Long live the data! Embedded data management at a long-term ecological research site

Nicole E. Kaplan1,†, Karen S. Baker1,2 and Helena Karasti1,3

1Rangeland Resources and Systems Research Unit, United States Department of Agriculture, Agricultural Research Service, Fort Collins, Colorado 80526 USA
2School of Information Sciences, University of Illinois Urbana Champaign, Champaign, Illinois 61820 USA
3IT University of Copenhagen, Rued Langgaards Vej 7, Copenhagen 2300 Denmark

Citation: Kaplan, N. E., K. S. Baker, and H. Karasti. 2021. Long live the data! Embedded data management at a long-term ecological research site. Ecosphere 12(5):e03493. 10.1002/ecs2.3493

Abstract. Open-access data associated with research efforts depend upon managing, packaging, and preserving data for sharing with collaborators and the public. The U.S. Long-Term Ecological Research (LTER) Network, established in 1980, provides an early example of embedded data management supporting long-term, place-based research and contributes to our understanding of the enactment of open data access within scientific research arenas. Here, we examine collective data activities enabled by embedding data management within the Shortgrass Steppe (SGS) research site. Study of the SGS LTER, a member of the U.S. LTER Network for more than three decades, provided a unique opportunity to investigate data management practices and challenges during the life cycle of a long-term project. It illustrates how a continuous, uninterrupted focus on data management positioned in dynamic interaction with researchers at a site as well as with an active network-wide data management committee can stimulate the growth of both data expertise and data infrastructure. We report on an ethnographic study by a collaborative team of researchers, all having been involved with the LTER network and well-positioned for investigating data management challenges faced during the periods of activation, maturation, and decommissioning of a project at a research site. Termination of the SGS site’s membership in the U.S. LTER Network prompted rethinking about long-term data management. During the decommissioning phase, we document how views on temporality and data management strategies shift from planning for a longitudinal, ongoing site to wrapping up a long-term project. Striving to ensure “long live the data” at the end, novel data arrangements, such as development of a digital legacy project collection, contribute to data stewardship. Lastly, from this study of a long-term research site, we offer five recommendations about data management and describe strategies pertinent to planning for data management and open access for other research projects.

Key words: (digital) legacy project collection; (embedded) data management; data infrastructure; data stewardship; long-term ecological research; open (access) data; research network.

Received 28 August 2020; revised 4 December 2020; accepted 18 December 2020. Corresponding Editor: Debra P. C. Peters.

Copyright: © 2021 The Authors. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
†E-mail: nicole.kaplan@usda.gov

INTRODUCTION

Access to data as a public resource and an institutional asset holds great promise for increasing research opportunities. Preparing data for reuse requires critical attention throughout the process of assembling and documenting data for sharing (Michener and Waide 2008, Johnson
et al. 2010, Peer and Green 2012, Moran et al. 2016, Popkin 2019) as digital research collections that can serve as resources for a variety of audiences (NSB 2005). While the value of data management for long-term ecological research is documented (Michener et al. 1994, Stafford et al. 1994, Sutter et al. 2015, Lamb 2017), Kitchin (2014) points out “we lack detailed case studies of open data projects in action, the assemblages surrounding and shaping them, and the messy, contingent and relational ways in which they unfold.” In this article, we seek to supplement these few studies with an example of a site within a research network that at its launch made an explicit long-term investment in data management.

Research sponsors continue to fund forward-looking ecological research programs that rely on open data (Robertson et al. 2008, Goring et al. 2014, Pickett et al. 2016, Balvanera et al. 2017, Chabbi and Loescher 2017, Mirtl et al. 2018, Angelstam et al. 2019) at a time when approaches to data management, data repositories, and data infrastructure are evolving. Our study of the Shortgrass Steppe Long Term Ecological Research site (SGS LTER) and its three-decade membership in the U.S. LTER Network informs such efforts with an account of the introduction and development of a data management component at a long-term research site. The U.S. LTER Network is a polycentric (Ostrom 2008) or decentralized network (Bocking 1997) with bottom-up governance. Each site works autonomously with site-based funding, leadership, and research aims while also contributing to network-wide efforts as members of the U.S. LTER Network Information Management Committee (IMC) and by sharing data in the LTER Network’s Information System (NIS), an ecological repository preserving U.S. LTER research datasets.

When the SGS LTER site learned about the termination of its long-term funding in 2010, a focus on ongoing data management pivoted to address decommissioning. However, decommissioning was not a well-documented or familiar event for data managers, especially with the limited attention given in the literature to “discontinuation” of data management and digital systems at the end of long-term initiatives (Fincham 2002, Krahe 2012, Jackson and Buyuktur 2014, Stegmaier et al. 2014). SGS LTER data management activities shifted to include developing opportunities that broadened the notion of stewardship to include project legacy. Initiating a partnership with Colorado State University (CSU) Libraries, known for its broad scope of collections accommodating a variety of kinds of materials within their digital repository, enabled design of a digital legacy project collection for preserving the supplementary materials accumulated over decades of research at the SGS LTER site.

**Embedded data management and data infrastructures**

In the age of collaborative research and “Big Data” with the emergence of new approaches to the handling and analysis of digital data, we observe in our study the evolution of embedded data management over three decades. In the U.S. LTER network, a designated role called data manager embeds data management within the site’s scientific research community. Designation of this role ensures a focus on assembly of data from several researchers, an activity that lays the groundwork for open access and collaborative learning, also referred to as mutual learning (Brown and Duguid 1991, Lave and Wenger 1991). Data management at each LTER site is framed by several U.S. LTER arrangements: (1) data management carried out within a collaborative ecological network, (2) a requirement to designate and fund a data manager on a continuing basis, (3) data management supported by the site as a synergistic component of the research team and by the U.S. LTER Network as a member of a data management community of practice (the IMC), and (4) by funding agencies cognizant of ecological complexity and the heterogeneity of data to be managed as well as of the LTER as an approach to enacting data management. Embedded data management in the LTER workplace ushered in new social and technical elements ensuring the evolution of data practices that were in sync with ongoing local research and the growth of data infrastructure.

The SGS LTER site, as is the case for each site within the U.S. LTER Network, provides an example of embedded data management supporting data activities locally. Local collective data management is described as an approach
that “facilitates assembly of data generated by individual researchers with a shared interest and contributes to the well-being of science by developing digital capabilities that support local knowledge work in addition to producing data for unanticipated reuse” (Baker and Karasti 2018). Furthermore, this description characterizes data management as a design-oriented and care-intensive activity responsive to changing research needs, technology options, and other situational factors. Embedded data management specifically involves a designated data management individual or team, engaged with a community of scientists involved in collaborative research investigations. Embedded data management nurtures an intertwined development of knowledge and data infrastructures (Baker and Mayernik 2020). Situated near the location of data generation and research work, embedded data management can support research practices. Development and design of collective data practices, flows, and policies take into account local logistics as well as relations with partners and systems more distant from the data origin (Millerand and Baker 2020). Embedded data management includes the notion of data stewardship. We refer to stewardship of data as a broad concept that involves an ethic of planning with a long view of providing care for data that reach beyond sole ownership to see data resources as a public good (Duer et al. 2004, Karasti et al. 2006, Baker and Yarmey 2009, Hartter et al. 2013, ICPSR 2013, Palmer et al. 2013).

Embedded data management is situated so as to contribute to the growth of data arrangements with the objective of accommodating data and information needs at various “temporal horizons” (Karasti and Baker 2008a, b), where the legacy of research and emergent data types are integrated over a long-term trajectory. Kitchin (2014) describes databases, data systems, and data infrastructures as “complex sociotechnical systems that are embedded within a larger institutional landscape of researchers, institutions, and corporations, while representing essential tools in the production of knowledge, governance, and capital.” Data infrastructures are noted to “host and link databases into a more complex sociotechnical structure.” Further, “data infrastructures do not simply support research, they fundamentally change the practices and organization of research—the questions asked, how they are asked, how they are answered, how the answers are deployed, who is conducting the research and how they operate as researchers.” This sets the stage for understanding the many responsibilities and opportunities for invention that are part and parcel of embedded data management. The multiple aspects of data management and the growth of local data infrastructures, though vital to everyday scientific research today, are frequently under-appreciated (Baker and Karasti 2018).

**Digital data collections and repositories**

Digital data collections and repositories are an integral part of contemporary data infrastructures supporting the sciences. Information technology and digital systems have spurred the advancement of data preservation and online services resulting in the development of digital data collections and repositories. However, agreed upon definitions and classifications of many of these assemblies are still unfolding. Consider two digital assemblies with different origins: (1) collections emanating from the fields associated with libraries and archives and (2) repositories emerging in association with scientific data generation and technology advances. Definitions of data collections and repositories are imprecise and sometimes coincide, which is why these terms are often used interchangeably by scientists. Cragin’s (2009) definition of scientific data collections describes them both as follows: “aggregates of data generated or collected by scientists or instruments which are grouped together because they share some common property.” They are typically characterized in terms of properties including size and structuredness of holdings, breadth of user base, degree of compliance with standards, sustainability, and funding (NSB 2005, Cragin and Shankar 2006). For example, the NSB (2005) identifies three kinds of “long-lived digital data collections” in the sciences—“research,” “resource,” and “reference”—characterizing them as increasing on continuums of the properties above. However, a good deal more variety in collections and repositories exists in practice.

For our investigation, we align with Cragin’s (2009) view of the relationship between collections and repositories: digital repositories as
brining together collections in the sciences. In this study, we encounter and characterize three kinds of data repositories: (1) a “local data repository,” known to SGS LTER participants as a site-based data system, for managing datasets as well as collecting metadata and information from the many SGS LTER researchers (Stafford et al. 2002); (2) a “community data repository,” known within the U.S. LTER Network as the Network Information System (NIS), for aggregating data packages from all the sites within the U.S. LTER Network (Michener et al. 2011, Servilla and Brunt 2011, Servilla et al. 2016); and (3) an “institutional repository,” such as the CSU Libraries digital repository, holding multiple diverse digital collections generated within an institutional setting (Lynch 2003).

