Summary and Keywords

The development of information infrastructures that make ecological research data available has increased in recent years, contributing to fundamental changes in ecological research. Science and Technology Studies (STS) and the subfield of Infrastructure Studies, which aims at informing infrastructures' design, use, and maintenance from a social science point of view, provide conceptual tools for understanding data infrastructures in ecology. This perspective moves away from the language of engineering, with its discourse on physical structures and systems, to use a lexicon more "social" than "technical" to understand data infrastructures in their informational, sociological, and historical dimensions. It takes a holistic approach that addresses not only the needs of ecological research but also the diversity and dynamics of data, data work, and data management. STS research, having focused for some time on studying scientific practices, digital devices, and information systems, is expanding to investigate new kinds of data infrastructures and their interdependencies across the data landscape. In ecology, data sharing and data infrastructures create new responsibilities that require scientists to engage in opportunities to plan, experiment, learn, and reshape data arrangements. STS and Infrastructure Studies scholars are suggesting that ecologists as well as data specialists and social scientists would benefit from active partnerships to ensure the growth of data infrastructures that effectively support scientific investigative processes in the digital era.

Keywords: environmental, data infrastructure, science and technology studies, infrastructure studies, infrastructuring, design, data work, data management

Introduction

The rise of digital technologies such as computers and the internet since the late 1950s has led to significant changes in all aspects of scientific research—from the research process and its scope to data and its management. These changes continue to impact science and pose significant challenges to the scientific community. Recent developments in collaboration, networking, and awareness of ecosystem complexity, while opening up opportunities for research, have evolved in conjunction with data issues that reside at the heart of contemporary science. There is a strong consensus that data produced by pub-

licly funded researchers should be considered a common good (e.g., Contreras & Reichman, 2015; National Research Council [NRC], 2004); and that public openness of data will contribute to the advancement of science. The exposure, circulation, and reuse of research data within and beyond a scientific community brings the promise of new discoveries and knowledge, as well as new forms of scientific collaboration (Olson & Olson, 2014; Parker, Vermeulen, & Penders, 2010; Sonnenwald, 2007).

"Information infrastructures," "cyberinfrastructures," "e-Research," and "e-Science" are some of the terms used in relation to digital infrastructures. Many such terms come from funding programs or science policies, such as the U.S. National Science Foundation (NSF) Office of Cyberinfrastructure established in the mid-2000s (Atkins et al., 2003) and European efforts focusing on research infrastructures (Wood, Andersson, Bachem, & Best, 2010). They refer to a wide array of service-oriented entities, such as local data gateways or grid computing. Bowker, Baker, Millerand, and Ribes (2010, p. 98) report that "information infrastructure refers loosely to digital facilities and services usually associated with the internet: computational services, help desks, and data repositories to name a few." This article focuses on data infrastructures as one type of information infrastructure that supports science by facilitating individual and collective work with data. Data infrastructures can be considered as a web of interconnected systems of subsystems, services, and arrangements that supports data work in the digital realm. A key service for researchers as well as a field of practice for computer scientists, technologists, and data professionals, infrastructures are also a research object for the social sciences (Frischmann, Madison, & Strandbury, 2014; Harvey, Bruun Jensen, & Morita, 2017; Jankowski, 2007; Ribes & Lee, 2010), especially the interdisciplinary field of Science and Technology Studies (STS), which examines the transformative power of science and technology to affect contemporary societies, and vice versa (Felt, Fouché, Miller, & Smith-Doerr, 2016; Vertesi & Ribes, 2019).

Infrastructure Studies (Bowker et al., 2010; Edwards, Bowker, Jackson, & Williams, 2009), a subfield of research in STS, has emerged to study the sociotechnical dynamics of information infrastructures and their impact on knowledge production. Moreover, recent work in this subfield suggests considering information infrastructures as "knowledge infrastructures" to focus on the changes in research and knowledge production associated with their development (Bowker, 2017; Edwards, 2010; Edwards et al., 2013; Leonelli, 2013). These works see knowledge infrastructures as "robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds" (Edwards, 2010, p. 17). With this framing, information and data infrastructures include more than digital technologies as they are also constituted of individuals, organizations, routines, shared norms, and practices (Bietz, Baumer, &Lee, 2010; Bietz, Ferro, & Lee, 2012; Edwards et al., 2013; Karasti et al., 2016). This represents a conceptualization of data infrastructure as continually realigning the relationships among people, technologies, and organizations; it is less about maintaining any particular technology than it is about being prepared to accommodate technological, scientific, and organizational change. Due to the nature of scientific inquiry, this openness to

change in relation to data infrastructure is particularly critical given that scientific investigations involve constant checking, feedback, updates, adaptation, and innovation.

The field of ecological research has experienced important epistemic changes since the 1970s, partly—although not exclusively—due to an ongoing technological "push." At the opening of the 21st century, however, with concerns about the sustainability of life on earth, some researchers who traditionally focused on local or regional scales were encouraged to become attuned to larger-scale thinking formulated in various disciplines including the environmental sciences. Additional dramatic changes in research have developed to include increasing interdisciplinarity, collaborative undertakings, and technological capabilities. U.S. national research programs have expanded from disciplinary "grand challenges," such as "biogeochemical cycles" in the environmental sciences (NRC, 2001) to boundary-spanning "big ideas" a decade later, a time when new interdisciplinary research fields such as the Data Sciences and Critical Zone Research emerged. It is significant that the U.S. NSF investment in "big ideas" includes "harnessing *data* for 21st century science and engineering" and "shaping the human-technology frontier" (Mervis, 2016, p. 755; NSF, 2016). The rapid development of data infrastructures has contributed to the profound changes that impact the practices and expectations of ecological science.

This article presents a review of seminal social science research contributions on the topic of data infrastructures, informed by a data management perspective. It draws mainly from the STS subfield of Infrastructure Studies that has shaped current understandings on the topic of data infrastructure in ecology. Our goal is to take stock of existing research, and to present some of the salient research themes and results as well as directions for future research. The article begins with a brief historical account of major turning points in the field of ecological research by linking them to significant developments in information infrastructure. The section ends with a synthesis of specific needs and challenges for data management and data infrastructures for contemporary ecological research. The article then moves to major contributions of Infrastructure Studies, summarizing six key research themes, and concludes with a discussion of four opportunities for future research.

Ecological Research and Information Infrastructure Developments

Changes in Ecology: Expanding the Scope of Ecological Research

The history of ecology is marked by an expansion in the scope of ecological research. Bocking (2010) presents the history of the concept of ecosystem science as a new approach in 1935 to the study of ecology, considering the flow of energy and nutrients in natural systems as complementary to the earlier studies of population and communities. It was followed by the development of an ethics-oriented notion of land stewardship. Further expansion of ecological insight occurred with the recognition of "coupled systems," particularly human and natural systems (e.g., Liu et al., 2007). New funding opportunities

were developed to support collaborative research including multi-institutional research, centers of excellence, and research coordination networks. Big ecology became evident with global-scale programs such as the International Biological Program (IBP), a program running from 1964–1974 that coordinated internationally (Aronova, Baker, & Oreskes, 2010; Kwa, 1987) as well as through development of ecological modeling (Jopp, Reuter, & Breckling, 2011; Shugart, 1976) and satellite remote sensing of the environment (Kwa, 2005; Kwa & Rector, 2010). The IBP was followed by various programs with new approaches that emphasized interdisciplinary, large-scale, and long-term perspectives in addition to collaboration and technology. In the United States these included the Long-Term Ecological Research (LTER) program, a site-based network made up of member sites established in 1980 (Hobbie, 2003; Waide & Thomas, 2013), and the Critical Zone Observatories, established in 2007 as laboratories focusing on the processes shaping the earth's surface (Anderson, Bales, & Duffy, 2008). In addition, the National Ecological Observatory Network started construction on a continental-scale 30-year ecological observatory platform in 2011, which focused on a network of instrumented field towers (Loescher, Kelly, & Lea, 2017; Zimmerman & Nardi, 2010). The new approaches prompted the development of new kinds of data practices and digital arrangements, specifically infrastructures for data.

Development of New Data Practices and Data Infrastructures

Ecological research has specific needs in terms of working with data. Synthesis of diverse data and knowledge is at the heart of ecology—an integrative discipline poised to address environmental challenges (Carpenter et al., 2009). Ecological research activities incorporate analysis and integration of many kinds of disciplinary data across space and time. Ecology is a field that has expanded from a small science focus to include big science. Today, ecological researchers study multicomponent, dynamic systems with associated digital data environments that are also dynamic and subject to nonlinearities and feedbacks (Harvey et al., 2017; Parsons et al., 2011). With increasing digital data and data capabilities, the concept of data infrastructure emerged.[1]

Digital data activities associated with short-term funded projects and instrument use expanded in the latter part of the 20th century to include interdisciplinary collaborations, data-rich digital collections, and platform design including not only satellites but also in-situ sensors. The increase in data and growth of data work (i.e., any efforts applied to data) has led to the development of new data practices and new kinds of infrastructures for assembling, processing, and managing data collectively. In ecology, data practices and policies have recently focused on data sharing (National Science and Technology Council, 2009). Data infrastructures providing access to shared data include research-centered local repositories, resource-centered synthetic centers, and large-scale archives that aggregate collections (Baker & Yarmey, 2009). Their scopes and structures differ: Large-scale data initiatives support access to highly structured data, whereas local scale efforts vary in size and accommodate a wider range of data types, methods, and analyses (Baker & Millerand, 2010).

Bocking (2010) provides a history of collaboration in ecology that examines motivations and the social relations associated with ecologists. It was the 1970s when geographic information systems using digital field data and Landsat satellite remote sensing data became instrumental in ecological collaboration. Personal computers were available widely in the 1980s followed by use of spreadsheets and small-scale databases in the 1990s, which allowed researchers to generate and analyze their data digitally in new ways. At the same time, field instrumentation was rapidly developing, becoming more available and beginning to provide data in digital form. Collaboration increased in the 1980s as scales and interdisciplinarity of fieldwork expanded. Large projects might hold a data workshop where researchers gathered in retreat-like setting to present and integrate their data over periods of days or weeks. Large-scale field programs exposed many researchers to large-scale collaborative investigations that were often interdisciplinary.

Availability of enabling technologies both to individual laboratories and computationally intensive centers created a continuum of digital environments that range from support for local data systems (e.g., a project-specific data repository) to larger-scale data facilities (e.g., a domain-wide data repository).[2] Ecological scientists often work with small-scale databases using a variety of instruments and sampling methods in generating data as well as a variety of analytic methods as they work with primary data (e.g., Leonelli, 2007; Michener & Brunt, 2000). Researchers in smaller research efforts were found to "lack the tools, infrastructure, and expertise to manage the growing amounts of data generated" (Zimmerman, Bos, Olson, & Olson, 2009, p. 222). Ecological science is part of a broad, ongoing transformation in science referred to as Open Science. "Open Science" encompasses many schools of thought aimed at increasing replicability, reducing bias, and increasing the impact of science by increasing access to data (Fecher & Friesike, 2014). It is a stimulus for change as well as a subject of debate in the digital era of earth and environmental sciences (Gewin, 2016; Hampton et al., 2015; Tenopir et al., 2015). National and international reports describe Open Science as a data access challenge, an economic force, a transformational opportunity, and a key to innovation (European Union, 2016; National Academies of Sciences, Engineering, and Medicine, 2018; Organisation for Economic Co-operation and Development [OECD], 2015; Royal Society, 2012).

A turn to long-term research in ecology brought attention to the temporal dimension of data issues including time-series data and long-term data management. In ecology, long-term research prioritized time-series data and data integration over a variety of time scales stretching from a short-term focus of hours, days, and months to longer periods of years, decades, centuries, and millennia (Likens, 1987; Magnuson, 1989, 1990). For example, the U.S. LTER Network focused on periods that varied in length from days to hundreds of years, thereby supporting long-term research that extended beyond most ecology studies of the time but not such long periods as the millennia of paleoecology. Managing these time-series data foregrounded data management and lifted the temporal horizon of data managers from the lifetime trajectories of individual researchers and laboratories to the notion of preservation of data and its use over longer periods. Social studies of science document data practices and scientific endeavors (e.g., Cragin & Shankar, 2006; Hine, 2006; Jackson, Ribes, Buyuktur, & Bowker, 2011; Zimmerman et al., 2009), in-

cluding tensions introduced by plans that must accommodate both short-term, knowledge-driven projects and longer-term, preservation-oriented data work (Baker & Millerand, 2010; Bowker et al., 2010; Karasti & Baker, 2004).

