and the Meaning of Loose Coupling: Connecting Institutional Sayings and (Not) Doings." *Strategic Organization* 14, no. 4: 407–440.
- Mongili, A., and G. Pellegrino. 2014. *Information Infrastructure (s): Boundaries, Ecologies, Multiplicity*. Cambridge, UK: Cambridge Scholars Publishing.
- NAP. 2018. "Open Science by Design: Realizing a Vision for 21st Century Research." <https://www.nap.edu/catalog/25116/open-science-by-design-realizing-a-vision-for-21st-century>.
- NRC. 1997. *Bits of Power: Issues in Global Access to Scientific Data*. Washington, DC: NRC.
- NSF CI Report. 2007. "National Science Foundation, Cyberinfrastructure Cision for 21st Century Discovery." <https://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf>.
- NSF CI Report. 2019. "National Cyberinfrastructure Coordination Service Conference: Rethinking NSF's Computational Ecosystem for 21st Century Science and Engineering." <https://beta.nsf.gov/events/rethinking-nsfs-computational-ecosystem-21st>.
- NSF Major Facilities Guide. 2019. "U.S. National Science Foundation: Major Facilities Guide, Report NSF19-68." Congressional Research Service. <https://www.nsf.gov/pubs/2019/nsf19068/nsf19068.pdf>.
- NSTC. 2022. *Desirable Characteristics of Repositories for Managing and Sharing Data Resulting from Federally Funded Research*. Washington, DC: NSTC.
- OECD. 2015. *Making Open Science a Reality*. Paris, France: OECD Publishing.
- Orton, J. D., and K. E. Weick. 1990. "Loosely Coupled Systems: A Reconceptualization." *Academy of Management Review* 15, no. 2: 203–223.
- OSTP. 2013. *Increasing Access to the Results of Federally Funded Scientific Research*. Washington, DC: Retrieved from Office of Science and Technology Policy, John P. Holdren, Memorandum for the Heads of Executive Departments and Agencies, White House.
- OSTP. 2022. *Ensuring Free, Immediate, and Equitable Access to Federally Funded Research. Memorandum for the Heads of Executive Departments and Agencies, August 25, 2022*. Washington, DC: White House.
- Pahlka, J. 2023. *(Re)coding America: Why Government Is Failing in the Digital Age and How we Can Do Better*. New York, NY: Macmillan.
- Peters, D. P. C., C. M. Laney, A. E. Lugo, et al. 2013. "Long-Term Trends in Ecological Systems: A Basis for Understanding Responses to Global Change." <http://www.ars.usda.gov/is/np/LongTermTrends/LongTermTrendsIntro.htm>.
- Pipek, V., H. Karasti, and G. Bowker. 2017. "A Preface to 'Infrastructuring and Collaborative Design'." *Computer Supported Cooperative Work (CSCW)* 26: 2130.
- Pipek, V., H. Karasti, and G. C. Bowker. 2018. "A Preface to 'Infrastructuring and Collaborative Design'." *Journal of Computer Supported Cooperative Work (CSCW)* 26: 2–5. <https://doi.org/10.1007/s10606-017-9271-3>.
- Pipek, V., and V. Wulf. 2009. "Infrastructuring: Toward an Integrated Perspective on the Design and Use of Information Technology." *Journal of the Association for Information Systems* 10, no. 5: 1–473.
- Pollock, N., and R. Williams. 2009. *Software and Organizations: The Biography of the Enterprise-Wide System or How SAP Conquered the World*. New York, NY: Rutledge.
- Pollock, N., and R. Williams. 2010. "E-Infrastructures: How Do We Know and Understand Them? Strategic Ethnography and the Biography of Artefacts." *Computer Supported Cooperative Work (CSCW)* 19: 521–556.
- Pomerantz, J., and R. Peek. 2016. "Fifty Shades of Open." *First Monday* 2: 5.
- Porter, J. H. 2010. "A Brief History of Data Sharing in the US Long Term Ecological Research Network." *Bulletin of the Ecological Society of America* 91, no. 1: 14–20.
- Porter, J. H., D. L. Henshaw, and S. G. Stafford. 1997. "Research Metadata in Long-Term Ecological Research (LTER)." Paper Presented at the IEEE Metadata Conference, 1999 February 2, Silver Spring, Maryland.

- Reichman, O. J., M. B. Jones, and P. Schildhauer. 2011. "Challenges and Opportunities of Open Data in Ecology." *Science* 331: 703–705. <https://doi.org/10.1126/science.1197962>.
- Ribes, D. 2014. "The Kernel of a Research Infrastructure." Paper Presented at the Computer Supported Cooperative Work Conference, CSCW 14, February 15–19, 2014, Baltimore, MD.
