Received: 9 February 2021 | Revised: 11 November 2021 | Accepted: 12 January 2022

DOI: 10.1002/asi.24620

RESEARCH ARTICLE

JASIST WILEY

Integrative data reuse at scientifically significant sites: Case studies at Yellowstone National Park and the La Brea Tar Pits

Andrea K. Thomer

School of Information, University of Michigan, Ann Arbor, Michigan, USA

Correspondence
Andrea K. Thomer, School of Information, University of Michigan, 105 S. State Street, Ann Arbor, MI 48104.
Email: athomer@umich.edu

Funding information
IMLS, Grant/Award Number: LG-06-12-0706-12

Abstract

Scientifically significant sites are the source of, and long-term repository for, considerable amounts of data—particularly in the natural sciences. However, the unique data practices of the researchers and resource managers at these sites have been relatively understudied. Through case studies of two scientifically significant sites (the hot springs at Yellowstone National Park and the fossil deposits at the La Brea Tar Pits), I developed rich descriptions of site-based research and data curation, and high-level data models of information classes needed to support integrative data reuse. Each framework treats the geospatial site and its changing natural characteristics as a distinct class of information; more commonly considered information classes such as observational and sampling data, and project metadata, are defined in relation to the site itself. This work contributes (a) case studies of the values and data needs for researchers and resource managers at scientifically significant sites, (b) an information framework to support integrative reuse at these sites, and (c) a discussion of data practices at scientifically significant sites.

1 | INTRODUCTION

Scientifically significant sites are localities that have attracted enough attention from researchers to merit special administration and/or protection, as well as the preservation of associated specimens and data collections. These sites are logical and often efficient points for data curation (Karasti & Baker, 2008; Millerand & Baker, 2010; Palmer et al., 2017; Thomer, Wickett, et al., 2018). Many scientifically significant sites feature corresponding museums, archives, and research stations tasked with the long-term management and curation of their associated data and artifacts, particularly in the natural sciences. Notable examples include the many research localities within the U.S. National Park Service (NPS); ecological field stations such as those in Long Term Ecological Research (LTER) network; and county- or state-managed lands such as lakes, marshes, and unique geological formations that are preserved for both recreation and research. However, despite the prevalence of site-based memory institutions, the work of site-based data curation has been relatively understudied and undertheorized in information science, where data curation research has primarily focused on the interests of scholarly institutions and archives.

A clear understanding of data practices and data properties at these sites is needed to support further data curation. Site-based data collections and scientific studies represent a different paradigm of scientific investigation than that of laboratory-based "bench science" or of the "big science, little science" framing that sees data practices through the lens of team size and infrastructure scale (de Solla Price, 1986). The closest corollary might be collection-based science (Strasser, 2012), but this framing does not account for the centrality of the site, and how that impacts data practices and data collections. These sites might be uniquely suited to facilitating the "integrative reuse" (Pasquetto et al., 2017) needed for systems science, but further work is needed to develop best practices and infrastructures to support such research and curation (Palmer et al., 2017).

To better understand and support data practices and curation needs at scientifically significant sites, I compared two case studies of site-based research and curation: geobiology research at Yellowstone National Park and paleontology research at the La Brea Tar Pits. Both cases were developed using participatory methods, in which I conducted interviews and focus groups to understand what of scientifically significant sites stakeholders most value in their work. I then worked with stakeholders to develop "minimum information frameworks" (MIFs) summarizing their most needed and valued data. This research has resulted in rich descriptions of the data practices of researchers and resource managers at these sites, as well as a high-level data model for future curation guidelines and infrastructure development.

2 | BACKGROUND

Much of the prior work supporting the data curation needs of scientifically significant sites has been undertaken within the informatics-oriented corners of site-based domain sciences (e.g., ecology, geology, biodiversity). For instance, research conducted at (and about) the LTER network and at independent biological field stations has resulted not only in a number of metadata standards for the management of data from these scientifically significant sites, notably the widely used Ecological Metadata Language (EML; Michener et al., 1997), but also significant research about the development of standards and curation of data more broadly (Brunt & Michener, 2009; Karasti et al., 2006; Karasti & Baker, 2008; Millerand & Baker, 2010; Millerand & Bowker, 2009; Ribes & Finholt, 2009). Additionally, the biodiversity informatics communities have developed standards such as the Environment Ontology (Buttigieg et al., 2013), the Biocollections Ontology (Walls et al., 2014), and Darwin Core (Wieczorek et al., 2012), all of which enable the detailed, semantically rich description of natural science data as well as the site from which they are collected.