The U.S. mandate for data sharing (Holdren, 2013) and the European Research Infrastructure Consortium (2009) are two milestones that launched widespread discussions of data management and data infrastructure and spurred changes in data practices. Recognizing and supporting data as a first-class research object that can be cited as a research product alongside research publications (Kratz & Strasser, 2014), prompted new practices. Persistent identifiers, such as digital object identifiers (DOIs), play an important role in this recognition process.[3] DOIs document data-related entities including instruments, platforms, people, and constructed digital objects called "artifacts." Research funders and organizations have responded to this call for data sharing by focusing on data management planning. New institutional efforts are developing not only within existing organizations but also across traditional organizational and domain boundaries to support research data as a public good as well as to articulate data principles (Allison & Gurney, 2015; Mons et al., 2017). In sharing research data and tending to data infrastructure (Parker et al., 2010; Ribes & Finholt, 2009), new forums have developed to consider the issues involved (Inter-university Consortium for Political and Social Research, 2013; NRC, 2014; OECD, 2017B), including funding, sustainability, and data arrangements (Maron & Loy, 2011; OECD, 2017A).

Current Challenges with Data Management and Data Infrastructures

Data management and data infrastructures present specific challenges in ecology (Hampton et al., 2013; Porter, 2017). With only a fraction of ecological data readily discoverable and accessible (Reichman, Jones, & Schildhauer, 2011), direct access to data remains the exception rather than the norm in ecology. Supporting ecological data access involves both technical and social challenges. As a field-intensive, data-rich research domain, ecology works with heterogeneous data. Technical challenges involve data issues such as harmonization (combining data from various sources that use different formats and naming conventions into a coordinated data set) and provenance (historical record of the origins and what happens to the data thereafter). Ecological data sets are often dispersed across thousands of researchers even with large, ongoing data collection efforts such as the Group on Earth Observation Biodiversity Observation Network (Scholes et al., 2008), Knowledge Network for Biocomplexity (Andelman, Bowles, Willig, & Waide, 2004), and federating initiatives such as the Data Observation Network for Earth (Michener, Vieglais et al., 2011). Ecological data remain highly diverse given the wide range of topics studied, methods used, and production contexts (Birnholtz & Bietz, 2003; Vertesi, 2014). Data provenance is fundamental to data synthesis, but remains difficult to trace particularly in light of the loss of information that occurs over time (Michener, 2000). In ecology, definition of biological units is a central concern. Further, the quality of data is intimately linked to local knowledge and data collection details that strongly constrain standardiza-

tion efforts, such as with the Ecological Metadata Language, and thereby the circulation of reusable data sets (Karasti et al., 2010; Millerand & Bowker, 2009; Zimmerman, 2008).

As noted by Reichman et al. (2011), technological issues may be challenging, but social challenges are even more daunting. Disciplines have different histories and specific data-sharing practices (Birnholtz & Bietz, 2003). The view that access to primary data collected by oneself is fundamental was initially an integral part of the training of the ecological scientist (Hampton et al., 2017; Zimmerman, 2008), but a growing number of ecologists are in general agreement about the new research opportunities offered by open data.[4] Open data practices enable the reuse of data. The reuse of data, however, requires research, development, and new policies. On one hand, the reward system linked to the production of data sets is not yet fully in place (National Academy of Sciences, National Academy of Engineering, & Institute of Medicine, 2009; Pryor & Donnelly, 2009) for ecology or for other disciplines. On the other hand, data infrastructures are developing that involve a complex coordination of often layered and intersecting efforts at a variety of levels, such as project, program, community, domain, state, national, and international levels.

The new work of data management, description, assembly, packaging, and curation struggles with embedded assumptions within the culture of science as well as in work with data (Cutcher-Gershenfeld, 2018). STS in general and Infrastructure Studies in particular contribute to the identification and explication of ubiquitous but limited or outmoded beliefs, including the following:

- Data are gathered, managed, and analyzed in a single laboratory.
- Work with data involves one set of well-established procedures.
- Data are moved in a linear path from their origin and delivered to a single destination.
- Data infrastructure is a single overarching entity to which researchers will adapt.

As the ecological sciences mature, there is wider understanding of the complexity and the interdependence of the elements of an ecosystem. There is greater recognition of the coupling of human and natural ecosystems with investigations expanding from "how to understand an ecological biome" to "how to understand a socio-ecological biome" from which emerges the question of "how to support a research community's data needs inclusive of research computing, data management, and data infrastructure."

Data infrastructure is an important research object in STS and Infrastructure Studies. The following section provides a brief overview of recent studies of research infrastructure and major findings from the studies.

Infrastructure Studies Research, Contributions to Infrastructure Making

An STS perspective focuses on how infrastructures are affected by society, politics, and culture and vice versa. An STS project may research the social and organizational ramifications of a technical decision in an infrastructure design process or the change in an organization's work practices after implementation of a new infrastructure. Infrastructure Studies may investigate how an infrastructure comes into being, evolves across time, lasts, or dies. This research perspective provides a theoretical understanding of infrastructure that aims to inform its design, use, and maintenance (Bowker et al., 2010). Such a view turns away from the language of engineering in talking about structure and systems. Instead, use of an informational-sociological-historical lens reframes discussion through use of a social rather than technical lexicon.

A variety of research fields in the social sciences contribute to our understanding of infrastructure. Historically associated with Library Science, Information Science takes information as its main object of study, thus focusing on information analysis, classification, retrieval, and so on, as well as on information systems. Computer-Supported Cooperative Work is an interdisciplinary design-oriented research field interested in the use of technology to support people in their work, with a strong interest in collaborative and cooperative work, and the Participatory Design field provides specific design methodologies that aim to actively involve stakeholders in design processes. In addition, Management Sciences and Policy Studies are the sources of digital age thinking on institutional change relating to the growing information workforce and the accelerating pace of technology in the workplace. Each of these fields has produced important works of interest to STS researchers investigating data and information infrastructures. STS historical and sociological approaches of sociotechnical systems are in turn incorporated into these fields. As an interdisciplinary research field, Infrastructure Studies is unique in that it brings a vast body of social science work to the study of infrastructure.

The following sections present some of the major research contributions of Infrastructure Studies, organized around six key research themes: definition and characteristics of infrastructures, invisibility and articulation work, metaphors of design, interdependence of data arrangements, collective data management and models, and infrastructure configurations.

Definition and Characteristics of Infrastructures

In a seminal study of one of the early scientific community's digital networking efforts, Star and Ruhleder (1996) provided a now classic definition of infrastructure and a list of its key characteristics. An initial view of infrastructure presents the concept as "something upon which something else 'runs' or 'operates,'" and as "something that is built and maintained, and which then sinks into an invisible background," such as electric power grids and the internet (Star & Ruhleder, 1996, p. 112). According to these authors, how-

ever, the initial view is not only inaccurate but also is of little use to understand what infrastructures really are and how they work. First, infrastructures are inherently social *and* technical, and should be considered as sociotechnical entities. Second, they constitute relevant research objects in themselves for understanding the relationship between work and technology (Star & Ruhleder, 1996), for example, to examine the uses of a new infrastructure in relation to new patterns of collaboration that are emerging in a research domain.

Infrastructure Studies theories emphasize the socially constructed aspect of infrastructure. An infrastructure such as an ecological database is, of course, a physical entity (e.g., sets of digital data sets hosted on servers), but it also implies a set of actors and institutions that contribute to its development, implementation, maintenance, and use (e.g., programmers, ecologists, research institutions), as well as socio-institutional arrangements (e.g., data standards, usage agreements). So, in order to design an infrastructure, one must take the technology, the actors and stakeholders, and the relationships and contexts into account (Star & Bowker, 2002). This theoretical perspective provides an understanding of infrastructure as a contextualized "relation" rather than a "thing." What is recognized as infrastructure differs. For a database developer, a research lab's local data system connected to an ecological data portal is a target object, whereas it may be just one element of support in an ecological scientist's research program. An infrastructure becomes one in practice, for someone, and in relation to some organized practices in specific contexts (Star & Ruhleder, 1996). This conceptualization is now widely recognized in the field of Infrastructure Studies (Karasti et al., 2016), although key characteristics continue to be examined and explored, enriching the language and stimulating the thinking of those working with data and infrastructure (Karasti & Blomberg, 2018; Mongili & Giuseppina, 2014).

The following have been identified as key characteristics of infrastructure (Star & Ruhleder, 1996, p. 113): (1) embeddedness in other sociotechnical structures (any infrastructure is laid upon another and interacts with others); (2) transparency (it is transparent to use and it supports tasks invisibly); (3) spatial and temporal reach or scope (it reaches beyond a single event or one-site practice); (4) learned as part of membership in a community (new members need to familiarize with it so that it becomes taken for granted); (5) shaped by conventions of practice (while also contributing to the shaping of these conventions); (6) plugged into other infrastructures and tools in a standardized fashion (maybe after being adjusted to possibly conflicting local conventions); (7) based on an installed base (infrastructures do not grow de novo but wrestle with the inertia of an installed base and inherit strengths and limitations from that base); and (8) normally invisible infrastructures become visible upon breakdown (a server failure, for instance).
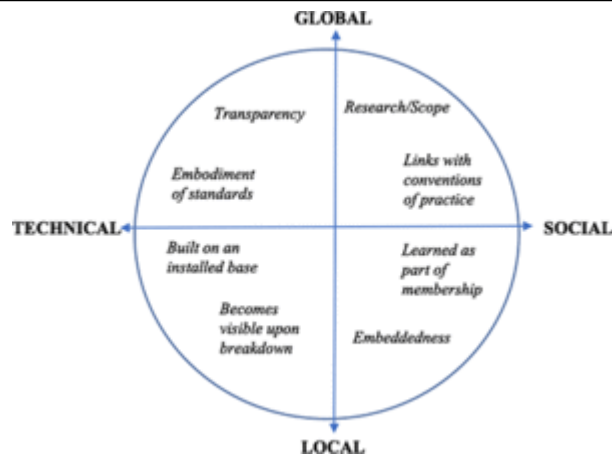
Figure 1. Eight key characteristics of infrastructure distributed across technical-social and local-global axes. The distribution of characteristics will change depending on the configuration or activity to be described.

Source: Bowker et al., 2010.

These infrastructural characteristics can be ordered along two axes, one technical/social, and the other one local/global, where infrastructure building is envisioned as a set of distributed activities along these axes (Figure 1). As Bowker et al. (2010, p. 102) explain: "the question is whether we choose, for any given problem, a primarily social or a technical solution, or some combination." For instance, metadata generation for an ecological data set can be an automated process, the responsibility of a local data manager, or an activity delegated to a national data archive. These choices imply activities that are distributed differently depending on whether the metadata production and expertise is expected to remain local or to reside at the domain level. Each infrastructure can therefore be understood as a configuration resulting from a different set of choices, the product of decisions that are not always thought through or whose implications are not always easy to anticipate.

Invisibility and Articulation Work

Invisibility is central to the infrastructural quality of a system (Star & Ruhleder, 1996), where infra means *below* in Latin. In Infrastructure Studies, invisibility may refer to one of three things: the invisible nature of the infrastructures themselves, the invisible work (of developing or maintaining it for instance), and the processes of making visible or invisible, deliberately or not, some work activities (Karasti et al., 2016).