These three kinds of repositories allow us to explore the intertwined relations of the SGS LTER “local data repository” for a single site and its submission of data packages to the U.S. LTER NIS, a “community repository.” CSU Libraries’ recent interest in expanding its “institutional repository” to include scientific research data became relevant to the SGS LTER site during its decommissioning, when the embedded data management was wrapping up and reconfiguring digital assets so that the longevity of data would be ensured. A digital collection of research data and materials was created at a time when the ongoing data care provided by LTER-embedded data management ceased to be viable for all the digital assets that needed to be preserved.

Research questions leading to lessons learned and recommendations

Examining the evolution of data management at SGS LTER, two research questions were asked about data activities over time: (1) How has the SGS LTER site formed its data management within the U.S. LTER Network? What were the major challenges that informed development of embedded data management? (2) How did data management change for the decommissioning period? What were the new challenges? To answer these questions, our study investigates embedded data management at a long-term research site within a network and details local data activities when the project comes to an end after three decades. In addition to describing how SGS LTER data management and data infrastructure formed and subsequently changed during the three phases of SGS LTER, we highlight lessons learned from the challenges of evolving data management activities. After analysis of SGS LTER data management strategies and rethinking the long term during decommissioning, we offer recommendations for data management that ensure data stewardship and the longevity of the data.

RESEARCH METHODS, DATA COLLECTION, AND ANALYSIS

We have conducted an ethnographic investigation of data management at the SGS LTER site. We chose an ethnographic approach to be able to ground our investigation on people’s experiences working collectively at a long-term ecological research site. We have used fieldwork methods of participant observation, interviewing, and collecting of primary source materials. Thus, data collected for this study included transcribed interviews, fieldnotes of interactions with participants, site’s historical documents, and materials from our personal archives. During decommissioning, our empirical activities offered opportunities for study participants and ourselves to recollect, reflect upon, and learn together about the SGS LTER site before experiences and details faded from memory. Though the study was conducted during this period, our data collection and analysis focused both on the long-term development of data management at the SGS LTER site and the transition in data management at its termination.

The authors have all been associated with LTER in differing capacities and their own lived experiences in LTER expanded collective observational capacity to conduct interviews as a reflective process with frequent dialogue among researchers (Campbell et al. 2016). Kaplan and Karasti conducted 33 interviews from fall 2012 through spring 2013, and Baker conducted 29 interviews from winter 2013 to summer 2015, with participants from a variety of roles including ecologists, staff, administrators, and other stakeholders who were or had been actively involved with SGS LTER. Interviews ranged from 45 min to 2 h with an average length of approximately an hour and a
half. Kaplan and Baker also met 15 times as members of a data migration working group during decommissioning (see transformation, phase 3) with each session lasting approximately 2 h. Study participants gave informed consent to be interviewed and their accounts included in this study either via Colorado State University Institutional Research Board (IRB) Protocol 12-3510H or University of Illinois Urbana-Champaign IRB #13595.

The authors set protocols for semi-structured interviews with open-ended questions constructed around major professional, site, and data management themes to guide those interviewed in sharing their reflections on the trajectory of careers and activities at the SGS LTER site. Interview questions were designed to evoke recollections of interviewees (Rubin and Rubin 2012) about the evolution of the site, the complexities, and challenges of managing research and data activities collectively over time, and issues related to decommissioning. As we progressed in gathering various ethnographic materials and documents of the SGS LTER site, we created synthetic overviews to prompt and assist interviewees in recalling events and activities over the past 30 yr (Ritchie and Lewis 2004). These overviews included summary tables, timelines, diagrams, and other graphics to illustrate, for example, the progression of leadership, the development of science themes, and the evolution of conceptual frameworks from the site’s six proposals granted during the SGS LTER life cycle. These visual devices were integrative, serving to elicit further discussion.

Interviews were audio recorded and transcribed for analysis. First steps of analysis were conducted separately by the authors, reading and rereading transcripts, thinking through the dialogue experienced in the moment with the study participants, and referring to our field notes (Rossman and Rallis 2011). Individually, we performed open coding of the transcripts to inductively identify themes and concepts related to our research interest. In joint sessions, we negotiated differences in interpretations and reconciled coding. We proceeded to focus coding with agreement among authors on a set of themes (Emerson et al. 2011), three of which are of relevance for this article: practices of and challenges related to data sharing, embedded data management, and data infrastructure.

We used triangulation as a method both to converge materials from multiple fieldwork methods and for comparative analysis in order to develop a more comprehensive understanding of the empirical phenomenon (Lindlof and Taylor 2011). We aimed to be able to both describe the situated phenomenon and identify factors related to research themes, for example, to identify the local troubles that represented examples of larger data management issues (Millerand et al. 2013). To guide the analysis and presentation of results of the SGS LTER trajectory of data management and infrastructure, we draw on the framework approach of Imperial et al. (2016) that identifies phases of network governance. We followed the themes of data sharing, embedded data management, and data infrastructure across the thirty-two years of the SGS LTER site life cycle through phases from activation to maturation and then decommissioning, also known as termination or as a period of transformation.

In focusing on SGS LTER, we utilized the U.S. LTER Network of 26 sites existing at the time for comparison. It was possible to relate the SGS LTER findings with the data management practices, arrangements, and approaches existing at the other sites. Given their similarities in size, scope, and data policies, the set of sites also made it possible to recognize site-specific idiosyncrasies. We were able to reason with knowledge of how data management evolved in these many cases, often with different timing and with different local circumstances or logistics. The range of this variety helped put SGS data activities into perspective and permitted identification of cross-site similarities and differences in data management. Perspectives rooted in concepts of data care as well as the relevance of mutual learning (Baker and Karasti 2018) were used to frame analysis of the many facets of data work as well as the dynamics at the intersection of science, technology, and data management.

**THE FORMATION OF SGS LTER DATA MANAGEMENT AND DATA INFRASTRUCTURE**

This section will describe how data management and data infrastructure were formed at the SGS LTER site during its three-decade membership
in the U.S. LTER Network. It was funded by the National Science Foundation (NSF) through Colorado State University (CSU), a federally funded land grant university established in 1870 (NRC 1995) with a broad mission to offer any citizen education in subjects including practical agriculture and engineering. From its start, the SGS LTER sampling was situated at the nearby Central Plains Experimental Range (CPER). The CPER, designated as an Agricultural Research Service (ARS) experimental range by the U.S. Department of Agriculture (USDA) in 1937, has hosted scientists working on a variety of research efforts funded by different agencies. A large networked research program launched in 1964 as part of the International Biome Program (IBP) provided early experience with data management at CSU and CPER. CPER became the core research location for the SGS LTER site in 1982.

Each member site of the LTER Network focuses on study of a biome, an ecological system at a particular “place” (Billick and Price 2010, Kingsland 2010, McNulty et al. 2017). The LTER Network formation was informed and shaped by breakthroughs in the ecological sciences (Worster 1997) and in digital technologies (Nielsen 2011) as well as by experiences in the IBP (1964–1974), a program that influenced the LTER approach to data management (Golley 1993, Aronova et al. 2010, Coleman 2010). In contrast to centralized data services for distributed data sources during IBP, each LTER site became responsible for designing and administering local data arrangements while participating in LTER Network data activities. Though each place-based site has unique features, all sites shared the same focus on ecological research themes and supported a designated data manager, as required by the NSF proposal call. As a result, embedded data management became a recognized component at each LTER site drawing scientists and data managers together to work collaboratively over time. Data management early on established a convention of using a locally developed template for defining variable definitions and units in a tab delimited format that subsequently added to the completeness of metadata for each SGS LTER dataset.

A brief overview of embedded data management at SGS LTER site is given in Table 1, where features of data management, technology, and data support are described. Embedded data management evolved through the phases from a part-time position to a full-time position, a trend that occurred across the U.S. LTER Network of sites. Also listed are some of the key activities tightly coupled with the LTER Network through each site’s participation in the all-site LTER IMC, a committee that prompts network-wide discussion and coordination of data issues by the site-based information management teams. The IMC is a standing committee of the LTER Network governance framework that helps guide information management direction and vision for the U.S. LTER Network.

**Activation (Phase 1, 1982–1995)**

At the SGS LTER site, a data management component was established at the launch of the site as required. This component was co-located in the lead investigator’s department at the CSU, close to the CPER field sampling and to the interdisciplinary researchers studying the grassland biome. Data management led data activities such as creating a field guide to document sampling programs as well as assisting in data processes including its cleaning and assembly of data on a shared server. The embedded data manager worked directly with researchers in designing data management processes, documenting datasets, and facilitating data organization among the SGS LTER investigators. Data management early on established a convention of using a locally developed template for defining variable definitions and units in a tab delimited format that subsequently added to the completeness of metadata for each SGS LTER dataset.

Within the first decade of SGS LTER, a balancing was evident in needing to address both site and network data activities. A professional
collegiality within the network-wide IMC emerged through collective discussions of site approaches, data issues, and potential cross-site activities. Embedded data managers responded to the constant pressure on sites to perform at the highest level, motivated by the risk of being perceived as the recipient of an unwarranted privilege of long-term funding. Furthermore, pressure arose both from the potential for comparison with other sites and from concerns of researchers regarding the possibility that data management objectives could distract from research activities. At the SGS LTER site, available technology support included limited access to a department's Unix server, while data processes were co-designed by the embedded data manager together with researchers then conveyed back to technology support staff within the department. Across the U.S. LTER Network, continuity in care of data and metadata was recognized initially as a need to identify, describe, assemble, and catalog time-series research datasets. At each site, local data systems developed incrementally over time for sharing data among local participants. These site-based systems stimulated assembly of the data from many researchers and promoted development of local conventions, provided initial experience with data sharing within a safe environment, and became the foundation for making data available to other repositories with wider audiences.