Within the information sciences, the focus has been on the data practices (as defined by Cragin et al., 2010) of domain scientists who work *at* these sites, rather than the curatorial needs of those managing the sites. Much of this research has focused on identifying barriers to data management and sharing (e.g., Akmon et al., 2011; Borgman, 2012; Borgman et al., 2007; Kowalczyk & Shankar, 2011; Stall et al., 2019; Tenopir et al., 2018, 2020; Wallis et al., 2013) or reuse (e.g., Darch & Borgman, 2016; Faniel & Jacobsen, 2010; Frank et al., 2015; Khan et al., 2019; Palmer et al., 2011; Weber et al., 2013; Yan et al., 2020; Zimmerman, 2008). While this prior research lays an important foundation to understanding the practices of the researchers working at these sites, it typically does not focus on the needs or practices of the managers of the sites or on the workflows necessary to build and sustain site-based data collections. Here, I aimed to bridge that gap by studying the data practices of the resource managers at scientifically significant sites—the park rangers, data managers, collection managers, and other staff who maintain the sites themselves, as well as their research infrastructure, over time.

Particularly, I focused on what it takes to facilitate *integrative reuse* of data at these sites. As Pasquetto and co-authors (2017) noted, data integration takes different—and often, more challenging—work than the independent reuse of an individual dataset. In their words:

> Reusing a single dataset in its original form is difficult, even if adequate documentation and tools are available, since much must be understood about why the data were collected and why various decisions about data collection, cleaning, and analysis were made. Combining datasets is far more challenging, as extensive information must be known about each dataset if they are to be interpreted and trusted sufficiently to draw conclusions. (Pasquetto et al., 2017, p. 4).

Integrative reuse is particularly challenging with "small science" datasets: those collected by individual researchers or small teams working on diverse projects and collecting heterogeneously structured data (as opposed to the "big science" of giant infrastructures such as the Large Hadron Collider or Hubble Space Telescope, which is conducted by large well-funded teams that often collect large amounts of homogeneously structured data; Price, 1986). By some measures, small science research projects makeup 80% of all science (Heidorn, 2008) and are expected to produce more data over time than big science (Carlson, 2006). But because of the complexity of its practice and culture, small science is very poorly served by curation and repository services (Cragin et al., 2010).

Scientifically significant sites might present a rare situation in which these small science datasets can find an easier road to integrative reuse. Although the range of studies conducted at a given site might be large, scientists are nevertheless collecting data from the *same geospatial region and geospatially linked phenomena*. The site and its unique physical qualities remain a common denominator and might facilitate distinct kinds of aggregation and integration (Palmer et al., 2017; Thomer, 2017), particularly in support of "systems science" that considers a natural system holistically (Yan et al., 2020).

One key to integrative reuse of site-based data is the development of data standards that model the relationships between the site's characteristics and those of the research conducted at the site. Data models are the mechanism by which we represent information in a sufficiently structured manner that a computer can understand, and map the relationships between, the roles and structures of objects and data (Kent, 1978). Data models are additionally critical to developing effective workflows between people and information systems (Dennis et al., 2015), and semantic technologies (Fox & Hendler, 2009; Narock & Fox, 2012). Prior work has shown that we must take the role of data modeling seriously if we are to build next-generation cyberinfrastructure; further, information modeling as an analytic technique can illuminate critical problems or quirks in data systems that must be addressed in cyberinfrastructure (Sacchi & Wickett, 2012; Wickett et al., 2012).

3 | METHODS: MULTISITE CASE STUDIES DEVELOPED THROUGH INTERVIEWS AND PARTICIPATORY ACTION

To better understand the diverse data practices at scientifically significant sites, I developed two cases of site-based research and data: first, geobiology research and resource management at Mammoth Hot Springs at Yellowstone National Park and, second, paleontology research and curation at the La Brea Tar Pits and Museum in Los Angeles. Each of these sites is managed by full-time curatorial and/or resource management staff, who oversees the site's maintenance and facilitate research access to the site. Each site also features associated data and/or specimen collections. Each case includes two embedded subcases: the work of scientists conducting research at the site and the work of curators and resource managers at the site.

Cases were developed in part through the participatory design of best practices for data curation, including high-level data models called MIFs (see Palmer et al., 2017; Thomer, Wickett, et al., 2018). Participatory research methods are rooted in the idea that the best way to understand something is to try to change it (Lewin, 1943). As reviewed by Hayes (2011), research-through-action places the researcher in an "ongoing dialogue" with the study community (rather than perhaps imposing the researcher's will on a community) and emphasizes the importance of sustainable change toward an agreed upon goal or the fulfillment of a pre-existing need (Hayes, 2011; Peters & Robinson, 1984). My approach here is similar to that of those using participatory design as part of the "infrastructuring" process (Karasti, 2014).