A "good" infrastructure is conspicuous by its ability to blend in with its surroundings, to become transparent, invisible. As Bowker (2014) likes to remind us, we do not think about the road when we drive our cars, except when there are potholes! In the same way, we often do not think about the database or the network when we do our research, except when access fails. Infrastructures become visible when they are incomplete, break down, or malfunction. The same goes for some work tasks, often referred to as invisible

work. A good example of invisible work is that of the person who edited this text to remove spelling and typographical errors. The irony is that his or her work will not be noticed until a typo appears in the text. This is the daily lot of people who may be in charge of the maintenance of systems and infrastructures or of creating metadata for data sets. These are the invisible workers of the everyday.

Invisible work often evokes a group of understudied activities, such as informal tasks, but it may specify a marginalized category of actors generally at the bottom of social and professional hierarchies such as data processors, also referred to as data cleaners (Plantin, 2018). It could also involve highly specialized or abstract work activities requiring the management of information or knowledge, such as the work of research data management in a laboratory (Millerand, 2012). As Star and Strauss (1999, p. 9) have shown in their studies of disease and hospitals: "no work is inherently visible or invisible; we 'see' work through a selection of indicators." Making work activities visible or invisible may be the subject of intense negotiations; making an activity visible may lead to social recognition of the work and the worker, but it may also result in the reification of work and open the way to new forms of control (Star & Bowker, 1999; Suchman, 1995).

With the work of infrastructure and its maintenance often carried out by undervalued or invisible workers, researchers have suggested the notion of human infrastructure (Denis & Pontille, 2012; Lee, Dourish, & Mark, 2006; Shapin, 1989; Star, 1991). Mayernik, Wallis, and Borgman (2013) refer to the human-in-the-loop to spotlight the human side of infrastructural projects, thus raising awareness of the distributed collective processes that require skill and attention within research communities. In a study based on a long-term ecological research community, Millerand (2012) has shown that, in the constitution of a network-wide database, different processes work to make information managers and their work visible. Visibility results when new roles and responsibilities are acknowledged; invisibility results when a large part of the work ensures production of a routine, expected service. More recently, Plantin (2018) has investigated the work of data processors in a large social science data archive. He showed that a combined configuration of visibility and invisibility defined the social value within the archive of these workers preparing data for sharing while leaving no traces of their own work on the data set for those outside the archive.

A large part of infrastructure development and maintenance falls under the broad categories of articulation work and articulation process, that is, all the activities that bind together a group or series of tasks, actors, or environments (Baker & Millerand, 2007; Millerand, 2012). The articulating of work is what makes it possible to accomplish it (Strauss, 1988). Concretely, it essentially rests on "(largely invisible) activities of planning, organization, evaluation, mediation and task coordination which enable the 'articulation' of elements by a larger group (tasks, individuals, technologies, environments, etc.)" (Millerand, 2012, p. 168). Articulation work often depends on recognized mediators or liaisons to negotiate, coordinate, and manage collective planning and activities. Communication can play a central role in facilitating or constraining interactions. The medi-

um, the format, or the chosen language can be more or less adapted to the situations at hand and especially to the nature of the relations between the persons involved.

Methods for making the articulation work visible have been developed in STS and in Infrastructure Studies. They include observing infrastructures during moments of breakdown (Star, 1999) or as they are being designed, used, or repaired (Jackson, 2014; Karasti et al., 2010; Star & Bowker, 2002). These moments make visible elements that are otherwise hard to uncover (Karasti et al., 2016). Another method consists in doing an "infrastructural inversion" (Bowker, 1994), that is, a figure/ground reversal that consists in going backstage (Star, 1999) and bringing to the forefront what has traditionally been taken for granted. And finally, the notion of creating and maintaining a category of "inclusion" in discussions of participants, designers, workers, and stakeholders of all kinds is a method that brings awareness of diversity and provides a prompt to recognize those whose work is easily forgotten in an information society.

Metaphors of Design

Developing data infrastructures involves important design issues that occur in a variety of circumstances. STS work carried out since the 1980s by historians, sociologists, and information scientists on infrastructures with the aim of understanding how infrastructures form, evolve, work and fail, has brought new metaphors and concepts for thinking about the design of infrastructures beyond just the technical issues (Edward, Jackson, Bowker, & Knobel, 2007). Words matter, and metaphors are powerful rhetorical resources that help define visions of the world (Hamilton, 2000), including design issues and ways of approaching infrastructure. Three metaphors relating to imagining data and infrastructure making are "growing" infrastructure (rather than "building" it), the process of "infrastructuring" (rather than a completed project), and a "data ecosystem" (rather than a data pipeline process).

Drawing from empirical research on infrastructure development, largely in the form of case studies, STS scholars have argued against the idea that infrastructures could be built in the sense of "deliberately designed and constructed to a plan" (Edwards et al., 2009, p. 369), in "a highly conscious, carefully controlled, and fully directed sort of way" (Jackson, Edwards, Bowker, & Knobel, 2007, p. 4). Jackson et al. (2007) found that, "a careful and historically-informed study of infrastructural dynamics and tensions" weighs against this common vision. Together with other scholars, they promote instead a metaphor of "growing" an infrastructure, "to capture the sense of an organic unfolding within an existing (and changing) environment" (Edwards et al., 2009, p. 369). Infrastructures are complex systems that can be better understood, as they say, as emergent phenomena whose properties appear as they develop—sometimes, if not always—in unexpected ways. Most projects fall into disuse at the prototype phase, while the successful ones appear at the end, less "built" than "nurtured" or "grown," having undergone constant adaptations (Edwards et al., 2009, p. 369).

Not only is there no recipe to guarantee, for instance, use of large-scale infrastructure for data sharing (Cutcher-Gershenfeld et al., 2016), but designers often work with a limited conception of tasks, requirements, and expected uses (Edwards et al., 2009). STS scholars have suggested the notion of "infrastructuring" as an alternative to a build-and-serve model for infrastructure (Star & Bowker, 2002; Star & Ruhleder, 1996). Infrastructuring has been used to refer to "a relational and processual (in-the-making) perspective and/or design-oriented interest towards Information Infrastructures," that is, as a "longitudinal, unfolding process" (Pipek, Karasti, & Bowker, 2017, pp. 1–2). Karasti and Blomberg (2018) describe the process of infrastructuring as having the characteristics of relational, connected, and invisible, highlighting the process as emerging and accruing as well as having associated intentions and interventions. In turning the word *infrastructure* into a transitive verb, the aim is to account for the complexity of infrastructure design and development, and also to emphasize the idea of an active, ongoing process.

Thinking about the design of infrastructure as a process rather than a task emphasizes its dynamic and unpredictable growth: "Infrastructures subtend complex ecologies: their design process should always be tentative, flexible and open" (Star & Bowker, 2002, p. 160). Research studies have used the notion of "infrastructuring" in many ways: Describing the work of ecological information management and environmental monitoring (Karasti & Baker, 2004; Karasti et al., 2006; Parmiggiani, 2015), supporting new kinds of data production and data curation (Baker & Millerand, 2010), enabling redesign and social innovation over extended periods of time (Bjorgvinsson, Ehn, & Hillgren, 2012; Karasti et al., 2010), encompassing many aspects of design activities (Pendleton-Julian and Brown, 2018; Pipek et al., 2017), and in conjunction with a participatory design approach (Karasti, 2014; Karasti & Blomberg, 2018; Simonsen & Robertson, 2013). Design thinking frequently opens up multidirectional interactions that support shared decision making, narratives from multiple perspectives, and negotiated digital arrangements that broaden the understanding of constraints and rationales for choices to be made. Various kinds of learning have been identified as significant in digital design activities, including incremental, continuing, mutual, and community learning.

Finally, with growing interest in making data discoverable, open, linked, useful, and safe, Parsons et al. (2011) suggest the value of conceptualizing a "data ecosystem." The idea of a data ecosystem captures the nonlinear dynamics of data arrangements (Parsons et al., 2011; Pollock, 2011). This metaphor of the data ecosystem is found in practice in data management. Like environmental ecosystems, a data ecosystem is made up of interdependent elements in a system of subsystems. The relations of subsystems within and across systems foregrounds the need to consider feedbacks and nonlinearities in an ecosystem, whether it is an environmental system or a data system.

Interdependence of Data Arrangements

In relation to data infrastructure in particular, STS investigations have delved into a variety of research topics to better understand how data infrastructures accommodate the interdependence of data arrangements. Three topics that impact the interdependence of

data arrangements are the development of standards, communities of practice, and data work arenas.

Standards

The study of standards and their role is a significant topic in the digital realm, yet it is often difficult to imagine their ramifications (Edwards, 2004; Lampland & Star, 2009). In scientific as well as digital projects where there is continuing change, Infrastructure Studies underscores standard making as an ongoing process, countering the cultural assumption that standards represent a permanent solution. For local collective undertakings, standards manifest as community conventions or local norms that enable comparability and integrative activities. A view of change and the politics of standards in science is detailed by Edwards (2004) with the case of the meteorological community's quest since 1839 to develop atmospheric observations, weather forecasting, and climate modeling on a global scale. Some standards organizations have emerged alongside community awareness of the need for standards-making processes such as multiphase report generation or a request for feedback that ensures continuing review and update of a standard (e.g., Consultative Committee for Space Data Systems [CCSDS], 2016; International Organization for Standardization, 2019).

With mandates for sharing data (e.g., Holdren, 2013), motivations for Open Science (Hampton et al., 2015), articulation of principles (Wilkinson et al., 2016), and the generative power of aggregated data (Peters, 2010), data workers and researchers are gaining collective experience with data as well as metadata. Conventions and standards play a significant role in increasing access to and understanding of scientific research data. They contribute to the functionality of data infrastructures as researchers develop new practices associated with reuse of data collected by others. Metadata describes the context and the provenance of what could otherwise be decontextualized numerical data (Greenberg, 2010). Although metadata aims to document data in support of knowledge claims (Mayernik, 2019), metadata descriptions are frequently minimal. Zimmerman (2008) also reports on the importance of knowing the context associated with generation of ecological data to assess its quality and to be able to consider its limitations. Studies of standards as coordination mechanisms have highlighted the development of metadata specifications as standards-making processes (Millerand, Ribes, Baker, & Bowker, 2013; Yarmey & Baker, 2013). The Ecological Metadata Standard, with its formal structure for describing a data set, illustrates a schema in the environmental sciences that has been updated continuously since its emergence in 1997 (Fegraus, Andelman, Jones, & Schildhauer, 2005; Millerand & Bowker, 2009).

Research in Infrastructure Studies highlights the ever-present tension between standardization and flexibility. Hanseth, Monteiro, and Hatling (1996) report on the value of making standards visible and subject to update as critical for the design of information infrastructures. Shared data vocabularies emerge at local levels as well as domain levels in many situations. If data are to be shared, they must be packaged and annotated in agreed-upon ways. The heterogeneity of data and legacy data practices belies this task. While work on basic and advanced data system terminology is ongoing (CCSDS, 2012; Re-

search Data Alliance [RDA], N.D.A), vocabularies exist or are emerging for data-related terms such as metadata, levels of data products, computational modeling, and data privacy. Parameter and unit lists that are being developed most frequently are project or organization specific and sometimes regional or discipline wide.

Recent developments in standards associated with a growing number of catalogs and registries are less studied infrastructural elements that have helped make data visible. In addition to registries for data, physical objects, and scientific researchers made possible by unique identifiers, there are catalogs for data repositories (e.g., Anderson & Hodge, 2014; Pampel et al., 2013), controlled vocabularies (e.g., National Information Standards Organization, 2017), and more. These catalogs inform design of information systems and foster insight and supporting analysis. Catalogs and registries benefit from controlled vocabularies as a mechanism for improving retrieval of entries (e.g., Vierkant et al., 2012). Metadata including tags such as "keywords" also benefit from content control in lists, thesauri, and ontologies. The number of metadata standards was conveyed early on as a visualization (Riley, 2010) and then by development of a catalog (Ball, 2013) that eventually became an international consortium effort (Ball, Greenberg, Jeffery, & Koskela, 2016; RDA, N.D.B).