**Maturation (Phase 2, 1996–2009)**

With the responsibilities of embedded data management increasing during this period, the data position was renamed “information management” in the U.S. LTER Network (Baker et al. 2000). By 2008, the role was funded full-time at the SGS LTER site. After the first three proposal cycles of funding for SGS in the first phase, LTER grants continued into the second or maturation phase to provide support for

Fig. 1. Evolution of data management goals in SGS LTER (1982–2014) with embedded data management shown in a three-phase framework. SGS LTER, Shortgrass Steppe Long Term Ecological Research site.
Table 1. Salient features of embedded data management for each of three phases of the SGS LTER site cycle.

| Project phases | Embedded data management focus | Technology support | Data support | Network data activities |
|----------------|--------------------------------|--------------------|--------------|------------------------|
| Phase 1        | 1/2-time data manager coordinated local fieldwork and data | Department-supported Unix server with limited access | Created a site field guide and protocols; adopted use of a variable definition template for each dataset included in the local data system’s database | Joined and participated in LTER Information Management Committee (IMC) |
| Phase 2        | Increasing support to full time in 2008 for data/information manager who collaborated with local team members and with U.S. LTER Network IMC | CSU-supported Web site with online searchable database; moved to Department-supported Microsoft server with site-managed database | Assembled datasets into relational database; prototyped submission of data packages to U.S. LTER Network Information System (NIS) | Adopted U.S. LTER Network metadata standard (EML), unit registry, and controlled vocabulary of keywords; continuing activity with LTER data sharing policy; and development of IMC bylaws |
| Phase 3        | 3/4-time information manager led closing data efforts | SGS-deprecated CSU-supported Web site and site-managed database; digital legacy project collection curated by CSU Libraries institutional repository | Partnered with CSU Libraries on data migration working group to create a digital collection; provided public and machine access to tabular datasets within legacy project collection | Published data packages to the LTER NIS with most current version of EML and updated validation requirements; became non-member in U.S. LTER Network; published a report on managing, transforming, and migrating data at termination |

Notes: CSU, Colorado State University; SGS LTER, Shortgrass Steppe Long Term Ecological Research site.

collaborative science, field sampling, embedded data management, and data sharing. This contributed to the growth of the research effort as the site attracted new collaborators. The number of participating scientists in SGS LTER increased from seven in the mid-1990s to 28 by 2007. One of the participating scientists describes the unique stability and intellectual resources of the SGS LTER research environment during this phase (Quote 1).

Quote 1. A research scientist reflecting on LTER science:

"It was this big picture, large spatial scale lens through which I started to see the Shortgrass Steppe. Those are themes that I really kept carrying throughout my work on the Shortgrass Steppe - spatial patterns as representing what controls ... ecosystem processes, long term data and manipulations, and big collaborations that bring together people who have different tool sets and different disciplinary approaches to their science. I would have never had all those things had I not gone to CSU. It was an amazing place to be dropped into as a scientist. But without the stability of the LTER, none of those people would have been there."

The need for robust online access to data and information required institutional resources from CSU for hosting a SGS LTER Web site. The Web site provided information about the site and a catalog of the site’s datasets that was a critical first step to making data visible and easily available. SGS LTER worked with the IMC to explore, provide feedback on, and adopt what became the Ecological Metadata Standard (EML) (Fegraus et al. 2005, Millerand and Bowker 2009). Metadata using local conventions were mapped to EML and bundled with datasets to become data packages in formats accepted by the Network Information System (NIS) (O’Brien et al. 2016, Servilla et al. 2016). Ongoing effort was required at each site to ensure continuing interoperability of local and network metadata and systems due to continuing refinement of conventions and standards (Yarmey and Baker 2013). In addition, SGS LTER worked with the IMC on network-level efforts such as a LTER unit registry (Karasti et al. 2010), a controlled vocabulary of keywords (Porter 2019), and an agreed upon U.S. LTER Network data sharing policy.
LTER SGS data practices:
in taking care of data (Quote 2).
agers describes the educational role they played
scientist Millerand and Baker 2011). Further, a research
2002, Benson et al. 2006, Michener et al. 2011,
soll et al. 1997, Baker et al. 2000, Stafford et al.
Inger
LTER data efforts by embedded data managers
limited resources available. Publications about
proposals became more extensive. The plans pro-
vided a crucial prompt to not only scrutinize but
also to articulate data management arrange-
ments. They created specific moments for
addressing scientific and data management
expectations and for negotiating needs given the
limited resources available. Publications about
LTER data efforts by embedded data managers
at sites began to appear in the literature (Inger-
soll et al. 1997, Baker et al. 2000, Stafford et al.
2002, Benson et al. 2006, Michener et al. 2011,
Millerand and Baker 2011). Further, a research
scientist’s reflection on the embedded data man-
gers describes the educational role they played
in taking care of data (Quote 2).

Quote 2. A research scientist reflecting on
LTER SGS data practices:
I think for me the data part, it’s been one of the really
valuable things that I’ve learned from being in the LTER.
That really helps in terms of understanding how to write
a good protocol, how to make a good datasheet. When you
actually understand the kind of problems that come with
the analysis and the data archiving and the writing the
metadata, you understand what the data has to look like
at that end. It really changes how you collect the data and
how you organize the data on the front end when you’re
actually collecting it. Thinking about the data stream
from field to analysis.

The availability of data managers to consult
with local scientists helps when needs associated
with data practices and data infrastructure liter-
acy arise (Hampton et al. 2017, Gray et al. 2018)
and in situations where open access and existing
data infrastructure are not enough to make data
easy to use (McNutt et al. 2016, Pasquetto et al.
2017). Local data management components
resulted in raising awareness of LTER particip-
ants about data issues and fostered familiarity
with the concepts and vocabulary needed for dis-
cussing and designing data infrastructure. As
researchers gain experience in managing data
collectively at long-term ecological study sites,
various facets of working with shared data begin
to be imagined and new expectations, practices,
and skills develop (Reichman et al. 2011, Hamp-
ton et al. 2013, Peters et al. 2014, Cheruvelil and
Soranno 2018).

Transformation (Phase 3, 2010–2014)
A period of uncertainty for SGS LTER began
in 2008 when after success with previous proposals,
its sixth proposal failed, and the site was put on
probation. U.S. LTER Network sites are funded
today by six-year renewable grants. Renewal
proposals are evaluated and designated either as
accepted for another six-year period of funding
or as accepted for a probation of two years. Prob-
bation is a time during which a site is expected to
address concerns raised by reviewers, to consider
ways to reorganize, and to write a new four-year
proposal. After failing its re-evaluation, a two-
year decommissioning period began at the end
of 2010. As the site was shut down, so were the
social and technical supports for managing data.

Following the LTER custom of learning from
the experience of other sites, the SGS LTER data
team revisited the closing of the three LTER sites
terminated prior to its decommissioning: North
Inlet Estuary in South Carolina (1980–1993),
Okefenokee Swamp in Florida (1981–1988), and
Illinois Rivers in Illinois (1981–1988). Unfortu-
nately, the way these early LTER sites ended
was not well documented so the ramifications
of their termination remain little known. U.S. LTER
Network protocols had been developed for
improved site management spurred by the clos-
ing of these sites. These protocols included
development of a probation period and mid-
term review team visits for each site, which pro-
vided sites with opportunities for reflection and
feedback. This contributed to no sites being dis-
continued for two decades until the SGS LTER
site was terminated.

Since the SGS LTER site did not have a pre-ex-
isting plan for site and data management closure,
developing, and submitting to NSF, a decommis-
sioning plan for data activities outlined in a bud-
get justification document was a critical first step
that ensured the third phase was a period of
transformation. SGS LTER’s list of data activities
during decommissioning is contained in Box 1,
which provides examples that address the long-
evity of data.
Box 1. Decommissioning budget justification of data activities as a road map for closure

1. Meet minimum data requirements of funding agency such as creating a dataset inventory and submitting data packages for off-site preservation in the LTER NIS, a network repository
2. Improve access and quality as well as increase quantity of data packages such as by delivering updated metadata in response to most up-to-date U.S. LTER Network standards that ensure congruence of metadata in the network repository
3. Preserve project legacy such as by identifying, assembling, and preserving site materials that provide context for the research datasets
4. Document data management during decommissioning by publishing an account of data activities together with the budget justification in an institutional report or research journal article

During decommissioning, new approaches and a budget were needed to create a viable data management strategy that could replace the existing data management approach and secure the longevity of all the SGS LTER data assets. Although site-based researchers expressed confidence that the embedded data manager would share data and metadata in accordance with established U.S. LTER Network requirements, there were concerns about the fate of the local data system because SGS LTER stopped allocating grant funds for replacing or updating services, technology, and applications. In addition, many materials surfaced during decommissioning including items initially stored on shelves in local offices as hard copies and subsequently digitized. Because the “long-term” or a “continuing” mindset familiar to those in the LTER program came to an end, data management had to devise ways of making certain all the site’s data assets would continue to have a persistent life beyond the life cycle of the SGS LTER.

Prompted by a sense of stewardship and professional curiosity, SGS LTER data management transformed from thinking of SGS as an LTER site, developing instead a view of what constituted digital data assets for a long-term project. This augmented earlier data efforts, moving from seeing a collection of data packages as the site’s core data responsibility to taking account of supplemental artifacts, a wide variety of materials that provide context for the hundreds of years of cumulative endeavors by researchers and staff at the site. Curation of this diversity of materials required developing new partnerships.

Situated at CSU, SGS LTER data management turned to CSU Libraries, a local institution that had provided help with digitization of some of the site’s artifacts. At the time of the SGS LTER decommissioning, CSU Libraries had interest in expanding the libraries’ role by mobilizing support for research data management across the campus. Inquiries made by SGS LTER-embedded data management about the library’s digital collections, opened discussion of new options for preservation of research data. Mutual interests motivated formation of a working group for data migration that met regularly from October 2013 to December 2014 to carry out a pilot study on moving project-related research data and materials to an institutional repository at a time the library was envisioning development of data management services (Lynch 2008, Tenopir et al. 2012). The design of data migration and of a research data collection were informed by the cumulative SGS LTER-embedded data management experience and the library experience as an institutional repository. In working with library partners, SGS LTER developed an understanding of collections as loosely structured assemblies of digital materials. The conceptualization of what would eventually be referred to as a digital legacy project collection represented an approach to organizing datasets and their metadata with links to supplemental SGS LTER materials.