- Ribes, D., and T. A. Finholt. 2007. "Tensions Across the Scales: Planning Infrastructure for the Long-Term." Paper Presented at the Proceedings of the 2007 International ACM Conference on Supporting Group Work. ACM.
- Ribes, D., and T. A. Finholt. 2009. "The Long Now of Infrastructure: Articulating Tensions in Development." *Journal of the Association for Information Systems* 10, no. 5: 375–398.
- Ribes, D., and J. B. Polk. 2014. "Flexibility Relative to What? Change to Research Infrastructure." *Journal of the Association for Information Systems* 15, no. 5: 287–305.
- Robbins, R. J. 2011. "Data Management for LTER 1980–2010, a Position Paper." https://lternet.edu/wp-content/uploads/2010/12/TheDecadalPlanReformattedForBook_with_citation.pdf.
- Salgado, L. A., L. Morel, and I. V  rilhac. 2018. "Towards a Better Understanding of the Concept of Design." *Projectics/Proyectica/Projectique* 20: 91–114.
- Servilla, M., and J. Brunt. 2011. "The LTER Network Information System: Improving Data Quality and Synthesis Through Community Collaboration." *AGU Fall Meeting Abstracts* 1: 1598.
- Servilla, M., J. Brunt, D. Costa, J. McGann, and R. Waide. 2016. "The Contribution and Reuse of LTER Data in the Provenance Aware Synthesis Tracking Architecture (PASTA) Data Repository." *Ecological Informatics* 36: 247–258. <https://doi.org/10.1016/j.ecoinf.2016.07.003>.
- Servilla, M., J. Brunt, I. San Gil, and D. Costa. 2006. "PASTA: A Network-Level Architecture Design for Automating the Creation of Synthetic Products in the LTER Network." <https://lternet.edu/stories/databits-newsletter-fall-2006/>.
- Servilla, M., D. Costa, C. Laney, I. San Gil, and J. Brunt. 2008. "The EcoTrends Web Portal: An Architecture for Data Discovery and Exploration." Paper presented at the Proceedings of the Environmental Information Management Conference.
- Servilla, M., M. O'Brien, and D. Costa. 2013. "Ensuring the Quality of Data Packages in the LTER Network Provenance Aware Synthesis Tracking Architecture Data Management System and Archive." Paper Presented at the AGU Fall Meeting Abstracts. <https://ui.adsabs.harvard.edu/abs/2013AGUFMIN53C1576S/abstract>.
- Simonsen, J., and T. Robertson. 2013. *Handbook of Participatory Design*. London, UK: Routledge.
- Stafford, S. 2021. "A Retrospective of Information Management in the Long Term Ecological Research Program, Chapter 13." In *The Challenges and Accomplishments of Long-Term Ecological Research: New Perspectives on the Past, Present, and Future of Ecological Science*, edited by R. B. Waide and S. E. Kingsland, 1–21. New York, NY: Springer.
- Star, S. L., and K. Ruhleder. 1996. "Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces." *Information Systems Research* 7, no. 1: 111–134.
- Steinhardt, S. B. 2016. "Breaking Down While Building Up: Design and Decline in Emerging Infrastructures." Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2198–2208.
- Steinhardt, S. B., and S. J. Jackson. 2014. "Reconciling Rhythms: Plans and Temporal Alignment in Collaborative Scientific Work." Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, 134–145.
- Swanson, F. J., and R. E. Sparks. 1990. "Long-Term Ecological Research and the Invisible Place." *Bioscience* 40, no. 7: 502–508.
- TRS. 2012. "Science as an Open Enterprise, Open Data For Open Science." Royal Society Science Policy Centre. <https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/>; http://royal.society.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf.
- UCSD. n.d. "University of California San Diego." <https://blink.ucsd.edu/HR/comp-class/classification/careertracks/index.html>.
- Uludag,   ., M. Kleehaus, C. Caprano, and F. Matthes. 2018. "Identifying and Structuring Challenges in Large-Scale Agile Development Based on a Structured Literature Review." Paper Presented at the 2018 IEEE 22nd International Enterprise Distributed Object Computing Conference (EDOC).
- Vision, T. 2010. "The Dryad Digital Repository: Published Evolutionary Data as Part of the Greater Data Ecosystem." *Nature Precedings* 20: 1–10.
- Waide, R. B., and S. E. Kingsland, eds. 2021. *The Challenges and Accomplishments of Long-Term Ecological Research: New Perspectives on the Past, Present, and Future of Ecological Science*. New York, NY: Springer.
- Winter, S., N. Berente, J. Howison, and B. Butler. 2014. "Beyond the Organizational 'Container': Conceptualizing 21st Century Sociotechnical Work." *Information and Organization* 24, no. 4: 250–269.