Communities of Practice

Communities of practice are investigated by researchers in many fields, including Infrastructure Studies, to better understand how to accommodate the interdependence of data arrangements. A community of practice is a group of people who share a concern or a passion for something they do and who learn how to do it better through regular interaction (Wenger, 1998). Denis and Pontille (2014, p. 164) hypothesize an increasingly important role in various fields for both "the elusiveness of the database and of collective work," which in broader terms includes coordinating data work across one or more communities, such as information managers at ecological research sites in the United States. The LTER Network joined together to form an information management community of practice. The community worked to translate local metadata conventions to a common metadata standard and to package data sets with their metadata so data could be assembled in a network repository, thereby extending their site-based data infrastructures. Communities of practice are new forms of institutional arrangements that may be considered infrastructural elements, often made visible by their mission statements and governance arrangements.

Infrastructure Studies has drawn attention to the collective work and learning that occurs via communities of practice. The notions of situated learning that occur in these communities (Lave, 1991; Lave & Wenger, 1991) represent a marked change from traditional classroom approaches to education and continuing learning. These ideas evolved into an approach to supporting knowledge work within organizations (Wenger, McDermott, & Snyder, 2002). In the contemporary research realm, communities of practice that cross the boundaries of organizations and disciplines bring together isolated data reposi-

tories. Such communities often become integrative forums that provide exposure to and communication about both similarities and differences in data and data systems.

Data Work Arenas

The third topic relating to interdependence of data arrangements is data work arenas. These arenas represent work unit groups that are much smaller than communities of practice. The increase in data and the growth of data work has led to new, often distributed and specialized arenas of work, stopping places in the flow of data. Data work arenas are sites where a particular set of data procedures and analyses are carried out by those with expertise in particular stages in the production of data. The expertise of a data processing group or a data visualization group may be distinct from that of data collection groups or analysis or metadata groups (Baker, 2017). Arenas exist in the field and the laboratory as well as within data centers and research institutes. They are associated with projects, programs, and departments. Data work arenas may be independent, nested, overlapping, intersecting, or interactive.

For data to be moved or migrated through various locations from the point of origin or generation to new locations (e.g., from a local database to a domain data archive), it is packaged so it can travel on what have been called "data journeys" (Leonelli, 2016). Any one path of the data using infrastructural supports can be described by a detailed data workflow (Bechhofer et al., 2013; Shaon et al., 2012). Data moving through these arenas often contradict idealized models that suggest a linear progression of management and analysis steps from field to archive. Although data may proceed through arenas sequentially (e.g., collection, calibration, processing, analysis), it may also proceed in a nonlinear manner where arenas may interact iteratively (as in the case of reprocessing and reanalysis), thereby creating feedback loops. An interdependent set of data work arenas may become recognized as a data infrastructure or as a component of a data infrastructure. Data work arenas may be established through ad hoc agreements or may be formalized through a growing number of mechanisms of coordination such as best practices or data policies.

Collective Data Management and Models

Concerns about data have changed as data infrastructure and data management practices expand and models develop. Initially, data work involved collecting observations and measurements for analysis as evidence in support of the scholarly production of knowledge. Important differences in data and data practices within disciplines and communities as well as between them are still being explored (e.g., Borgman, 2015), and there is growing recognition that data as a first-class research object gives rise to a new process, that of data production (Baker & Millerand, 2010). Understanding changes in data practices involves paying attention to new geographic scales, temporal extents, institutional arrangements, documentation issues, and distributions of data work. STS work has provided understandings of the impact of these changes on data practices, particularly with

respect to long-term research and collective data management that has led to the emergence of new data roles.

Collective data management activities involving the aggregation and organization of data from many community members brought recognition of the need for data systems, repositories, and platforms (e.g., Palmer, 2001). Site-based projects and networks in ecology provide early examples of collaborative work and approaches to data integration and data interoperability (Baker et al., 2000; Benson, Hanson, Chipman, & Bowser, 2006; Ingersoll, Seastedt, & Hartman, 1997; Michener, Porter et al., 2011). In these cases, data and science insights were presented and compared at data workshops that informed research narratives and knowledge making while plans were being made for coauthored papers and further field work.

An information environment supports shared resources and promotes dialogue critical to collective endeavors (Leonelli, 2016). Shared resources such as databases, data systems, project websites, and mailing lists are elements of information environments. Similar to the common information space discussed by Schmidt and Bannon (1992), an information environment provides a place for negotiation of agreed-upon interpretations of data and data procedures where differences are identified and accommodated and sometimes resolved. Such environments both enable sharing interpretations of data and facilitate cooperative decision making.

Figure 2 illustrates changes in data work spurred by collective data management and the growth of information environments. The dyad model is shown in the lower part of the diagram: a researcher working with an assistant helping with field work, instrumentation, or data analysis. As collaborative activities grow and data are aggregated within laboratories and other groups, a data management model develops to support group use of data via a designated data manager. When the development of longer-term perspectives and longer-term funding arrangements coincide in the digital age, an information management model develops to support community data use, a community-specific information environment, and a local data system supporting online data access. Strategic leadership and data vision in some cases leads to stability and funding for an informatics team model where the team is able to optimize for both immediate data use locally and for data delivery to external facilities, thereby creating the opportunity for data reuse over time. Data packaged in accordance with the requirements of other data facilities enables the data to travel upward from any of the lower levels (Leonelli, 2009, 2016). Each level in the figure represents added infrastructure supporting work with digital data.
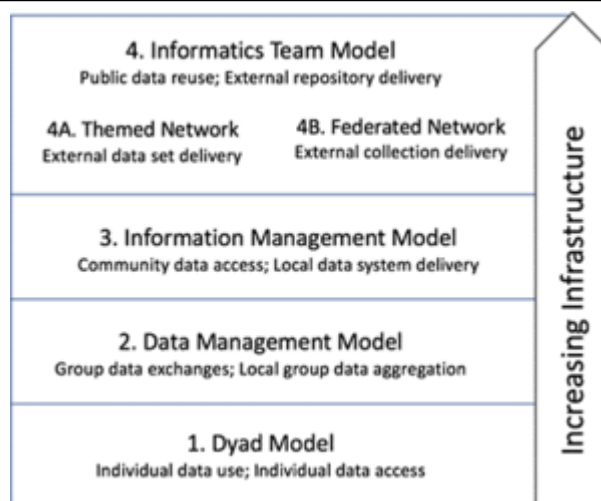
Figure 2. Data work models: Transitioning from an individual model to collective models.

Adapted from Baker and Millerand, 2010.

The increase in kinds and amounts of data as well as the advent of collective data management have led to the emergence of new kinds of analysis, expertise, and liaison work. Titles such as data manager, metadata specialist, data scientist, and software engineer designate specialists who work with data. Data or information management have been identified as dealing with a "trajectory" comprising many interdependent arenas of action that require continuous attention to managing data, supporting science, and designing technology (Karasti & Baker, 2004). Data specialists often perform mediation or liaison work. Data roles associated with data management are described in reports and case studies (Borgman, Wallis, & Enyedy, 2006; Karasti and Baker, 2008; Karasti et al., 2006; Meyer, 2009; NRC, 2015; Thompson, 2015), while the growth and distribution of data work remain topics of study critical to establishing and maintaining data quality.

A Multiplicity of Infrastructure Configurations

The term "data landscape" describes the larger context within which data resides and includes support elements such as data work arenas, repositories, centers, archives, and partners that may be positioned locally or remotely. Any arrangement of one or more of these elements may be designated a "data infrastructure" where data work occurs and across which data moves, ultimately arriving at one or more destinations. Potential partnering organizations and communities in academic, government, commercial, and public sectors may develop and support some of the elements. In practice, elements must align both internally and laterally to enable the movement of data from one location to another. The variety of configurations of infrastructure within the data landscape is large and only now beginning to be understood in terms of management and sustainability in addition to fostering invention and innovation. One example of infrastructure within an organization involves a researcher having support available to initiate submissions of a conference paper or poster to their institutional repository; another example with infrastructure that crosses institutional boundaries is the case of a researcher who annotates a data set with

metadata and sends this data package to an ecological repository external to their local institution. The many data infrastructure arrangements, each made up of multiple, interacting elements, aim to support the dynamics, complexity, and interdependencies of data in support of research efforts.

Expanding from the two-axis infrastructure (Figure 1) to four axes makes visible more of the work and decision making associated with configuring a data infrastructure. The two panels in Figure 3 describe two different cases by indicating some of the choices made—although they are typically expected to change over time. The first case illustrates data use (case A) where the data work targets support for a set of defined science-driven data practices for a small project that has little existing infrastructure available. These researchers are positioned to analyze and use the data they have generated in a field project but do not have a data management plan. This configuration shows minimal activity in terms of the larger data landscape and data reuse. In the second panel illustrating data use and reuse (case B), there is a formal data management plan and a locally grown infrastructure with capabilities and functionality developed for their local environment—conventions, procedures, and systems—that facilitate local data activities and inform practices of local researchers resulting in less time dedicated to support of individual needs. The local, collective infrastructure still requires maintenance and attention to update and redesign, but if it is well planned and carried out in partnership with others it may free up resources. The project may find new, more cost-effective approaches to internal services or identify external partners that can provide resources and expertise such as for web services. With less time dedicated to individual needs, more resources are available at a collective level as well as for long-term, global efforts where extra-institutional infrastructure provides support and guidance. Whereas case A is centered around local knowledge production, case B is concerned with the production of both knowledge and data.
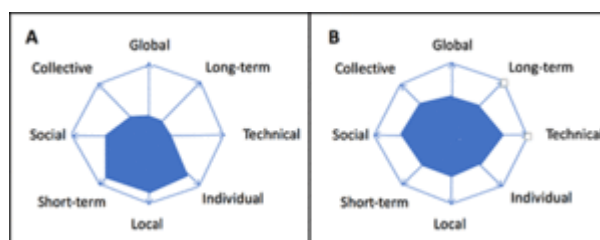

Figure 3. Two of the many possible configurations of data infrastructure illustrated by showing their differing distribution of choices pertaining to data effort across eight characteristics on four axes: (A) case of data use with ad hoc plans for data work and (B) case of data use and reuse with a formal data management plan that includes data preservation and access.

A few examples illustrate the scope of each axis in Figure 3:

(1) The sociotechnical axis representing a continuum from *social* to *technical* refers to data work practices that may draw on socially oriented approaches, such as individual-to-individual informal data exchanges, or may involve more technical arrangements, such as the ability to create and deliver local metadata digitally in the standard format of a partner.

(2) The audience axis of *local* and *global* reach defines the scope of services. One end of the spectrum focuses on participants familiar with the data generation and data context as well as with conventions tailored to the local venue. At the other end of the spectrum, *global* refers to more general services remote from the data origin, encompassing a large geographic or thematic scope and involving standards in terms of formats and metadata.

(3) The coordination axis stretching between *individual* and *collective* refers to the notion of targeting assembly and management of data for participants individually for the benefit of a designated project, community, or domain.

(4) The design axis represents a continuum from *short-term* to *long-term activities*. Short-term refers to data use for knowledge production and long-term adds to local data work by engaging in data reuse via data production for preservation and access. Design choices involve sampling, processing, and data generation, and time needed for data packaging and submission to repositories. A short-term case involves informal assembly and use of data, whereas those with longer-term plans are more formal in data systems aimed at providing a public interface for access to preserved data.

The shaded polygon for case B is more centered than for case A. In case A, data generation is locally oriented to serve needs of individual researchers, and data work remains informal with no long-tern vision for data reuse. In case B, the added demands on time, budgets, and coordination that involve assembling, packaging, partnering, preserving, and disseminating data for reuse are evident.

Future Research Opportunities, Supporting the Data Landscape

Given the diversity of data and data configurations, one view of current efforts is that they represent pilot studies providing experience to designers as well as researchers struggling with data practices, workflows, and infrastructure making. Data sharing is prompting change and a growing awareness of the diversity, dynamics, and change inherent to the research data landscape and to high-quality scientific research. A number of unsettled subjects of ongoing discussion and debate were mentioned earlier—open data, data ethics, and data interoperability. Four additional topics, important opportunities for future research that support the data landscape, are considered in this section: the sustainability of infrastructures; the politics of data infrastructure; the well-being of natural,

social, and information systems as a trio of interdependent systems; and data infrastructure literacy and partnering.