Data management challenges and SGS LTER responses

There were challenges addressed by embedded data management in all three phases of the SGS LTER life cycle. Those discussed above are summarized in Table 2 together with the SGS LTER responses to these challenges. The need to balance responses to both science needs and data needs heads the list. Next, the term design is introduced to underscore the planning and
Table 2. Five challenges addressed by embedded data management listed together with examples of SGS LTER responses.

| Data management challenges                                                                 | SGS LTER responses                                                                                   |
|-------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|
| 1. Supporting science and data management needs locally as they evolve                     | Surveyed data practices, articulated data needs, planned data flows,                                 |
|                                                                                           | designed data systems, and negotiated data arrangements                                               |
| 2. Designing and maintaining a local data repository                                       | Developed local metadata conventions, data system, and Web site with a catalog of site datasets       |
| 3. Packaging datasets with useful metadata                                                 | Posted metadata on the Web site and enabled download of data dynamically                               |
| 4. Meeting expectations of open-access data and interoperability                           | Mapped local metadata to U.S. LTER Network standards to submit machine-readable data packages to external repositories |
| 5. Prioritizing a growing professionalism in data management                              | Continued learning on-the-job experientially and pursued opportunities for communication about data management |

Note: SGS LTER, Shortgrass Steppe Long Term Ecological Research site.

...ingenuity required in managing data and information within a site's constraints. The term “local data repository” is introduced to refer to both the local data system and Web site that became important elements supporting communication associated with data and information. These are followed by strategic planning to become part of the larger data landscape by addressing data packaging, access, and interoperability. The final challenge points out the evolution of the role of data management.

RETHINKING THE LONG TERM IN DATA MANAGEMENT

A profound shift occurred in the understanding of “long term” during SGS LTER’s decommissioning phase. This unexpectedly involved a significant shift from thinking about SGS LTER as a “continuing site” to reasoning about it as a long-term “project” with a termination date. Thinking about the long term for data with an end date in mind required a different orientation to “long term.” While continuing to participate in the U.S. LTER Network activities, the ever-evolving design of the SGS LTER site’s local data infrastructure was ended and the local data system was abandoned during this period in the SGS LTER life cycle. A thoughtful approach to the termination event created a critical time of transformation during which persistent open access to the site’s digital data within a comprehensive “legacy project collection” was formulated together with a plan for migration of all SGS LTER digital assets to an institutional repository. It was the continuing attention to data arrangements by embedded data management in the first phases of the SGS LTER that enabled creation of a proactive approach to decommissioning. Five SGS data management strategies made evident during decommissioning are discussed below.

Invest in data management locally

The continuous, uninterrupted focus on data management by the LTER sites and their network stimulated the growth of data expertise, data practices, and data infrastructure. Data management as an embedded component at an LTER research site ensured that data were tended collectively in a coherent and cogent manner throughout the phases. Local conventions were developed, metadata templates designed, and the flow of data co-designed with an end goal in mind. Addressing fieldwork documentation, metadata descriptions, Web site design, and data system development via incremental design resulted in improvements in data practices and processes that contributed to improving SGS LTER data organization, findability, and access over the decades.

The SGS LTER expanded the lead data management role from part time to full time while drawing in and engaging other personnel at the site including field technicians, administrative staff, students, and data and technical specialists in other groups at CSU as part of a team familiar with and contributing to managing the SGS LTER data. The team helped in maintaining field manuals documenting sampling procedures for ongoing projects, in entering data and metadata into the data system, and in establishing a...
dynamic Web site that delivered data and metadata online.

Embedded data managers work in concert with technologists, data providers, and staff by supporting “consultation, mediation, advocacy, integration, synthesis, translation, and mutual learning” (Baker and Chandler 2008), which contributes both to managing data in the present and to planning for its future use. For a long-term site, embedded data management sustained collective data management efforts, ensuring accomplishments were planned and achieved over time (Karasti et al. 2010). Such a data component is positioned to facilitate and readjust relationships between people, facilities, data sources, technologies, and data services (Baker and Millerand 2007, 2010, Baker and Karasti 2018) while negotiating between site data and U.S. LTER Network needs in addition to new data mandates.

Collaborate and learn within your community

The IMC had continuous support and an active role in designing and enacting data arrangements at both site and network levels. Consequently, the term “community” is used here to refer not only to the SGS LTER site participants but also includes the U.S. LTER Network participants. A number of strategies were employed by SGS LTER for communicating and learning collaboratively.

Two LTER data management communication dynamics.—Two critical communication dynamics shape an LTER site’s data management. First, a site-network data management dynamic ensures ongoing interactions between each site’s data management team with the other site teams to envision and enact network-wide data management approaches. Second, a science-data management dynamic is created by continuous interactions between research science and data management participants. These interactions result in frequent reviews of data practices buttressed by the social dynamics of collaborative learning.

The IMC underpins the site-network data management dynamic. Ongoing collaboration of site data managers creates a unique opportunity for discussion, planning, and prototyping activities that is key to broadening site horizons through sharing of local experiences and ideas for network-wide actions. The IMC propels ongoing communications—monthly video teleconference calls, working group planning, and face-to-face annual meetings—among the data management teams at LTER member sites. Participation in working groups as well as multi-site prototyping efforts contributes to awareness and development of network-wide data insights and policies. A common practice among embedded data managers is to draw insights from membership in the U.S. Network “where each site is a ‘laboratory’ with its local specificities” (Karasti and Baker 2004). Each LTER site benefits from and is shaped by this interactive forum that stimulates learning while presenting new options and strategies. The diversity of U.S. LTER Network sites requires constant attention to differences which in turn fosters insight into data issues, design ramifications, and data infrastructure options. The site-network data management dynamic cultivates breadth of understanding, attention to differences, and mindfulness of design ramifications.

The second communication dynamic, a science-data management interaction, arises at the site level and is anchored by the everyday interplay that occurs at LTER sites in planning and carrying out data activities. The embedded data manager works directly with those generating the data, activating collaborative learning, and joint priority setting for continued design-in-use (Henderson and Kyng 1991). Due to communication and collaboration with other sites as well as their understanding of changing scientific practices including data practices, SGS LTER-embedded data management was able to provide feedback on ways to work with data (Quote 3).

Quote 3. A research scientist describing SGS LTER help with changing data practices:

“There have been times when we’ll be working on something, and [the data manager] suggests why aren’t you doing it this way. This is the way they do it at so and so other place. That gives me a fresh look. Well, this is just the way I’ve always done it. Sometimes just having that new insight, I’ll learn new things that will improve the way that I do it [work with data]. I think there’s that value about the network of data managers bringing that information back to their site. There clearly is a feed-back there. 
There are plans, processes, and systems needed in order to document, aggregate, store, and share data (Baker and Karasti 2004, Karasti and Baker 2008b) at an LTER site. Tending to communication supports its evolution, so that engagement can broaden, and feedback channels develop. For synthesis and data accessibility within a distributed research network, it was recommended: “Develop and maintain transparency by fostering communication and feedback” (Laney et al. 2013). Though data activities represent only one component at a site, the science-data management interactions are an important part of keeping data issues integrated into a site’s scientific efforts. It is an opportunity for data managers to explore how to minimize disruption of existing practices and resources as well as to highlight a judicious approach to introduction of new data practices and technologies (Karasti and Baker 2004). Such an approach is typically tailored to local needs and modest in terms of technical hardware and software.

The science-data management dynamic also evolves across the U.S. LTER Network, taking place between the LTER science leadership groups and the IMC. A primary channel of communication was through a non-voting representative from the IMC serving as a consultant on network-level data management as well as a liaison between network science and site-based data management. For instance, the IMC representative brought proposals for the development of modules in the NIS and for the adoption of a network-wide metadata standard. In addition, in response to a request from leadership, a working group of the IMC developed criteria for data management as well as for Web sites in order to facilitate the review of sites at the end of their six-year proposal cycles.

Collaborative learning.—The collaborative approach ingrained as part of the LTER culture stimulates learning across the broader research network. The sites and the network feature positive group processes such as hearing a diversity of perspectives and reaching common ground. Data management and research participants share site-based understandings of local history as well as current research efforts while data management participants “have in common concerns with data work and with care of scientific data” (Baker et al. 2011). Shared experiences generated by the site and network dynamics described above are examples of learning opportunities enabled by LTER-embedded data management. A lively learning environment, existing throughout the SGS LTER life cycle, facilitated alignment of researcher and data manager expectations as well as coordination of local conventions with highly structured U.S. LTER Network standards. Collaborative learning by data management as well as researchers at the SGS LTER site occurred through discussions of data issues that accompanied the intertwined planning, negotiating, and envisioning of research and the data work supporting collaborative research. (Quote 4).

Quote 4. A research scientist reflecting on LTER SGS data practices:

I’m much more aware in looking at long term data of just all the sorts of problems you can have in the field and how those if they don’t get fixed early they just blend into the background and can create all kinds of problems. … It makes you think a lot more about the quality and about fixing the issues and the huge need to be entering the data soon after collection. Those are all things that are just fundamental in my data practice now. I make my students, even putting data into excel, clean it up and do that all really correctly and not let bad data sit around. … A big project with lots of different data coming in, different shapes, different types of data, you have to have a person just working with the data, like in LTER.

LTER is an example of a bottom-up networked community able to integrate an informal but effective collaborative learning environment into a scientific research workplace. This expands the notion of collaborative learning (Voss et al. 2009) beyond a focus on technical arrangements, data skills, and computational capacities to include design and data management as well as data care (Baker and Karasti 2018). Given experience with the dynamic social interactions required for the development of a research site with a long-term mindset, the LTER Network has established “the kinds of social networks, infrastructures, and situated knowledge” that have been identified as critical to “technology comprehension” (Smith et al. 2020).