Sustainability of Infrastructures

Questions about approaches and costs for the sustainability of infrastructures are topics of active discussion. Although print archives and museums have focused on preservation of materials for some time, the need for long-term preservation and subsequent provision of public access to digital data sets and to research project collections are new concerns. A mistaken assumption about infrastructure's role in supporting data work is that it can be considered a permanent solution to a problem. Rather, the existence of infrastructure requires, in the long view, watchfulness (Orr, 1996) and care for its fragility and continuing maintenance (Mol, 2008; Puig de la Bellacasa 2011). Edwards (2003) describes infrastructure as an "artificial environment" that, unlike geophysical systems that change slowly, fail because their development assumes design and maintenance are orderly, dependable, and separate from human, technical, and environmental environments.

Sustainability planning requires the development of shared ontologies (Geels, 2010; NRC, 2014) as well as shared understandings of the various characteristics and configurations of infrastructure. For instance, by the 1980s the need to coordinate satellite imagery and remote sensing efforts became evident, spurring development of agreements and vocabularies as the concepts of information systems and archives were brought together by the international satellite community (Albani, Molch, Maggio, & Cosac, 2015; CCSDS, 2012). In commenting on sustainability of digital access to data and information, the Blue Ribbon Task Force (2010) report identified risks and presented recommendations on multiple dimensions: organizational, technical, public policy, and education/outreach. To address contemporary worldwide data efforts, the United Nations (UN, 2014) reported on the necessity of "mobilising the Data Revolution for Sustainable Development" to address environmental change. Support for this aim includes coordination of data efforts at individual and community levels as well as at international levels. Some global-scale efforts include the Belmont Forum supported by international science councils and funding (Allison & Gurney, 2015), the nation-based International Council for Science Committee on Data (2014), and the multistakeholder-based Research Data Alliance (Berman, 2014).

In planning for long-term infrastructure development, whether for data generation, production, or preservation, the design notion of iterative development emerges as one key to sustainability. The iterative nature of design work occurring over time highlighted by Greenbaum and Kyng (1991) is being explored via the concept of infrastructuring (e.g., Karasti & Baker, 2004; Karasti et al., 2018; Pipek & Wulf, 2009), with emphasis on the processual nature of infrastructure making.

The Politics of Infrastructure

An investigation of the politics of infrastructure expands on the concern for sustainability to consider the ethics associated with the support and use of data and data infrastruc-

ture. Amid the politics of design (Dourish, 2010; Parmiggiani, 2017), gateways (Edwards et al., 2007; Zimmerman & Finholt, 2007), and data care (Baker & Karasti, 2018), swirl questions in the sciences about who gets to learn by designing, what gets reinforced via technology (Kraemer & King, 2006), when balances are struck, and how infrastructure can accommodate change. Forms of intervention and engagement are the subject of ongoing discussion in STS research (e.g., Ribes & Baker, 2007), including activist approaches that foreground the politics of neglected things (Baker & Karasti, 2018). Social science researchers with qualitative methodologies can raise awareness and share knowledge about the politics of infrastructure, including the impact of categories, standards, and information systems as well as the processes of design and use within a community (e.g., Karasti et al., 2018; Pendleton-Julian & Brown, 2018).

Larkin (2013, p. 330) sees infrastructure holistically as something with an ontology involving power, culture, and imaginative practices. He declares the concept of infrastructure "unruly," pointing out that "the act of defining an infrastructure is a categorizing moment. Taken thoughtfully, it comprises a cultural analytic that highlights the epistemological and political commitments involved in selecting what one sees as infrastructural (and thus causal) and what one leaves out." Envisioning a data infrastructure as only technical elements, or as merely ecological science and computer-IT people, excludes other key participants such as data managers and data service specialists. Trigg and Ishimaru (2012, p. 219) detail the challenges of gaining organizational support for creating new "structures of participation: informal and formal liaisons, formal committees, working groups formed around specific projects, and our own incorporation [as social science researchers] into a department of Information Management." Finally, Jensen and Morita (2015) see today's infrastructures as experiments with an outcome of relating to politics and power as well as an experience base from which to imagine new possibilities.

Well-Being of Natural Systems, Social Systems, and Information Systems

In responding to grand challenges and creating data infrastructures during the anthropocene, there is great value in considering the "well-being" of a trio of interdependent systems. Not long ago, the twin concepts of ecosystem services (e.g., Daily et al., 1997) and of coupled human-natural systems (e.g., Brymer et al., 2020; Liu et al., 2007) opened up research on the earth's environment. The recognition of the well-being of natural or earth systems together with human needs and desires as an interplay of the earth and of our superimposed social systems, represents a profound conceptual move. Here, "social systems" refers to human, economic, and political developments. The addition of "information systems" as a third component in an overarching socio-info-natural system establishes the role of information as a primary agent and captures the interplay of the three systems in the digital era. The term "information systems" is used broadly, inclusive of digital technology, data and information infrastructures, and sociotechnical factors. By making information systems explicit rather than subsuming them within "social systems," a richer context is created within which decisions must be made and balances established

to ensure the well-being of natural systems, human populations, and the data that undergird society's knowledge.
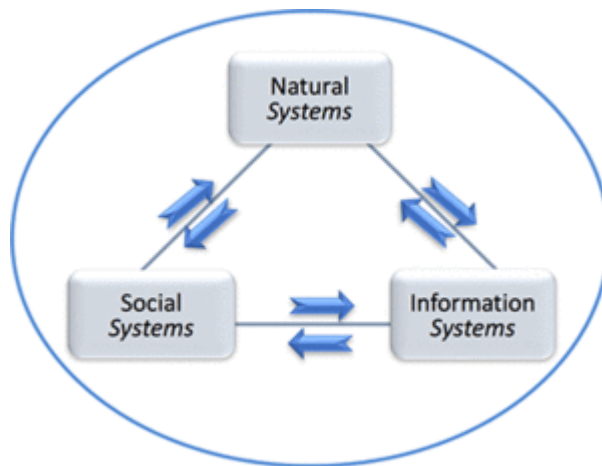

Figure 4. Conceptualizing a trio of interdependent systems.

Conceptualizing this trio of interdependent systems—natural systems, social systems, information systems—ties together the confluence of forces that impact data work for which policies may be developed (Figure 4). Within information systems writ large, data is now associated with emergent information services and data products of central importance to scientific research. Further, there is a new sense of urgency about increasing efficiency while maintaining the effectiveness of data infrastructures because the data undergird scientific research that is being called upon to support societal policy making at the scale required for a managed earth. One aim of data infrastructure is to transform the perceptions of data from that of a deluge into a readily findable and accessible set of resources.

Data Infrastructure Literacy and Partnering

In developing the concept of "data infrastructure literacy", the aim is "to make space for inquiry, experimentation, imagination and intervention." Gray, Gerlitz, and Bounegru (2018, p. 1). These authors remind us that "critical engagement with data infrastructures has been central to various interdisciplinary perspectives from the past several decades, including infrastructure studies, data studies, science and technology studies, the history and philosophy of science, human-computer interaction, computer supported cooperative work, ethnomethodology, the history and sociology of quantification, software studies, platform studies, new media studies, critical design studies and associated fields" Gray, Gerlitz, and Bounegru (2018, p. 3).

Despite this long list, there are more items to add, such as sociotechnical systems, information sciences, change management science, and institutional science in addition to ecology, which has together with technology and library efforts developed project-, program-, and discipline-oriented research and data infrastructures. Many fields are acquir-

ing experience and developing ontologies aiming to capture the variety of data arrangements that can enrich growth of infrastructure if mingled rather than retained separately. All are striving to understand the dynamic interplay of data management and data infrastructures in a data landscape that encompasses local scale efforts, midsize collectives, and larger-scale endeavors.

An STS perspective questions whether there is wide understanding about the aggregation and evolution of data management, the data roles emerging in response to the data work distributed across data work arenas, or how data flows across the heterogenous data landscape in the interdisciplinary field of ecology. Amid differing disciplinary perspectives such as the aims of elegant universal solutions by computer science or of diverse epistemic cultures as well as local data practices recognized by STS research, there is an ongoing emergence of insights. It behooves ecologists to develop a critical eye for infrastructure design and formation because, as interest grows in supporting data infrastructures, larger-scale enterprises develop that enable normative forces such as monitoring, auditing, and governing within layers of infrastructure management where administrative values and bureaucratic narratives flourish. Data sharing and data infrastructures create new responsibilities that require ecologists to engage in opportunities to plan, experiment, learn, and reshape data infrastructures to ensure that their science and its inquiry-driven, innovative core are served appropriately in the 21st century. Ecologists, data specialists, and social scientists such as STS researchers will benefit from active partnerships to ensure data infrastructures effectively support scientific investigative processes.

Further Reading

Cheruvelil, K. S., & Soranno, P. A. (2018). Data-intensive ecological research is catalyzed by open science and team science. *BioScience*, *68*(10), 813–822.

Edwards, P. N., Bowker, G. C., Jackson, S. T., & Williams, R. (Eds.). (2009). Special issue on infrastructure studies. *Journal of the Association for Information Systems*, *10*(5).

Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L. (2011). Science friction: Data, metadata, and collaboration. *Social Studies of Science*, *41*(5), 667–690.

Gosz, J. R. (1999). Ecology challenged? Who? Why? Where is this headed? *Ecosystems*, *2*(6), 475–481.

Heidorn, B. (2008). Shedding light on dark data in the long tail of science. *Library Trends*, *57*(2), 280–299.

Hine, C. (2006). *New infrastructures for knowledge production: Understanding e-science*. London, UK: Information Science Publishing.

Lee, C. P., Ribes, D., Bietz, M. J., Karasti, H., & Jirotka, M. (Eds.). (2010). Special issue: Sociotechnical studies of cyberinfrastructure and e-research—Supporting collaborative research. *Computer Supported Cooperative Work*, *19*(3–4).

Monteiro, E., Pollock, N., & Williams, R. (2014). Special issue on innovation in information infrastructures. *Journal of the Association for Information Systems*, *15*(4).

Ribes, D., & Finholt, T. A. (2007). **Tensions across the scales: Planning infrastructure for the long-term**. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work* (pp. 229–238). New York, NY: Association for Computing Machinery.

Robertson, G. P., Collins, S. L., Foster, D. R., Brokaw, N., Ducklow, H. W., Gragson, T. L., . . . Moore, J. C. (2012). Long-term ecological research in a human-dominated world. *BioScience*, *62*(4), 342–353.

Zimmerman, A. (2007). Not by metadata alone: The use of diverse forms of knowledge to locate data for reuse. *International Journal on Digital Libraries*, *7*(1–2), 5–16.

References

Albani, M., Molch, K., Maggio, I., & Cosac, R. (2015). **Long term preservation of earth observation space data, preservation workflow (version 1)** CEOS-WGISS.

Allison, L., & Gurney, R. (2015). ***A place to stand: E-infrastructures and data management for global change research—Belmont forum e-infrastructures and data management community strategy and implementation plan***.

Andelman, S. J., Bowles, C. M., Willig, M. R., & Waide, R. B. (2004). Understanding environmental complexity through a distributed knowledge network. *BioScience*, *54*(3), 243–249.

Anderson, N., & Hodge, G. (2014). ***Repository registries: Characteristics, issues and futures***.

Anderson, S. P., Bales, R. C., & Duffy, C. J. (2008). Critical zone observatories: Building a network to advance interdisciplinary study of Earth surface processes. *Mineralogical Magazine*, *72*(1), 7–10.

Aronova, E., Baker, K. S., & Oreskes, N. (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*(2), 183–224.

Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., . . . Wright, M. H. (2003). ***Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure***. Arlington, VA: National Science Foundation.

Baker, K. S. (2017). ***Data work configurations in the field-based natural sciences: Mesoscale infrastructures, project collectives, and data gateways*** (PhD diss., University of Illinois at Urbana-Champaign).

Baker, K. S., Benson, B. J., Henshaw, D. L., Blodgett, D., Porter, J. H., & Stafford, S. G. (2000). Evolution of a multisite network information system: The LTER information management paradigm. *BioScience*, *50*(11), 963–978.

Baker, K. S., & Duerr, R. E. (2017). Data and a diversity of repositories. In L. Johnston (Ed.), *Curating research data; Vol. 2, A handbook of current practice* (pp. 139–144). Chicago, IL: Association of College and Research Libraries.

Baker, K. S., & Karasti, H. (2018). **Data care and its politics: Designing for a neglected thing**. In *Proceedings of the 15th Participatory Design Conference (PDC '18). Genk, Belgium, August 2018* (pp. 1–12). New York, NY: Association for Computing Machinery.

Baker, K. S., & Millerand, F. (2007). **Articulation work supporting information infrastructure design: Coordination, categorization, and assessment in practice**. In *Proceedings of the 40th Hawaii International Conference on System Sciences, 3–6 January 2007 Waikoloa, Big Island, Hawaii* (pp. 1–10). Washington, DC: IEEE Computer Society.

Baker, K. S., & Millerand, F. (2010). Infrastructuring ecology: Challenges in achieving data sharing. In J. N. Parker, N. Vermeulen, & B. Penders (Eds.), *Collaboration in the new life sciences* (pp. 111–136). Farnham, UK: Ashgate.

Baker, K. S., & Yarmey, L. (2009). Data stewardship: Environmental data curation and a web-of-repositories. *International Journal of Digital Curation*, *4*(2), 12–27.

Ball, A. (2013). ***UK research data registry mapping schemes***.

Ball, A., Greenberg, J., Jeffery, K., & Koskela, R. (2016). ***RDA metadata standards directory working group: Final report***. Research Data Alliance.

Bechhofer, S., Buchanb, I., Roure, D. D., Missier, P., Ainsworth, J., Bhagat, J., . . . Goble, C. (2013). **Why linked data is not enough for scientists**. *Future Generation Computer Systems*, *29*(2), 599–611.

Benson, B. J., Hanson, P. C., Chipman, J. W., & Bowser, C. J. (2006). Breaking the data barrier: Research facilitation through information management. In J. J. Magnuson, T. K. Kratz, & B. J. Benson (Eds.), *Long-term dynamics of lakes in the landscape* (pp. 259–279). Oxford, UK: Oxford University Press.

Berman, F. (2014). **Building global infrastructure for data sharing and exchange through the research data alliance**. *D-Lib*, *20*(1/2).

Bietz, M. J., Baumer, E. P., & Lee, C. P. (2010). Synergizing in cyberinfrastructure development. *Computer Supported Cooperative Work*, *19*(3–4), 245–281.

Bietz, M. J., Ferro, T., & Lee, C. P. (2012). **Sustaining the development of cyberinfrastructure: An organization adapting to change**. In *Proceedings of the ACM 2012 Con-*

ference on computer supported cooperative work, 11–15 February 2012 (pp. 901–910). New York, NY: Association for Computing Machinery.

Birnholtz, J., & Bietz, M. (2003). **Data at work: Supporting sharing in science and engineering**. In *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work*, Sanibel Island, FL, November 2003 (pp. 339–343). New York, NY: Association for Computing Machinery.

Björgvinsson, E., Ehn, P., & Hillgren, P.-A. (2012). Design things and design thinking: Contemporary participatory design challenges. *Design Issues*, *28*(3), 101–116.

Blue Ribbon Task Force. (2010). ***Blue Ribbon Task Force report on sustainable economics for a digital planet: Ensuring long-term access to digital information***. Washington, DC: National Science Foundation.

Bocking, S. (2010). Organizing the field: Collaboration in the history of ecology and environmental science. In J. N. Parker, N. Vermeulen, & B. Penders (Eds.), *Collaboration in the new life sciences* (pp. 15–36). Farnham, UK: Ashgate.

Borgman, C. (2015). *Big data, little data, no data*. Cambridge, MA: MIT Press.

Borgman, C., Wallis, J. C., & Enyedy, N. (2006). Building digital libraries for scientific data: An exploratory study of data practices in habitat ecology. In J. Gonzalo, C. Thanos, M. F. Verdejo, & R. C. Carrasco (Eds.), *Research and advanced technology for digital libraries* (pp. 170–183). Berlin, Germany: Springer.

Bowker, G. C. (1994). *Science on the run: Information management and industrial geophysics at Schlumberger, 1920–1940*. Cambridge, MA: MIT Press.

Bowker, G. C. (2014). The infrastructural imagination. In A. Mongili & G. Pellegrino (Eds.), *Information infrastructure(s): Boundaries, ecologies, multiplicity*. Newcastle, UK: Cambridge Scholars Publishing.

Bowker, G. C. (2017). How knowledge infrastructures learn. In P. Harvey, C. Bruun Jensen, & A. Morita (Eds.), *Infrastructures and social complexity: A companion* (pp. 391–404). New York, NY: Routledge.

Bowker, G. C., Baker, K. S., Millerand, F., & Ribes, D. (2010). Toward information infrastructure studies: Ways of knowing in a networked environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International handbook of internet research* (pp. 97–117). Dordrecht, The Netherlands: Springer.

Brymer, A. L. B., Toledo, D., Spiegal, S., Pierson, F., Clark, P., & Wulfhorst, J. (2020). Social-ecological processes and impacts affect individual and social well-being in a rural western US landscape. *Frontiers in Sustainable Food Systems*, *4*(38), 1–16.

Carpenter, S. R., Armbrust, E. V., Arzberger, P. W., Chapín, F. S., Elser, J. J., Hackett, E. J., . . . Zimmerman, A. S. (2009). **Accelerate synthesis in ecology and environmental sciences**. *BioScience*, *59*(8), 699–701.

Consultative Committee for Space Data Systems. (2012). ***Reference model for an Open Archival Information System (OAIS)*** (CCSDS No. 650.0-M-2). Washington, DC: Consultative Committee for Space Data Systems.

Consultative Committee for Space Data Systems. (2016). ***Organization and processes for the Consultative Committee for Space Data Systems* (CCSDS No. A02.1-Y-4)**. Washington, DC: Consultative Committee for Space Data Systems.

Contreras, J. L., & Reichman, J. H. (2015). Sharing by design: Data and decentralized commons. *Science*, *350*(6266), 1312–1314.

Cragin, M. H., & Shankar, K. (2006). Scientific data collections and distributed collective practice. *Computer Supported Cooperative Work*, *15*(2–3), 185–204.

Cutcher-Gershenfeld, J. (2018). **Assumptions wrangling: An experiment in culture change**. *Heller Magazine*, (Winter).

Cutcher-Gershenfeld, J., Baker, K. S., Berente, N., Carter, D. R., DeChurch, L. A., Flint, C. C., . . . Zaslavsky, I. (2016). **Build it, but will they come? A geoscience cyberinfrastructure baseline analysis**. *Data Science Journal*, *15*, 8.

Daily, G. C., Alexander, S., Ehrlich, P. R., Goulder, L., Lubchenco, J., Matson, P. A., . . . Woodwell, G. M. (1997). Ecosystem services: Benefits supplied human societies by Natural Ecosystems. *Ecology*, *2*, 1–16.

Denis, J., & Pontille, D. (2012). **Workers of writing, materials of information**. *Revue d'anthropologie des connaissances*, *6-1*(1), a–s.

Denis, J., & Pontille, D. (2014). Parasite users? The volunteer mapping of cycling infrastructures. In A. Mongili & G. Pellegrino (Eds.), *Information infrastructure(s): Boundaries, ecologies, multiplicity*. Newcastle, UK: Cambridge Scholars Publishing.

Dourish, P. (2010). **HCI and environmental sustainability: The politics of design and the design of politics**. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems, Aarhus, Denmark, August 2010* (pp. 1–10). New York, NY: Association for Computing Machinery.

Edwards, P. N. (2003). Infrastructure and modernity: Force, time, and social organization in the history of sociotechnical systems. In T. J. Misa, P. Brey, & A. Feedberg (Eds.), *Modernity and technology* (pp. 185–225). Cambridge, MA: MIT Press.

Edwards, P. N. (2004). "A vast machine": Standards as a social technology. *Science*, *304*(5672), 827–828.

Edwards, P. N. (2010). *A vast machine: Computer models, climate data, and the politics of global warming*. Cambridge, MA: MIT Press.

Edwards, P. N., Bowker, G. C., Jackson, S. J., & Williams, R. (2009). Introduction: An agenda for infrastructure studies. *Journal of the Association for Information Systems*, *10*(5), 364–374.

Edwards, P. N., Jackson, S. J., Bowker, G. C., & Knobel, C. P. (2007). **Understanding infrastructure: Dynamics, tensions, and design**.

Edwards, P. N., Jackson, S. J., Chalmers, M. K., Bowker, G. C., Borgman, C. L., Ribes, D., . . . Calvert, S. (2013). **Knowledge infrastructures: Intellectual frameworks and research challenges**.

European Research Infrastructure Consortium. (2009). ***Community legal framework for a European Research Infrastructure Consortium (ERIC)*** (Council Regulation [EC] No. 723/2009). Brussels, Belgium: European Commission.

European Union. (2016). ***Open innovation, open science, open to the world: A vision for Europe***. Brussels, Belgium: European Commission.

Faniel, I. M., & Zimmerman, A. (2011). Beyond the data deluge: A research agenda for large scale data sharing and re-use. *International Journal of Digital Curation*, *6*(1), 59.

Fecher, B., & Friesike, S. (2014). Open science: One term, five schools of thought. In *Opening Science* (pp. 17–47). New York, NY: Springer.

Fegraus, E. H., Andelman, S., Jones, M. B., & Schildhauer, M. (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*(3), 158–168.

Felt, U., Fouché, R., Miller, C. A., & Smith-Doerr, L. (Eds.). (2016). *The handbook of science and technology studies* (4th ed.). Cambridge, MA: MIT Press.

Frischmann, B. M., Madison, M. J., & Strandbury, K. J. (2014). *Governing knowledge commons*. New York, NY: Oxford University Press.

Geels, F. W. (2010). Ontologies, socio-technical transitions (to sustainability), and the multi-level perspective. *Research Policy*, *39*(4), 495–510.

Gewin, V. (2016). An open mind on open data. *Nature, 529*(7584), 117–119.

Gray, J., Gerlitz, C., & Bounegru, L. (2018). **Data infrastructure literacy**. *Big Data and Society*.

Greenbaum, J., & Kyng, M. (Eds.). (1991). *Design at work: Cooperative design of computer systems*. Hillsdale, NJ: Lawrence Erlbaum.

Greenberg, J. (2010). Metadata and digital information. In *Encyclopedia of library and information sciences* (3rd ed., pp. 3610–3623). New York, NY: Taylor & Francis.

Hamilton, A. (2000). Metaphor in theory and practice: The influence of metaphors on expectations. *ACM Journal of Computer Documentation*, *24*(4), 237–253.

Hampton, S. E., Anderson, S. S., Bagby, S. C., Gries, C., Han, X., Hart, E. M., . . . Michener, W. K. (2015). The Tao of open science for ecology. *Ecosphere*, *6*(7), 1–13.

Hampton, S. E., Jones, M. B., Wasser, L. A., Schildhauer, M. P., Supp, S. R., Brun, J., . . . Gross, L. J. (2017). Skills and knowledge for data-intensive environmental research. *BioScience*, *67*(6), 546–557.

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., . . . Porter, J. H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, *11*(3), 156–162.

Hanseth, O., Monteiro, E., & Hatling, M. (1996). Developing information infrastructure: The tension between standardization and flexibility. *Science, Technology and Human Values*, *21*(4), 407–426.

Harvey, P., Bruun Jensen, C., & Morita, A. (Eds.). (2017). *Infrastructures and social complexity: A companion*. New York, NY: Routledge.

Hine, C. (2006). Databases as scientific instruments and their role in the ordering of scientific work. *Social Studies of Science*, *36*(2), 269–298.

Hobbie, J. E. (2003). Scientific accomplishments of the long term ecological research program: An introduction. *BioScience*, *53*(1), 17–20.