The LTER Network has created a familiarity with data infrastructures “as distributed accomplishments constituted by an evolving set of relationships between people and devices, software
and standards, words and instruments” described recently as “data infrastructure literacy” (Gray et al. 2018), a concept intended “to make space for collective inquiry, experimentation, imagination and intervention.” The establishment and growth of U.S. LTER Network data management, with dynamic community interactions and each site’s local data repository, enabled learning that has contributed to local knowledge as well as improvements in data sharing practices. Embedded data management has functioned as both a service and an intervention to traditional scientific data practices of individual researchers. It initiated a redistribution of data responsibilities and contributed to both site and network digital data capacity building.

**Address data issues proactively and creatively**

Rather than seeing data issues simply as problems to overcome, a data manager recognizes them as opportunities for creative design and learning. Three of the SGS LTER strategies for addressing data issues proactively and creatively are discussed below.

*Create a site Web site.*—Through the interaction of researchers and data management, a site Web site served as a familiar forum for sharing research priorities and historical documents as well as the ongoing activities and research findings at the site. The SGS LTER digital assembly of supplementary materials was enabled by the development of the Internet and the custom of research initiatives creating web pages. The SGS LTER data management provided continuing support for the Web site with a recognition of both its identity-building capability and the Web site’s effectiveness as a communication device for the site’s distributed stakeholders. The Web site was initiated by local participants, both those generating and those using the information. Initially, this digital forum was to provide an overview of research while summarizing ongoing activities and their background. Published content provided research context including overarching objectives and experimental designs as well as information about sampling methods. During decommissioning, data management recognized the Web site first as a shared conceptual space where socially negotiated knowledge is captured (Roschelle and Teasley 1995) and eventually as a pre-archive, collection-building resource. Many of the Web site materials would have been lost, if they had not been first identified and assembled on the Web site and subsequently migrated as a collection to an institutional repository.

*Embed data management to enable local responsiveness.*—As part of everyday activities at the site, embedded data management created an opportunity for the development of local data expertise at SGS LTER. This included having data management immersed in local activities acquiring familiarity with site research as well as with local data practices, concerns, and logistics. A locally designated data manager was able to constantly gather information about site needs and existing resources so able to plan for both longer term objectives and those requiring immediate action. Continuous engagement with researchers both as data generators and as data users provided opportunities for data discussions and collaborative learning. Interacting both formally and informally with research team members, data work could be carried out with forethought and incorporation of feedback, yet plans could be altered as needed in response to changing circumstances.

The community environment created frequent occasions for learning-by-doing. For instance, for SGS LTER, documentation of fieldwork activities as well as description of datasets informed design of local data management processes. Work on the packaging of data and metadata laid the groundwork for discussion of machine-readable tabular data and facilitated open access to data. These activities exposed researchers as well as partners at CSU Libraries to the social and technical aspects associated with data sharing. With partners at the IMC and LTER NIS, SGS LTER contributed to operationalizing open access to data through participation in the development, use, and revisions of the Ecological Metadata Language (EML), a community metadata standard required for submission of data packages to the LTER NIS. The experience gained from leading local data activities and participating in network-wide data management activities was confidence building, providing the insight and competence needed to formulate and carry out a response to the SGS LTER termination.

*Assemble a digital “legacy project collection”.*—SGS LTER, a site that provided access to datasets
on their Web site as well as through the LTER NIS, worked during decommissioning with CSU Libraries to preserve data and materials as a digital collection and an institutional asset. Data management at the site developed a broad view of what constituted site data, augmenting their data efforts from production of a collection of data packages, that is, tabular datasets described with EML metadata. By planning to manage and make available Web site content as well as recently digitized materials that contributed to the story of how, where, and why datasets were collected, an understanding emerged of a collection as an assembly that “houses a greater institutional memory” (Lamb 2017). Such a collection represents a valuable and comprehensive data product, providing additional contextualization that can inform interpretation during reuse of datasets by audiences unfamiliar with the field site.

Supplemental materials provide context for the hundreds of years of cumulative endeavors by researchers. They included historical photographs, species lists, annual field crew manuals, field protocols, pasture treatment maps, field data sheets, reports, and project proposals (Fig. 2). These materials inform the complex process of assessing the appropriateness of datasets for reuse in specific situations (Yoon 2017) particularly for those unfamiliar with the circumstances of data generation. In the end, the SGS LTER collection was created by data managers, researchers, technicians, and academic librarians working to deliver datasets machine-to-machine as well as to provide easy access to variety of materials in one location for browsing a broad, site-generated source of information about SGS LTER. The legacy project collection co-designed with local knowledge of an embedded data manager and librarians was aligned with FAIR principles to ensure data would be findable, accessible, interoperable, and reusable (Wilkinson et al. 2016). Though many of the physical artifacts digitized were then discarded, CSU Libraries archival book storage curates originals of some of the printed materials. Physical soil and plant samples, however, were stored locally in old dormitory buildings at the CPER.

A report, documenting the collection formation process, made it visible both to the LTER and to the other communities (Kaplan et al. 2014). Furthermore, the internal communication of the data migration working group was markedly improved by the identification and visualization of migration tasks at hand. Frequently, the data manager and information scientist thought they understood the material presented by the librarian members of the working group, and vice versa. However, some text or a figure that was first created for sharing with the group and then for inclusion in the report (Kaplan et al. 2014) would often reveal differing understandings or interpretations that led to critical clarifications. Further, informed by the report about partnering with CSU Libraries for the formation of a legacy project collection, other LTER sites terminated after LTER SGS inquired about collection-making that led to partnering with their local libraries (e.g., the Coweta LTER with the University of Georgia’s Athenaeum and the University of New Mexico digital repository contains data from the first decades of Sevilleta LTER prior to 2016, at which time the site experienced a termination followed by a subsequent successful LTER proposal).

The SGS LTER presents a novel example of a long-term data collection that does not fit neatly into kinds of collections described in the literature. The NSB (2005) report that uses the term data to “refer to any information that can be stored in digital form” defines—as already mentioned in the beginning—three functional categories for digital data collections—research, resource, and reference—that in turn loosely characterize collections as having project, community, and global users, respectively. In addition to the audience, the categories are distinguished by the scope and diversity of data, the extent of conformance to standards, as well as to the intent and funding support for preserving the collection. The SGS LTER data collection featured characteristics from all three of these categories, though the fits changed as the site matured over time. For instance, the SGS LTER local data repository could be considered a “research collection” with its assembly of digital materials together with its datasets initially made available via the site Web site for the purpose of communication among project participants. The datasets could be accessed by the public, and the structure frequently conformed to conventions that established an expected format for local
users. On the one hand, with the adoption of EML and availability of data from all LTER sites through the LTER NIS, the SGS LTER dataset collection could be considered an ecological community dataset collection. In the NIS, the SGS LTER dataset collection is available in a recognized ecological repository together with other ecological data collections conforming to the same community standards. In this system, users can query across the diversity of biomes, data types, and geographic settings for one or more of the collections or variables. On the other hand, it could one day be considered a resource dataset collection with the data available and useful to researchers in many disciplines. The SGS LTER digital legacy project collection, however, resides in an institutional repository that includes both highly structured datasets and metadata along with loosely coordinated supplemental materials. The classification for such an arrangement is unclear using the NSB data collection categories. The aim to classify collections and their repositories with their differing trajectories requires further investigation. To begin to add to our understanding of the variety involved, Cragin and Shankar (2006) have explored additional properties of collections pertinent to the three kinds of data repositories encountered in this study including the complexity of interdependencies, authority structures, institutional reward structures, and the transitional nature of collections.
Consider relationships with those outside your community

SGS LTER partnered with others within their institution to consider interoperability of repositories to facilitate the flow of data and information online. Social and institutional roles and relations between partners were also taken into consideration.

Consider strategic partnering.—SGS LTER partnering included university organizational units and CSU Libraries. On-campus units included the SGS LTER administrative unit, a campus-wide service providing support for Web sites, and a research department that provided LTER data systems and database applications. For SGS LTER, new partnerships and new forms of connectivity took time to cultivate but supported migration of long-term data and ultimately an exemplar of a long view of data stewardship. Partners hold the potential to provide redundancy in data repositories as well as some measure of sustainability during all phases of research projects (Eschenfelder et al. 2016).

The CSU Libraries is known today for offering services to individuals and organizations generating information and knowledge about the region (Monaghan et al. 2019). The SGS LTER data management partnership with the library in pursuing the formation of a legacy project collection enabled cross-fertilization and a proactive co-design approach. The partners pursued novel options for addressing continued access to SGS LTER assets at a time when concerns relating to sustainability of data repositories represented an emergent challenge of heightened importance (OECD 2017a, b). With many facets to sustainability of digital materials to consider (Eschenfelder et al. 2016), CSU Libraries was identified as an appropriate host to take long-term responsibility for a digital legacy project collection due to its potential stability within academic institutions as well as its proximity to the SGS LTER site and its interest in partnering with SGS LTER data management to prototype migration of scientific data. Partnering was one approach to preserving data in a world of changing science, technologies, and institutional arrangements.

Several larger-scale, technology-oriented interactions initiated at the network level and influencing LTER data management development were ongoing but beyond the scope of this paper. These include activities that were carried out in partnership with centers such as the National Center of Ecological Analysis and Synthesis (NCEAS) and the U.S. LTER Network Office (LNO) together with programs such as the Knowledge Network for Biodiversity (KNB, Andelman et al. 2004) and the Science Environment for Ecological Knowledge (SEEK, Michener et al. 2007). They broadened community understanding of metadata standards that subsequently became critical to larger-scale open-access ecological repositories (Waide et al. 2017).

Tend to relations between external repositories.—The multi-repository data arrangements for the SGS LTER data are shown in Fig. 3. The SGS LTER local data repository (green) on the left includes a data system and supplemental materials posted on the site’s Web site. The LTER NIS is the community repository (blue, top right) with a highly structured architecture that supports metadata-enabled queries across all sites’ datasets. SGS LTER submission of data to this LTER Network data repository makes its data available to a wider audience. A subsequent partnership was formed with an institutional repository at CSU (red, bottom right), housing more loosely structured collections. The SGS LTER submission, consisting of data packages (residing also in the LTER NIS) together with supplemental materials, is structured for browsing rather than queries across datasets. The dashed line between the two remote, or non-local, repositories indicates the need for future work on describing relations that enable meaningful exchange of information about data that reside within more than one repository. Establishing relations between a community repository and an institutional repository creates a configuration that provides users with the ability to query about datasets in one repository and to browse contextual information that supports reuse of the data in another repository.

SGS LTER collaboration with CSU Libraries led to examining the relations between repositories. With data packages from the SGS LTER local data repository submitted for many years to the LTER NIS, this community repository initially was considered by SGS LTER as its primary repository. The site’s perspective on data repositories shifted dramatically, however, during decommissioning. The CSU institutional
repository became seen as the primary repository for SGS LTER with its holdings of supplemental materials in addition to data packages. As a secondary repository, the LTER NIS metadata was updated during decommissioning to refer to the site’s status as a closed LTER site and to document the end of data submission to the NIS.

The SGS LTER case illustrates open access to data with data residing in two different repositories. The development of persistent uniform resource locators represents a technical advance in data findability where the NIS community repository uses internationally registered Digital Object Identifiers (DOI, Brase et al. 2015) and the CSU institutional repository uses a service with unique local handles. For cases such as the SGS LTER data with more than one unique locator, however, policies and practices were not yet in place to establish interconnections between the repositories that made clear the relations between the repositories and their differing approaches to presenting data collections (OSTP 2020). Though it was not yet possible to navigate easily across facilities within a “web-of-repositories” (Baker and Yarmey 2009, Waide et al. 2017) as part of the larger landscape of data infrastructures, a link was provided in the interim within the NIS SGS LTER dataset abstract to the legacy project collection at CSU Libraries.

The vision for an SGS LTER legacy project collection included staging machine-readable data packages that could be accessed by and ingested into the LTER NIS repository for discovery. In addition to describing relations between the two repositories, the dashed line in Fig. 3 is an indicator of the potential for one repository to harvest data from another. This approach to aligning and linking the CSU institutional repository with the LTER NIS exceeded project and funding agency expectations while introducing library staff to opportunities for interoperability by identifying and addressing new technical possibilities.

**Cultivate a long view of data management**

A long-term project with embedded data management such as SGS LTER exemplifies approaches to the development of data systems, resources, and services that ensure long-term continuity of open data and information. Change in data management at the site was evident with the increase in support for the role from half time to full time, the development of a local data repository including a site-based Web site, and the delivery of datasets enhanced first by the delivery of the data locally and then via the NIS multi-site collections of data packages. The investment in embedded data management and the funding that led to development of continuing, incremental design at SGS LTER enabled the evolution of approaches to data and data infrastructure carried out at SGS LTER. Collaborating within the site and with new partners enabled proactivity in terms of identifying and addressing data issues. This included planning both for embedded, long-term management of data and subsequently a rethinking of the long-term when facing decommissioning that compelled the

---

**Fig. 3. Relations of the SGS LTER data repository with two other repositories are described: (1) creation of data packages and submission of SGS LTER data packages to the LTER Network Information System (NIS), and (2a, 2b) the migration during site decommissioning of data packages as well as supplemental materials to form a digital “legacy project collection” at the CSU Libraries institutional repository. The dashed line between the LTER NIS and the CSU institutional repository indicated there are still specifics within the interoperability of the systems not fully resolved, but the descriptions of the relationship are included within the abstract of each data package shared between the institutional and community repositories. CSU, Colorado State University; SGS LTER, Shortgrass Steppe Long Term Ecological Research site.**
proactive response of finding new ways to ensure the longevity of data.

Decommissioning of the site prompted data management to formulate a new view of data stewardship. Questions were asked by SGS LTER first internally and then directed to the IMC and the Network Office during its decommissioning. Administrative and social uncertainties with a discontinued site and its data management issues became evident. Questions arose such as whether the SGS LTER site was to participate in IMC meetings during this time, whether the community would address topics such as how to present terminated projects in the NIS, and how to relate the SGS collection in the LTER NIS to the SGS LTER legacy project collection at the CSU Libraries. Finding ways to address termination not just as a closing up of the site but as a time of transformation required local inventiveness and strategic planning that resulted in a long view that prompted new perspectives on open data and data stewardship.

Long Live the Data!

Having described how embedded data management grew within the U.S. LTER Network during the SGS LTER site’s 32-yr life cycle of activation, maturation, and then transformation, major challenges that emerged are summarized above in Table 2. In answer to questions of how this study of embedded data management contributes to our understanding of the enactment of open data access as well as of how to handle “the end of long term,” we offer five data management recommendations. These recommendations are summarized in Table 3 together with a reminder of the strategies discussed above that were developed by SGS LTER-embedded data management.

The U.S. LTER Network, with relationships anchored by site-based fieldwork and data sharing as well as by working groups that met routinely, involved various sites, individuals, projects, disciplines, and roles. The heterogeneity of arrangements, activities, and interactions establishes an environment of community communications that support collaborative research efforts (Goring et al. 2014). Data-related LTER relations existing within-site, cross-site, and network-wide are identified as a site-network data management dynamic as well as a multi-level science-data management dynamic that contribute to the U. S. LTER as a collaborative learning environment. The formation of the digital legacy project collection is just one example of learning from a new perspective on the longevity of data, followed by an inventive response, which occurred during decommissioning. It exceeded the NSF minimum open data requirement for LTER sites to submit dataset packages to a community repository. Embedded data management discerned that collections with datasets and metadata alone do not capture the rich scientific backstory accumulated during a collaborative site-based research effort. The SGS LTER legacy project collection, an approach complementary to merely archiving datasets with metadata, provided a more comprehensive suite of materials that were generated at the site. During the collection formation process, the SGS LTER Web site was reframed as a valuable source of research information. It served as a “pre-archive collection.” The formation of a legacy project collection provides a glimpse of how embedded data management is equipped to respond to change.

The role of data management in facilitating data sharing and data reuse is fundamental. In exploring embedded data management as an approach able to address the challenges of collective data management, this study illustrates the development of open data, the decommissioning of a long-term project, and the evolution of local data infrastructures. Documenting the sociotechnical evolution of an embedded data management approach highlights the contributions to data care made at the site of data generation. Our investigation captures issues faced by research communities in producing open access to data. With data management situated locally and working collaboratively with scientists, relationships that result in constant communication are key in planning for the continuing changes observed in data systems, data infrastructures, and data stewardship. Embedded data management fostered an ethic of care through persistent participation and by championing design of data arrangements that fit the local circumstances while ensuring the multiple forms data live on. Despite termination of the SGS as a long-term LTER site, embedded data management together...
Table 3. Recommendations for data management are given with examples of SGS LTER strategies.

| Data management recommendations | SGS LTER strategies |
|---------------------------------|---------------------|
| 1. Invest in data management locally | Supported a data manager continuously to ensure capable leadership for ongoing design and development of embedded data management |
| 2. Collaborate and learn within your community | Engaged in site-network data management and science-data management dynamics to inform collaborative learning |
| 3. Address data issues proactively and creatively | Recognized collaborative learning as informing embedded data management through shared data activities that foster situated creativity |
| 4. Consider relationships with those outside your community | Identified partners having experience with collections and institutional repositories while considering the relations between data repositories |
| 5. Cultivate a long view with data management | Designed for evolution of data, data management, and data infrastructure including a plan for decommissioning to ensure longevity of data |

Note: SGS LTER, Shortgrass Steppe Long Term Ecological Research site.

with membership in a network of sites prepared SGS LTER data management for continuously rethinking data management strategies. With this approach, the meaning of “long live the data” was enriched to include both “long live the data-sets” and “long live a legacy project collection.”

ACKNOWLEDGMENTS

We thank members of the communities involved for their participation in this study. We appreciate Justin Derner and Hailey Wilmer for reviewing earlier versions of the manuscript. As part of this ethnographic research, Institutional Review Board review processes were followed from the University of Illinois Urbana-Champaign and Colorado State University. We acknowledge our funders: the LTER SGS Supplement at CSU (NSF Award DEB #1027319) and the grants “The Challenge of the Long-Term Perspective for Data-Intensive Collaboration in e-Research” and “Multi-scoped infrastructure: Forming knowledge infrastructure for the ILTER Network” at the University of Oulu (The Academy of Finland #218189, #285903 respectively). We also acknowledge assistance from the University of Illinois School of Information Sciences and Rangeland Resources and Systems Research Unit, Agricultural Research Service, USDA. All authors conducted fieldwork and analyzed data individually and collectively. Nicole E. Kaplan and Karen S. Baker wrote initial drafts. Helena Karasti provided theoretical insight and analytic input during the writing process. Nicole E. Kaplan was affiliated with the Shortgrass Steppe Long-Term Ecological Research project as the information manager at Colorado State University, Fort Collins, CO. Her current affiliation is with the Long-Term Agroecosystem Research project at the USDA Agricultural Research Service, Rangeland Resources and Systems Research Unit, Fort Collins, CO. Karen S. Baker is affiliated with the iSchool at the University of Illinois Urbana-Champaign, IL. Dr. Baker was also a previous information manager with two other U.S. LTER sites. Helena Karasti is affiliated with the IT University of Copenhagen, Denmark, and has studied data management and research infrastructure formation in U.S., Finnish, and European LTER Networks.

LITERATURE CITED

Andelman, S. J., C. M. Bowles, M. R. Willig, and R. B. Waide. 2004. Understanding environmental complexity through a distributed knowledge network. BioScience 54:240–246.

Angelstam, P., et al. 2019. LTSER platforms as a place-based transdisciplinary research infrastructure: learning landscape approach through evaluation. Landscape Ecology 34:1461–1484.

Aronova, E., K. S. Baker, and N. Oreskes. 2010. Big science and big data in biology: from the International Geophysical Year through the International Biological Program to the Long Term Ecological Research (LTER) network, 1957–present. Historical Studies in the Natural Sciences 40:183–224.