Holdren, O. P. (2013). Increasing access to the results of federally funded scientific research [Memorandum]. Washington, DC: US Office of Science and Technology Policy.

Ingersoll, R. C., Seastedt, T. R., & Hartman, M. (1997). A model information management system for ecological research. *BioScience*, *47*(5), 310–316.

International Council for Science. (2014). *Review of CODATA, the Committee on Data for Science and Technology; Report to the ICSU Committee on Scientific Planning and Review*.

International Organization for Standardization. (2019). *Guidance on the systematic review process in ISO*.

Inter-university Consortium for Political and Social Research. (2013). *Sustaining domain repositories for digital data: A call for change from an interdisciplinary working group of domain repositories*.

Jackson, S. (2014). Rethinking repair. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 221–240). Cambridge, MA: MIT Press.

Jackson, S. J., Edwards, P. N., Bowker, G. C., & Knobel, C. P. (2007). **Understanding infrastructure: History, heuristics and cyberinfrastructure policy**. *First Monday*, *12*(6).

Jackson, S. J., Ribes, D., Buyuktur, A., & Bowker, G. C. (2011). **Collaborative rhythm: Temporal dissonance and alignment in collaborative scientific work**. In *Proceedings of the ACM 2011 Conference on computer supported cooperative work, Hangshou, China, March 2011* (pp. 245–254). New York, NY: Association for Computing Machinery.

Jankowski, N. W. (2007). **Exploring e-science: An introduction**. *Journal of Computer-Mediated Communication*, *12*(2), 549–562.

Jensen, C. B., & Morita, A. (2015). **Infrastructures as ontological experiments (I)**. *Engaging Science, Technology, and Society*, *1*, 81–87.

Jones, M. B., Schildhauer, M. P., Reichman, O. J., & Bowers, S. (2006). The new bioinformatics: Integrating ecological data from the gene to the biosphere. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 519–544.

Jopp, F., Reuter, H., & Breckling, B. (Eds.). (2011). *Modelling complex ecological dynamics: An introduction into ecological modelling for students, teachers and scientists*. Heidelberg, Germany: Springer.

Karasti, H. (2014, October). **Infrastructuring in participatory design**. In *Proceedings of the 13th Participatory Design Conference, Windhoek Nambia* (pp. 141–150). New York, NY: Association for Computing Machinery.

Karasti, H., & Baker, K. S. (2004). **Infrastructuring for the long-term: Ecological information management**. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, Big Island, HI, 5-8 January 2004* (pp. 1–10). Piscataway, NY: IEEE.

Karasti, H., & Baker, K. S. (2008). Digital data practices and the long term ecological research program growing global. *International Journal of Digital Curation*, *3*(2), 42–58.

Karasti, H., Baker, K. S., & Halkola, E. (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*, *15*(4), 321–358.

Karasti, H., Baker, K. S., & Millerand, F. (2010). Infrastructure time: Long-term matters in collaborative development. *Computer Supported Cooperative Work*, *19*(3–4), 377–415.

Karasti, H., & Blomberg, J. (2018). **Studying infrastructuring ethnographically**. *Computer Supported Cooperative Work*, 27(2), 233–265.

Karasti, H., Millerand, F., Hine, C. M., & Bowker, G. C. (2016). Knowledge infrastructures: Part I. *Science and Technology Studies*, *29*(1), 2–12.

Karasti, H., Pipek, V., & Bowker, G. C. (2018). **An afterword to "Infrastructuring and collaborative design"**. *Journal of Computer Supported Cooperative Work*, *27*(2), 267–289.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Los Angeles, CA: SAGE.

Kraemer, K. L., & King, J. L. (2006). **Information technology and administrative reform: Will e-government be different?** *International Journal of Electronic Government Research*, *2*(1), 1–20.

Kratz, J., & Strasser, C. (2014). **Data publication consensus and controversies**. *F1000Research*, *3*(94).

Kwa, C. (1987). Representations of nature mediating between ecology and science policy: The case of the International Biological Programme. *Social Studies of Science*, *17*(3), 413–442.

Kwa, C. (2005). **Local ecologies and global science: Discourses and strategies of the International Geosphere-Biosphere Programme**. *Social Studies of Science*, *35*(6), 923–950.

Kwa, C., & Rector, R. (2010). A data bias in interdisciplinary cooperation in the sciences: Ecology in climate change research. In J. N. Parker, N. Vermeulen, & B. Penders (Eds.), *Collaboration in the new life sciences* (pp. 161–176). Farnham, UK: Ashgate.

Lampland, M., & Star, S. L. (2009). Reckoning with standards. In M. Lampland & S. L. Star (Eds.), *Standards and their stories: How quantifying, classifying, and formalizing practices shape everyday life* (pp. 3–24). Ithaca, NY: Cornell University Press.

Larkin, B. (2013). The politics and poetics of infrastructure. *Annual Review of Anthropology*, *42*, 327–343.

Lave, J. (1991). Situated learning in communities of practice. In L. B. Resnick, J. M. Levine, & S. D.Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 63–82). Washington, DC: American Psychological Association.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

Lee, C. P., Dourish, P., & Mark, G. (2006). **The human infrastructure of cyberinfrastructure**. In *Proceedings of the 2006 20th anniversary conference on computer supported cooperative work, Banff, Canada November 2006* (pp. 483–492). New York, NY: Association for Computing Machinery.

Leonelli, S. (2007). *Weed for thought: Using Arabidopsis thaliana to understand plant biology*. (PhD diss., Vrije Universiteit Amsterdam).

Leonelli, S. (2009). On the locality of data and claims about phenomena. *Philosophy of Science*, *76*(5), 737–749.

Leonelli, S. (2013). Global data for local science: Assessing the scale of data infrastructures in biological and biomedical research. *BioSocieties*, *8*(4), 449–465.

Leonelli, S. (2016). *Data-centric biology: A philosophical study*. Chicago, IL: University of Chicago Press.

Likens, G. E. (Ed.). (1987). *Long-term studies in ecology: Approaches and alternatives*. New York, NY: Springer.

Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E., . . . Lubchenco, J. (2007). Complexity of coupled human and natural systems. *Science*, *317*(5844), 1513–1516.

Loescher, H., Kelly, E., & Lea, R. (2017). National ecological observatory network: Beginnings, programmatic and scientific challenges, and ecological forecasting. In A. Chabbi & H. W. Loescher (Eds.), *Terrestrial ecosystem research infrastructures: Challenges, new developments and perspectives* (pp. 16–42). New York, NY: CRC Press.

Magnuson, J. J. (1989). Sustained or long-term ecological research. *Ecology*, *70*(5), 1553–1554.

Magnuson, J. J. (1990). Long-term ecological research and the invisible present. *BioScience*, *40*(7), 495–501.

Maron, N. L., & Loy, M. (2011). *Funding for sustainability: How funders' practices influence the future of digital resources*.

Mayernik, M. S. (2019). Metadata accounts: Achieving data and evidence in scientific research. *Social Studies of Science*, *49*(5), 732–757.

Mayernik, M. S., Wallis, J. C., & Borgman, C. L. (2013). Unearthing the infrastructure: Humans and sensors in field-based scientific research. *Computer Supported Cooperative Work*, *22*(1), 65–101.

Mervis, J. (2016). **NSF director unveils big ideas, with an eye on the next president and congress**. *Science*, *352*(6287), 755–756.

Meyer, E. T. (2009). Moving from small science to big science: Social and organizational impediments to large scale data sharing. In N. W. Jankowski (Ed.), *E-research: Transformation in scholarly practice* (pp. 222–239). New York, NY: Routledge.

Michener, W. K. (2000). Metadata. In W. K. Michener & J. W. Brunt (Eds.), *Ecological data: Design, management and processing* (pp. 92–116). Oxford, UK: Blackwell Science.

Michener, W. K., Porter, J., Servilla, M., & Vanderbilt, K. (2011). Long term ecological research and information management. *Ecological Informatics*, *6*, 13–24.

Michener, W., Vieglais, D., Vision, T., Kunze, J., Cruse, P., & Janée, G. (2011). DataONE: Data observation network for earth-preserving data and enabling innovation in the biological and environmental sciences. *D-Lib Magazine*, *17*(1), 3.

Michener, W. K., & Brunt, J. W. (Eds.). (2000). *Ecological data: Design, management and processing*. Oxford, UK: Blackwell Science.

Millerand, F. (2012). **La science en réseau, Les gestionnaires d'information "invisibles" dans la production d'une base de données scientifiques**. *Revue d'Anthropologie des Connaissances*,*6*(1), 163–190.

Millerand, F., & Bowker, G. C. (2009). Metadata standards: Trajectories and enactment in the life of an ontology. In M. Lampland & S. L. Star (Eds.), *Standards and their stories: How quantifying, classifying, and formalizing practices shape everyday life* (pp. 149–165). Ithaca, NY: Cornell University Press.

Millerand, F., Ribes, D., Baker, K., & Bowker, G. C. (2013). **Making an issue out of a standard: Storytelling practices in a scientific community**. *Science, Technology and Human Values*, *38*(1), 7–43.

Mol, A. (2008). *The logic of care: Health and the problem of patient choice*. Abingdon, UK: Routledge.

Mongili A., & Giuseppina, P. (Eds.). (2014). *Information infrastructure(s): Boundaries, ecologies, multiplicity*. Newcastle, UK: Cambridge Scholars Publishing.

Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). **Cloudy, increasingly FAIR: Revisiting the FAIR Data guiding principles for the European Open Science Cloud**. *Information Services and Use*,*37*(1), 49–56.

Nadim, T. (2016). Data labours: How the sequence databases GenBank and EMBL-Bank make data. *Science as Culture*, *25*(4), 496–519.

National Academies of Sciences, Engineering, and Medicine. (2018). *Open science by design: Realizing a vision for 21st century research*. Washington, DC: National Academies Press.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2009).. *Ensuring the integrity, accessibility, and stewardship of research data in the digital age*. Washington, DC: National Academies Press.

National Information Standards Organization. (2017). ***National Information Standards Organization: Issues in vocabulary management*** (Technical Report No. TR-06-2017). Baltimore, MD: NISO.

National Research Council. (1999). *Our common journey: A transition toward sustainability*. Washington, DC: National Academies Press.

National Research Council. (2001). *Grand challenges in environmental sciences*. Washington, DC: National Academies Press.

National Research Council. (2004). *Open access and the public domain in digital data and information for science: Proceedings of an international symposium*. Washington, DC: National Academies Press.

National Research Council. (2014). *Sustainable infrastructures for life science communication: Workshop summary*. Washington, DC: National Academies Press.

National Research Council. (2015). *Preparing the workforce for digital curation*. Washington, DC: National Academies Press.

National Science Foundation. (2016). ***NSF's 10 big ideas: National Science Foundation report***.

National Science and Technology Council. (2009). ***Harnessing the power of digital data for science and society: Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council***. Washington, DC: NCOO NITRD. .

Olson, J. S., & Olson, G. M. (2014). *Working together apart: Collaboration over the internet*. San Rafael, CA: Morgan & Claypool.

Organisation for Economic Co-operation and Development. (2015). ***Making open science a reality*** (OECD Science, Technology and Industry Policy Paper No. 25). Paris, France: OECD.

Organisation for Economic Co-operation and Development. (2017a). ***Business models for sustainable research data repositories*** (Report No. 47). Paris, France: OECD.

Organisation for Economic Co-operation and Development. (2017b). ***Strengthening the effectiveness and sustainability of international research infrastructures*** (Report No. 48). Paris, France: OED.

Orr, J. E. (1996). *Talking about machines: An ethnography of a modern job*. Ithaca, NY: Cornell University Press.

Palmer, C. L. (2001). *Work at the boundaries of science: Information and the interdisciplinary research process*. Dordrecht, The Netherlands: Springer.

Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., . . . Dierolf, U. (2013). **Making research data repositories visible: The re3data.org registry**. *PLoS One*, *8*(11).

Parker, J. N., Vermeulen, N., & Penders, B. (Eds.). (2010). *Collaboration in the new life sciences*. Farnham, UK: Ashgate.

Parmiggiani, E. (2015). ***Integration by infrastructuring: The case of subsea environmental monitoring in oil and gas offshore operations*** (PhD diss., NTNU, Norway).

Parmiggiani, E. (2017). **This is not a fish: On the scale and politics of infrastructure design studies**. *Computer Supported Cooperative Work*, *26*(1–2), 205–243.

Parsons, M. A., & Berman, F. (2013). The Research Data Alliance: Implementing the technology, practice and connections of a data infrastructure. *Bulletin of the American Society for Information Science and Technology*, *39*(6), 33–36.

Parsons, M. A., Godøy, Ø., LeDrew, E., De Bruin, T. F., Danis, B., Tomlinson, S., & Carlson, D. (2011). A conceptual framework for managing very diverse data for complex, interdisciplinary science. *Journal of Information Science*, *37*(6), 555–569.

Pendleton-Julian A. M., & Brown, J. S. (2018). *Design unbound: Designing for emergence in a white water world. Vol. 1, Designing for emergence*. Cambridge, MA: MIT Press.

Peters, D. P. (2010). Accessible ecology: synthesis of the long, deep, and broad. *Trends in Ecology & Evolution*, *25*(10), 592–601.

Pipek, V., Karasti, H., & Bowker, G. C. (2017). A preface to "Infrastructuring and Collaborative Design." *Computer Supported Cooperative Work*, *26*(1), 1–5.

Pipek, V., & Wulf, V. (2009). Infrastructuring: Toward an integrated perspective on the design and use of information technology. *Journal of the Association for Information Systems*, *10*(5), 447.

Plantin, J.-C. (2018). Data cleaners for pristine datasets, visibility and invisibility of data processors in social science. *Science, Technology and Human Values*, *44*(1), 52–73.

Plantin, J.-C., Lagoze, C., & Edwards, P. N. (2018). Re-integrating scholarly infrastructure: The ambiguous role of data sharing platforms. *Big Data and Society*, *5*(1), doi: 10.1177/2053951718756683.

Pollock, R. (2011, March 31). *Building the (Open) data ecosystem* [OFK Blog post].

Porter, J. H. (2017). Scientific databases for environmental research. In F. Recknagel & W. Michener (Eds.), *Ecological informatics* (pp. 27–53). Berlin, Germany: Springer.

Pryor, G., & Donnelly, M. (2009). Skilling up to do data: Whose role, whose responsibility, whose career? *International Journal of Digital Curation*, *4*(2), 158–170.

Puig de la Bellacasa, M. (2011). Matters of care in technoscience: Assembling neglected things. *Social Studies of Science*, *41*(1), 85–106.

Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. *Science*, *331*(6018), 703–705.

Research Data Alliance. (n.d.a). ***Data foundation and terminology working group***.

Research Data Alliance. (n.d.b). ***Metadata directory of standards***.

Ribes, D., & Baker, K. S. (2007). Modes of social science engagement in community infrastructure design. In C. Steinfield, B. Pentland, M. S. Ackerman, & N. Contractor (Eds.), *Communities and technologies 2007: Proceedings of the Third Communities and Technologies Conference* (pp. 107–130). London, UK: Springer.

Ribes, D., & Finholt, T. A. (2009). The long now of technology infrastructure: Articulating tensions in development. *Journal of the Association for Information Systems*, *10*(5), 375.

Ribes, D., & Lee, C. P. (2010). **Sociotechnical studies of cyberinfrastructure and e-research: Current themes and future trajectories**. *Computer Supported Cooperative Work*, *19*(3), 231–244.

Riley, J. (2010). ***Seeing standards: A visualization of the metadata universe*** [Poster].

Royal Society. (2012). ***Science as an open enterprise: Open data for open science*** (Royal Society Science Policy Centre report No. 02/12). London, UK: Royal Society.

Schiermeier, Q. (2015). Research profiles: A tog of one's own. *Nature*, *526*(7572), 281–283.

Schmidt, K., & Bannon, L. (1992). Taking CSCW seriously: Supporting articulation work. *Computer Supported Cooperative Work*, *1*(1), 7–40.

Scholes, R. J., Mace, G. M., Turner, W., Geller, G. N., Jürgens, N., Larigauderie, A., . . . Mooney, H. A. (2008). Toward a global biodiversity observing system. *Science*, *321*(5892), 1044–1045.

Shaon, A., Giaretta, D., Crompton, S., Conway, E., Matthews, B., Marelli, F., . . . Guarino, R. (2012). **Towards a long-term preservation infrastructure for earth science data**. In *iPRES 2012: 9th International Conference on Preservation of Digital Objects*, 1–5 Oct 2011 (pp. 89–96). Toronto, Canada: Digital Curation Institute, iSchool University of Torontos.

Shapin, S. (1989). The invisible technician. *American Scientist*, *77*(6), 554–563.

Shugart, H. H. (1976). The role of ecological models in long-term ecological studies. In G. E. Likens (Ed.), *Long-term studies in ecology: Approaches and alternatives*. Millbrook, NY: Springer.

Simonsen, J., & Robertson, T. (Eds.). (2013). *Routledge international handbook of participatory design*. New York, NY: Routledge.

Sonnenwald, D. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology*, *41*(1), 643–681.

Star, S. L. (1991). The sociology of the invisible: The primacy of work in the writings of Anselm Strauss. In D. Maines (Ed.), *Social organization and social process: Essays in honor of Anselm Strauss* (pp. 265–283). Hawthorne, NY: Aldine de Gruyter.

Star, S. L. (1999). The ethnography of infrastructure. *American Behavioral Scientist*, *43*(3), 377–391.

Star, S. L., & Bowker, G. C. (1999). *Sorting things out*. Cambridge, MA: MIT Press.

Star, S. L., & Bowker, G. C. (2002). How to infrastructure. In L. A. Lievrouw & S. Livingstone (Eds.), *Handbook of new media: Social shaping and consequences of ICTs*. London, UK: SAGE.

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, *7*(1), 111–134.

Star, S. L., & Strauss, A. (1999). **Layers of silence, arenas of voice: The ecology of visible and invisible work**. *Computer Supported Cooperative Work*, *8*(1), 9–30.

Strauss, A. (1988). The articulation of project work: An organizational process. *Sociological Quarterly*, *29*(2), 163–178.

Suchman, L. (1995). **Making work visible**. *Communications of the ACM*, *38*(9), 56–64.

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., . . . Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS One*, *10*(8), e0134826.

Thompson, C. A. (2015). Building data expertise into research institutions: Preliminary results. In *Proceedings of the Association for Information Science and Technology*, *52*(1), 1–5.

Trigg, R., & Ishimaru, K. (2012). Integrating participatory design into everyday work at the Global Fund for Women. In J. Simonsen & T. Robertson (Eds.), *Routledge international handbook of participatory design* (Vol. 1, pp. 213–234). New York, NY: Routledge.

United Nations. (2014). ***A world that counts: Mobilising the data revolution for sustainable development***. New York, NY: United Nations.

Vertesi, J. (2014). Seamful spaces: Heterogeneous infrastructures in interaction. *Science, Technology and Human Values*, *39*(2), 264–284.

Vertesi, J., & Ribes, D. (2019). *digitalSTS: A field guide for science & technology studies*. Princeton, NY: Princeton University Press.

Vierkant, P., Spier, S., Rücknagel, J., Gundlach, J., Fichtmüller, D., Pampel, H., . . . Scholze, F. (2012). ***Vocabulary for the registration and description of research data repositories***. Potsdam, Germany: German Research Centre for Geosciences.

Waide, R. B., Brunt, J. W., & Servilla, M. S. (2017). Demystifying the landscape of ecological data repositories in the United States. *BioScience*, *67*(12), 1044–1051.

Waide, R. B., & Thomas, M. O. (2013). Long-Term Ecological Research Network. In J. Orcutt (Ed.), *Earth System Monitoring*. New York, NY: Springer.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, UK: Cambridge University Press.

Wenger, E., McDermott, R. A., & Snyder, W. (2002). *Cultivating communities of practice: A guide to managing knowledge*. Boston, MA: Harvard Business Press.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., . . . Bourne, P. E. (2016). **The FAIR guiding principles for scientific data management and stewardship**. *Scientific Data*, *3*(1), 160018.

Wood, J., Andersson, T., Bachem, A., & Best, C. (2010). *Riding the wave: How Europe can gain from the rising tide of scientific data—Final report of the High Level Expert Group on Scientific Data*. Brussels, Belgium: European Commission.

Yarmey, L., & Baker, K. (2013). Towards standardization: A participatory framework for scientific standard-making. *International Journal of Digital Curation*, *8*(1), 157–172.

Zimmerman, A. (2008). New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology, and Human Values*, *33*(5), 631–652.

Zimmerman, A., Bos, N., Olson, J. S., & Olson, G. M. (2009). The promise of data in e-research. In N. W. Jankowski (Ed.), *E-research: Transformation in scholarly practice* (pp. 222–239). New York, NY: Routledge.

Zimmerman, A., & Finholt, T. A. (2007). **Growing an infrastructure: The role of gateway organizations in cultivating new communities of users**. In *Proceedings of the 2007 International ACM conference on supporting group work, Sanibel Island, FL, November 2007* (pp. 239–248). New York, NY: Association of Computing Machinery.

Zimmerman, A., & Nardi, B. (2010). Two approaches to big science: An analysis of LTER and NEON. In J. N. Parker, N. Vermeulen, & B. Penders (Eds.), *Collaboration in the new life sciences* (pp. 65–84). Farnham, UK: Ashgate.

Notes:

(1.) Data observations and measurements in ecology that were initially recorded by hand evolved into more complex arrangements featuring use of spreadsheets on computers (Jones, Schildhauer, Reichman, & Bowers, 2006). Data practices evolved as researchers

were exposed to new data management including data formatting, packaging, steward-ship, and metadata description in the 21st century. The notion of data management evolved as research projects with many research participants aiming to analyze and ex-change data grew alongside the availability of new technologies. Star and Ruhleder (1996) identified social and technical dimensions of a community's communication, data arrange-ments, and information systems support as infrastructure. Over the next decades the con-cept of infrastructure grew to encompass digital systems and collective data efforts as well as the flow of data between locations (Kitchin, 2014; Leonelli, 2013; Parsons & Berman, 2013). The sharing of information and assembly of data moved from centralized servers to web accessible locations followed thereafter by development of databases, complex data systems or platforms, and data repositories (e.g., Nadim, 2016; Plantin, Lagoze, & Edwards, 2018; Waide, Brunt, & Servilla, 2017).

(2.) An overview of some of the data repositories in ecology's data landscape is presented in a table organized by kind of repositories and funding agencies (Waide et al., 2017). These data-aggregating repositories are designated first-order, second-order, aggregator, and super-aggregator. Another set of categories developed to describe the population of U.S. data repositories includes those at federally funded data centers, research centers, national libraries, state and local agencies, thematic repositories, domain repositories, in-stitutional repositories, replication repositories, software repositories, commercial archives, and private archives (Baker & Duerr, 2017). The re3data.org data registry in 2018 showed 1,004 repositories worldwide in the natural sciences, with 46 repositories specific to ecology.

(3.) A persistent identifier system for data objects enables citation of data in a manner similar to the knowledge captured in scholarly publications identified with internationally administered ISBNs. The issuing of digital object identifiers (DOIs) creates stable, unique labels that enable larger-scale availability and findability to human participants as well as to automated system services (Faniel & Zimmerman, 2011). Additional identifier systems that are emerging include an International Geo Sample Number (IGSN) system discussed in 2008 and the Open Researchers and Contributors Identification (Schiermeier, 2015) created in 2012 as a unique tag or identifier system for individual researchers and re-search organizations. More recently, a registry to establish links between research fund-ing grants and published products is under development.

(4.) The term "open" in "open data" "means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)" Open Knowledge Foundation.

Florence Millerand

Université du Québec à Montréal - Département de communication sociale et publique Montréal Quebec, Canada

Karen S. Baker

University of California San Diego