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

Baker, K. S., and C. L. Chandler. 2008. Enabling long-term oceanographic research: changing data practices, information management strategies and informatics. Deep Sea Research Part II: Topical Studies in Oceanography 55:2132–2142.

Baker, K. S., and H. Karasti. 2004. The long-term information management trajectory: working to support
data. Science and Technology. http://escholarship.org/uc/item/7d6d4x0bd
Baker, K. S., and H. Karasti. 2018. Data care and its politics: designing for local collective data management as a neglected thing. In Proceedings of the 15th Participatory Design Conference: Full Papers 1:1–12.
Baker, K. S., H. Karasti, K. Vanderbilt, and D. Pennington. 2011. LTER information management and collaborative learning environments. Scripps Institution of Oceanography Technical Report July 2011. http://escholarship.org/uc/item/9k2879fp
Baker, K. S., and M. Mayernik. 2020. Disentangling knowledge production and data production. Ecosphere 11:1–12.
Baker, K. S., and F. Millerand. 2007. Articulation work supporting information infrastructure design: coordination, categorization, and assessment in practice. Pages 242a in IEEE 40th Annual Hawaii International Conference on System Sciences (HICSS‘07).
Baker, K. S., and F. Millerand. 2010. Infrastructuring ecology: challenges in achieving data sharing. Pages 111–138 in J. N. Parker, N. Vermeulen and B. Penders, editors. Collaboration in the new life sciences. Ashgate, Farnham, UK.
Baker, K. S., and L. Yarmey. 2009. Data stewardship: environmental data curation and a web-of-repositories. International Journal of Digital Curation 4:12–27.
Balvanera, P., et al. 2017. Key features for more successful place-based sustainability research on social-ecological systems: a Programme on Ecosystem Change and Society (PECS) perspective. Ecology and Society 22:14.
Benson, B. J., P. C. Hanson, J. W. Chipman, and C. J. Bowser. 2006. Breaking the data barrier: research facilitation through information management. Pages 259–279 in J. J. Magnuson, T. K. Kratz, and B. J. Benson, editors. Long-term dynamics of lakes in the landscape: long-term ecological research on north temperate lakes. Oxford University Press, Oxford, UK.
Billick, I., and M. V. Price. 2010. The ecology of place: contributions of place-based research to ecological understanding. University of Chicago Press, Chicago, Illinois, USA.
Bocking, I. 1997. Ecologists and environmental politics: a history of contemporary ecology. Yale University Press, New Haven, Connecticut, USA.
Brase, J., M. Lautenschlager, and I. Sens. 2015. The tenth anniversary of assigning DOI names to scientific data and a five year history of DataCite. D-Lib Magazine 21:10.
Brown, J. S., and P. Duguid. 1991. Organizational learning and communities-of-practice: toward a unified view of working, learning, and innovation. Organization Science 2:40–57.
Campbell, L. K., E. S. Svendsen, and L. A. Roman. 2016. Knowledge co-production at the research-practice interface: embedded case studies from urban forestry. Environmental Management 57:1262–1280.
Chabbi, A., and H. W. Loescher. 2017. Terrestrial ecosystem research infrastructures: challenges, new developments and perspectives. CRC Press Taylor & Francis Group, New York, New York, USA.
Cheruvilil, K. S., and P. A. Soranno. 2018. Data-intensive ecological research is catalyzed by open science and team science. BioScience 68:813–822.
Coleman, D. C. 2010. Big ecology: the emergence of ecosystem science. University of California Press, Oakland, California, USA.
Cragin, M. H. 2009. Scientific data collections: use in scholarly communication and implications for data curation. Dissertation. University of Illinois at Urbana-Champaign, Champaign, Illinois, USA.
Cragin, M. H., and K. Shankar. 2006. Scientific data collections and distributed collective practice. Computer Supported Cooperative Work 15:185–204.
Duerr, R. E., M. A. Parsons, M. Marquis, R. Dichtl, and T. Mullins. 2004. Challenges in long-term data stewardship. Pages 47–67 in Twenty-First IEEE Conference on Mass Storage Systems and Technologies (MSST). University of Maryland, College Park, Maryland, USA.
Emerson, R. M., R. I. Fretz, and L. L. Shaw. 2011. Writing ethnographic fieldnotes. University of Chicago Press, Chicago, Illinois, USA.
Eschenfelder, K. R., K. Shankar, R. Williams, A. Lanham, D. Salo, and M. Zhang. 2016. What are we talking about when we talk about sustainability of digital archives, repositories and libraries? Proceedings of the Association for Information Science and Technology 53:1–6.
Fegraus, E. H., A. Andelman, M. B. Jones, and M. Schildhauer. 2005. Maximizing the value of ecological data with structured metadata: an introduction to Ecological Metadata Language (EML) and principles for metadata creation. Bulletin of the Ecological Society of America 86:158–168.
Fincham, R. 2002. Narratives of success and failure in systems development. British Journal of Management 13:1–14.
Golley, F. B. 1993. A history of the ecosystem concept in ecology: more than the sum of the parts. Yale University Press, New Haven, Connecticut, USA.
Goring, S. J., K. C. Weathers, W. K. Dodds, P. A. Soranno, L. C. Sweet, K. S. Cheruvilil, J. S. Kominoski,
INNOVATIVE VIEWPOINTS

J. Rüegg, A. M. Thorn, and R. M. Utz. 2014. Improving the culture of interdisciplinary collaboration in ecology by expanding measures of success. Frontiers in Ecology and the Environment 12:39–47.

Gray, J., C. Gerlitz, and L. Bounegru. 2018. Data infrastructure literacy. Big Data & Society 5:1–11.

Hampton, S. E., et al. 2017. Skills and knowledge for data-intensive environmental research. BioScience 67:546–557.

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

Hartter, J., S. J. Ryan, C. A. MacKenzie, J. N. Parker, and C. A. Strasser. 2013. Spatially explicit data: stewardship and ethical challenges in science. PLOS Biology 11:e1001634.

Henderson, A., and M. Kyng. 1991. There’s no place like home: continuing design in use. Pages 219–240 in J. Greenbaum and M. Kyng, editors. Design at work: cooperative design of computer systems. Lawrence Erlbaum Associates, Hillsdale, New Jersey, USA.

ICPSR [Interuniversity 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. https://www.icsu-wds.org/files/DomeinRepositories_Call_to_Action_16_Sept_2013.pdf.

Imperial, M. T., E. Johnston, M. Pruett-Jones, K. Leong, and J. Thomsen. 2016. Sustaining the useful life of network governance: life cycles and developmental challenges. Frontiers in Ecology and the Environment 14:135–144.

Ingersoll, R. C., T. R. Seastedt, and M. Hartman. 1997. A model information management system for ecological research. BioScience 47:310–316.

Jackson, S., and A. Buyuktur. 2014. Who killed WATERS? Mess, method, and forensic explanation in the making and unmaking of large-scale science networks. Science, Technology & Human Values 39:285–308.

Johnson, J. C., R. R. Christian, J. W. Brunt, C. R. Hickman, and R. B. Waide. 2010. Evolution of collaboration within the US Long Term Ecological Research Network. BioScience 60:931–940.

Kaplan, N. E., K. S. Baker, D. C. Draper, and S. Swauger. 2014. Packaging, transforming, and migrating data from a scientific research project to an institutional repository: the SGS LTER collection. Colorado State University, Fort Collins, Colorado, USA. https://mountainscholar.org/handle/10217/87239

Karasti, H., and K. S. Baker. 2004. Infrastructuring for the long-term: ecological information management. In IEEE 37th Annual Hawaii International Conference on System Sciences (HICSS’04).

Karasti, H., and K. S. Baker. 2008a. Community design: growing one’s own information infrastructure. Pages 217–220 in Proceedings of the Participatory Design Conference 2008 Sep 30–Oct 04.

Karasti, H., and K. S. Baker. 2008b. Digital data practices and the Long Term Ecological Research program growing global. International Journal of Digital Curation 3:42–58.

Karasti, H., K. S. Baker, and E. Halkola. 2006. Enriching the notion of data curation in e-science: data managing and information infrastructuring in the Long Term Ecological Research (LTER) Network. Computer Supported Cooperative Work 15:321–358.

Karasti, H., K. S. Baker, and F. Millerand. 2010. Infrastructure time: long term matters in collaborative development. Computer Supported Cooperative Work 19:377–415.

Kingsland, S. E. 2010. The role of place in the history of ecology. Pages 15–39 in I. Billick and M. V. Price, editors. The ecology of place: contributions of place-based research to ecological understanding. University of Chicago Press, Chicago, Illinois, USA.

Kitchin, R. 2014. The data revolution: big data, open data, data infrastructures and their consequences. Sage Publications, Los Angeles, California, USA.

Krahe, D. 2012. The ill-fated NBS: a historical analysis of Bruce Babbitt’s vision to overhaul interior science. Pages 160–165 in S. Weber, editor. Rethinking Protected Areas in a Changing World: Proceedings of the 2011 GWS Biennial Conference on Parks, Protected Areas, and Cultural Sites. The George Wright Society, Hancock, Michigan, USA. http://georgewright.org/proceedings2011

Lamb, D. 2017. Long-term ecological monitoring and institutional memories. Ecological Management & Restoration 18:200–204.

Laney, C. M., K. S. Baker, D. P. C. Peters, and K. W. Ramsey. 2013. Recommendations for data accessibility. Pages 216–255 in D. P. C. Peters, C. M. Laney, A. E. Lugo, S. L. Collins, C. T. Driscoll, P. M. Goffman, J. M. Grove, A. K. Knapp, T. K. Kratz, M. D. Ohman, R. B. Waide, and J. Yao, editors. Long-term trends in ecological systems: a basis for understanding responses to global change. United States Department of Agriculture, Agricultural Research Service, Las Cruces, New Mexico, USA.

Lave, J., and E. Wenger. 1991. Situated learning: legitimate peripheral participation. Cambridge University Press, Cambridge, UK.
Lindlof, T. R., and B. C. Taylor. 2011. Qualitative communication research methods. Sage Publications, Los Angeles, California, USA.

Lynch, C. A. 2003. Institutional repositories: essential infrastructure for scholarship in the digital age. Libraries and the Academy 3:327–336.

Lynch, C. 2008. The institutional challenges of cyberinfrastructure and e-research. EDUCAUSE Review 43:74–76.

McNutt, M., K. Lehnert, B. Hanson, B. A. Nosek, A. M. Ellison, and J. L. King. 2016. Moving to an open-data culture, Promote reproducibility by sharing data and materials. Science 351:1024–1026.

McNulty, S. A., D. White, M. Hufty, and P. Foster. 2017. The organization of biological field stations at fifty. Bulletin of the Ecological Society of America 98:359–373.

Michener, W. K., J. H. Beach, M. B. Jones, B. Ludäschner, D. D. Pennington, R. S. Pereira, A. Rajasekar, and M. Schildhauer. 2007. A knowledge environment for the biodiversity and ecological sciences. Journal of Intelligent Information Systems 29:111–126.

Michener, W. K., J. W. Brunt, and S. G. Stafford. 1994. Environmental information management and analysis: ecosystem to global scales. CRC Press, Boca Raton, Florida, USA.

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

Michener, W. K., and R. B. Waide. 2008. The evolution of collaboration in ecology: lessons from the US Long-Term Ecological Research program. Pages 297–309 in G. M. Olson, A. Zimmerman, and N. Box, editors. Scientific collaboration on the internet. Volume 1. The MIT Press, Cambridge, Massachusetts, USA.

Millerand, F., and K. S. Baker. 2011. Ocean informatics monograph, Ocean informatics initiative: an ethnographic study (2002–2006). Scripps Institution of Oceanography Technical Report. San Diego, California, USA. https://escholarship.org/uc/item/67x0f1g7

Millerand, F., and K. S. Baker. 2020. Data infrastructures in ecology: an infrastructure studies perspective. In H. Shugart, editor. Oxford research encyclopedia of environmental science. Oxford University Press, Oxford, UK. https://doi.org/10.1093/acrefore/9780199389414.013.554

Millerand, F., and G. C. Bowker. 2009. Metadata standards: trajectories and enactment in the life of an ontology. Pages 149–165 in M. Lampland and S. L. Star, editors. Standards and their stories. How quantifying, classifying, and formalizing practices shape everyday life. Cornell University Press, Ithaca, New York, USA.

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

Mirtl, M., et al. 2018. Genesis, goals and achievements of long-term ecological research at the global scale: a critical review of ILTER and future directions. Science of the Total Environment 626:1439–1462.

Monaghan, A. J., B. Milash, T. Magle, T. Hauser, T. E. Cheatham, P. J. Burns, and H. J. Siegel. 2019. Improving regional cyberinfrastructure services through collaboration: cyberteam for the Rocky Mountain Advanced Computing Consortium. Pages 1–4 in Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines. Boulder, Colorado, USA.

Moran, M. S., P. Heilmann, D. P. Peters, and C. Holifield Collins. 2016. Agroecosystem research with big data and a modified scientific method using machine learning concepts. Ecosphere 7:1–11.

Nielsen, M. 2011. Reinventing discovery: the new era of networked science. Princeton University Press, Princeton, New Jersey, USA.

NRC [National Research Council]. 1995. Colleges of agriculture at the land grant universities: a profile. National Academies Press, Washington D.C., USA. https://www.nap.edu/catalog/4980/colleges-of-agriculture-at-the-land-grant-universities-a-profile

NSB [National Science Board]. 2005. Long-lived digital data collections: enabling research and education in the 21st century. National Science Board Report NSB-05-04, September 2005. http://www.nsf.gov/pubs/2005/nsb05040

O’Brien, M., D. Costa, and M. Servilla. 2016. Ensuring the quality of data packages in the LTER network data management system. Ecological Informatics 36:237–246.

OECD [Organisation for Economic Co-operation and Development]. 2017a. Business models for sustainable research data repositories, Report No. 47 (December 2017). OECD Publishing, Château de la Muette in Paris, France.

OECD [Organisation for Economic Co-operation and Development]. 2017b. Strengthening the effectiveness and sustainability of international research infrastructures, Report No. 48 (December 2017). OECD Publishing, Château de la Muette in Paris, France.

OSTP [Office of Science and Technology]. 2020. Request for public comment on draft desirable characteristics of repositories for managing and sharing data resulting from federally funded
research. The White House, Washington D.C., USA. https://www.federalregister.gov/d/2020-00689

Ostrom, E. 2008. Polycentric systems as one approach for solving collective-action problems. Paper (2008-11), 02. Indiana University, Bloomington, School of Public & Environmental Affairs Research, Bloomington, Indiana, USA.

Palmer, C. L., N. M. Weber, T. Munoz, and A. H. Reenear. 2013. Foundations of data curation: the pedagogy and practice of “purposeful work” with research data. Archive Journal 3. http://hdl.handle.net/2142/78099

Pasquetto, I., B. Randles, and C. Borgman. 2017. On the reuse of scientific data. Data Science Journal 16:1–9.

Peer, L., and A. Green. 2012. Building an open data repository for a specialized research community: process, challenges and lessons. International Journal of Digital Curation 7:151–162.

Peters, D. P., H. W. Loescher, M. D. SanClements, and K. M. Havstad. 2014. Taking the pulse of a continent: expanding site-based research infrastructure for regional-to-continental-scale ecology. Ecosphere 5:1–23.

Pickett, S. T. A., et al. 2016. Dynamic heterogeneity: a framework to promote ecological integration and hypothesis generation in urban systems. Urban Ecosystems 20:1–14.

Popkin, G. 2019. Data sharing and how it can benefit your scientific career. Nature 569:445.

Porter, J. H. 2010. A brief history of data sharing in the US Long Term Ecological Research Network. Bulletin of the Ecological Society of America 91:14–20.

Porter, J. H. 2019. Evaluating a thesaurus for discovery of ecological data. Ecological Informatics 51:151–156.

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

Ritchie, J., and J. Lewis. 2004. Qualitative research practice. A guide for social science students and researchers. Sage Publications, Los Angeles, California, USA.

Robertson, G. P., et al. 2008. Long-term agricultural research: a research, education, and extension imperative. BioScience 58:640–645.

Roscione, J., and S. D. Teasley. 1995. The construction of shared knowledge in collaborative problem solving. Pages 69–97 in C. E. O’Malley, editor. Computer supported collaborative learning. Springer-Verlag, Berlin, Germany.

Rossman, G. B., and S. F. Rallis. 2011. Learning in the field: an introduction to qualitative research. Sage Publications, Los Angeles, California, USA.

Rubin, H. J., and I. S. Rubin. 2012. Qualitative interviewing: the art of hearing data. Third edition. SAGE Publications, Los Angeles, California, USA.

Servilla, M., and J. Brunt. 2011. The LTER Network Information System: improving data quality and synthesis through community collaboration. AGU Fall Meeting Abstracts 2011:N51C–1598.

Servilla, M., J. Brunt, D. Costa, J. McGann, and R. Waide. 2016. The contribution and reuse of LTER data in the Provenance Aware Synthesis Tracking Architecture (PASTA) data repository. Ecological Informatics 36:247–258.

Smith, R. C., C. Bossen, C. Dindler, and O. S. Iversen. 2020. When participatory design becomes policy: technology comprehension in Danish education. Pages 148–158 in Participatory Design Conference, Biennial Participatory Design Conference. ACM-Association for Computing Machinery, New York, New York, USA.

Stafford, S. G., J. W. Brunt, and W. K. Michener. 1994. Integration of scientific information management and environmental research. Pages 3–20 in W. K. Michener, J. W. Brunt, and S. G. Stafford, editors. Environmental information management and analysis: ecosystem to global scale. Taylor & Francis, London, UK.

Stafford, S. G., N. E. Kaplan, and C. W. Bennett. 2002. Through the Looking Glass: What do we see, what have we learned, what can we share? Information management at the Shortgrass Steppe Long Term Ecological Research site. In International Institute of Informatics and Systematics (SCI2002), July 14–18, 2002, Orlando, Florida.

Stegmaier, P., S. Kuhlmann, and V. R. Visser. 2014. The discontinuation of socio-technical systems as a governance problem. In S. Borrás and J. Edler, editors. The governance of socio-technical systems: explaining change. Edward Elgar Publishing, Cheltenham, UK.

Sutter, R. D., S. B. Wainscott, J. R. Boetsch, C. J. Palmer, and D. J. Rugg. 2015. Practical guidance for integrating data management into long-term ecological monitoring projects. Wildlife Society Bulletin 39:451–463.

Tenopir, C., B. Birch, and S. Allard. 2012. Academic libraries and research data services: current practices and plans for the future. An ACRL white paper. Association of College and Research Libraries, a division of the American Library Association, Chicago, Illinois, USA.

Voss, A., R. Procter, R. Slack, M. Hartswook, and M. Rouncefield. 2009. Design as and for collaboration: making sense of supporting practical action. Page 31–58 in A. Voss, R. Procter, R. Slack, M.
Hartwood, M. Rouncefield, and M. Buscher, editors. Computer supported cooperative work configuring user-designer relations. Springer-Verlag, London, UK.
Waide, R. B., J. W. Brunt, and M. S. Servilla. 2017. Demystifying the landscape of ecological data repositories in the United States. BioScience 67:1044–1051.
Wilkinson, M. D., et al. 2016. The FAIR guiding principles for scientific data management and stewardship. Scientific Data 3:1–9.
Worster, D. 1997. Nature’s economy: a history of ecological ideas. Cambridge University Press, Cambridge, UK.
Yarmey, L., and K. S. Baker. 2013. Towards standardization: a participatory framework for scientific standard-making. International Journal of Digital Curation 8:157–172.
Yoon, A. 2017. Data reusers’ trust development. Journal of the Association for Information Science and Technology 68:946–956.