At both sites, I conducted semi-structured interviews with stakeholders about their data needs, priorities, and wish lists and reviewed key data artifacts and documentation. I then collaborated with stakeholders to model data collection and curation workflows (described in Thomer, Weber, & Twidale, 2018) and to develop MIFs outlining the data types they would ideally like to access and how those data might be organized in a repository or database. The collaboration took place through site visits, workshops, and remote meetings. Further details of data collection are in Table 1.

Each case was analyzed using an "explanation building" logic (Yin, 2009), which is done through iterative construction of narrative case reports, in which the researcher makes an initial proposition, compares the initial findings of a case against that proposition, revises the proposition, compares other details of the case against the revision, and then compares this revision to the facts of a second or more cases. This research was reviewed by the University of Illinois Institutional Review Board and found to pose no more than minimal risk to participants and therefore was exempt from oversight.

4 | RESULTS

Next I present short case narratives of each site. I also briefly present the MIFs I developed for each site, to facilitate the comparison to curatorial processes and data at each site.

4.1 | Yellowstone case

The grounding case for this study—geobiology research in the hot springs at Yellowstone National Park—was initially identified and developed through work on the

TABLE 1 Data collection

Data collection category	Yellowstone	La Brea
Interviews and other stakeholder engagement	Focus group with nine participants; approximately 20 telecons with Yellowstone resource managers; four follow-up interviews w/researchers; one follow-up interview with a resource manager	Interviews with 10 La Brea staff and 12 researchers who use the La Brea collections and data; ongoing collaboration with two core La Brea staff; three site visits
Artifact inventory and workflow modeling	Inventory and systems analysis of Fouke data; workflow modeling of data collection methods	Inventory of La Brea collections data sources; data collection workflow modeling

Notes: Data collection for the Yellowstone case was completed from 2013 to 2016. Data collection for the La Brea case was completed from 2015 to 2017 and was designed to roughly mirror that of the Yellowstone case.

Site-Based Data Curation project (Palmer et al., 2017; Thomer et al., 2014; Thomer, Wickett, et al., 2018). Yellowstone is home to nearly 40,000 geothermal features, such as hot springs and geysers, which are popular tourist attractions. Many rare thermophilic (heat-loving) microbes live in the superheated waters of these springs. Research on these microbes has led to important breakthroughs in the field geobiology, as well as critical advances in biotechnology (Fouke, 2011).

The "resource manager" stakeholder group at Yellowstone includes the park rangers working in the Research Permitting Office, as well as information professionals working in the archives, library, and natural history collections. The researcher group includes visiting researchers from academia and industry. Researchers must be granted a Scientific Research and Collecting Permit before they can conduct studies in the park. Resource managers and researchers primarily interact with each other through this permitting process and through annual reporting processes. All participants in this study expressed a desire to create shared data collections. However, although researchers are asked to send their data to the park after completing their projects, few do. My conversations and MIF design focused on understanding what each group needs from a data collection. Next, I describe each stakeholder group's roles and data needs, and then present the MIF we developed together.

4.1.1 | Resource manager roles and values: Research support and permit condition enforcement

The resource managers saw their roles as divided among supporting researchers in their work, managing park resources, and enforcing NPS permitting conditions and regulations. They emphasized that their job is to protect the park for recreational visitors as well as for researchers, and that they are managers rather than project directors. They felt comfortable enforcing current NPS guidelines, but not developing their own: "The types of positions that we have are managing the information that is given to us but not necessarily determining what is important to get" (YNP-RM-2).

Many of our conversations centered on issues of governance, compliance, and enforcement: If data collection or sharing standards were developed, who would be responsible for enforcing them? Resource managers noted that they already have problems enforcing existing permitting conditions, largely because they have very little leverage after a permit is granted:

> This is the frustrating part of the permitting process. It really focuses on what happens on the front end, making sure that the person is providing all the deliverables before and during the [permit application] process and that they are maintained, and not impacting resources during their [fieldwork]. (YNP-RM-2).

That is, the only real enforcement mechanisms that resource managers (particularly, the permitting staff) have over visiting researchers is their ability to (a) deny researchers a permit or (b) kick researchers out of the park while they are conducting fieldwork (which is very serious and would only be done for egregious safety violations or damage to natural resources). Resource managers consequently had to come up with creative workarounds to obtain the information they need for their work. For instance, because they received so few theses and papers sent back to them by researchers, they turned to setting search alerts for Yellowstone:

> I don't think we'd get 60% of the publications we get if [YNP-RM-3] and I and the gals didn't have Google Scholar alerts for Yellowstone. We went from having maybe 40–50 a year submitted to over 120 because we are actually looking for them. (YNP-RM-2)

4.1.2 | Resource manager data needs

Resource managers said that while primary data from researchers would be ideal, they would be just as happy with more information *about* research activities rather than access to the data themselves—essentially, this would be metadata about research activities going on in the park as the "larger context" of data collection and project activities (YNP-RM-4). Several resource managers noted that they simply need to know where researchers' data are being stored, and that they would like to store copies of data management plans, especially if these plans are already being written for various funding agencies:

> What would be helpful to get in the short term … is the data management plan. What are they going to do with the data when they are no longer active and people and can't find them anymore to ask directly? And I think now that they're going to have to often decide what repository they are going to put the data in sooner rather than later. That could be really useful. Because if they do retire and they can't be contacted anymore by somebody following up, we know at least where they thought they were going to deposit that information and that gives us a lead to provide to the researchers. (YNP-RM-4)

Resource managers were also interested in data that could help with strategic planning. Despite their not necessarily being responsible for this sort of work, they recognized it as something that they need to support, if not personally direct. These data would include a record of how many projects had been conducted at different sites so that the resource managers could better monitor the impact of research activities on springs over time. Data about the springs' structure and condition over time would help in this work.

4.1.3 | Researcher roles and values: Independence and access

Yellowstone researchers viewed the park as a "living laboratory"—a place to conduct experiments and collect data. Participants' research involved multiple kinds of field observations (biological, chemical, physical, geological, and genomic) and sample collection. Researchers valued their ability to access data for broad, integrative studies—while also wishing to control access to their personal data stores.

Occasionally tensions arose when researchers' goals were seen as in conflict with the park's role as a site of recreation. For instance, one Yellowstone permitting condition requires that researchers must "carry out all [research] activities out of public view" unless specifically authorized otherwise. One researcher described having to literally camouflage himself and his students and having to duck out of view whenever a park visitor approached. The researcher thought that this was excessive: Why not let park visitors see that scientific research was active and ongoing at the park? The resource managers, however, wanted to ensure that the park visitors could see the natural features—and did not get the idea that it was acceptable or safe to venture beyond the boardwalk.

4.1.4 | Researcher data needs

Whereas the resource managers primarily wanted metadata about projects and their impacts, researchers identified concrete categories of primary data that are highly reusable:

Genetic data. DNA and RNA sequences were unanimously seen as highly reusable, as well as next-generation genomic and metagenomic data. Access to these data facilitated comparative studies across sites over time.

Basic geochemistry. Data about the hot springs—particularly the water temperature and pH, and particularly geochemistry data collected through well-established methods, would be easier to reuse (or perhaps more trustworthy) than data collected through less commonly used methods.

Photographs, at a range of scales. Researchers agreed that simple point-and-shoot photographs of the spring and sampling sites made it possible for them to quickly assess site conditions at the time of sampling. Photographs also provided a basic understanding of how active (or inactive) a spring was, and of other seasonal conditions that might impact interpretation (e.g., snow cover, microbial mat growth, etc.).

Context. Finally, researchers also valued some more abstract qualities about their data, and about Yellowstone as a site. The concept of "context" came up repeatedly; well-contextualized data sets were more valuable and usable than non-contextualized datasets. Context could include information about data collection methods, environmental conditions, or project goals and hypotheses.

4.1.5 | Yellowstone MIF

In synthesizing the data needs presented here, I worked with site stakeholders over the course of a year to develop the following three-class data model. This model was described in Palmer et al. (2017), but it is adapted in Figure 1 to facilitate comparison with the La Brea case. The three classes of information needed for both researchers and resource managers include:

1. The *Field Campaign* class. This class contains metadata about a research project overall. Both stakeholder groups agreed that information about projects was important to capture, albeit for different reasons: resource managers wanted these data to monitor the use of a site over time, and researchers wanted the data to provide context for integrative reuse.
2. The *Hot Spring Structure* class. This includes information that characterizes the structure, geochemistry, behavior, and extant of a hot spring over time. Information in this class would allow researchers to integrate data from many studies and would support resource managers in monitoring springs over time.
3. The *Sample Sites and Measurements* class. This class includes information about each sample or observation collected, as well as information about each specific sampling site, including unique contextual information about each sampling site, such as the sample site's appearance, water temperature, and pH. As in the Hot Spring Structure class, this information would help characterize each sample's specific geochemical context, thereby facilitating data reuse and integration.

Just as the researchers' and resource managers' interactions center on the site, so do their data. The site and its characteristics over time are foundational to integrating data from both groups. The Field Campaign class includes the data primarily of interest to the resource managers, whereas the Sample Measurements data are primarily of interest to the researchers, but they could be integrated via the characteristics of the sites themselves.

4.2 | La Brea case

The second site for this study, the La Brea Tar Pits, was identified through my prior work as an excavator at the site itself. Like Yellowstone, La Brea is a site of both research and recreation. The La Brea Tar Pits are a cluster of rich fossil deposits located in the heart of Los Angeles, CA. During the ice age (~10,000–50,000 years ago), plants, animals, insects, and other organisms became trapped and fossilized in a quicksand-like mix of

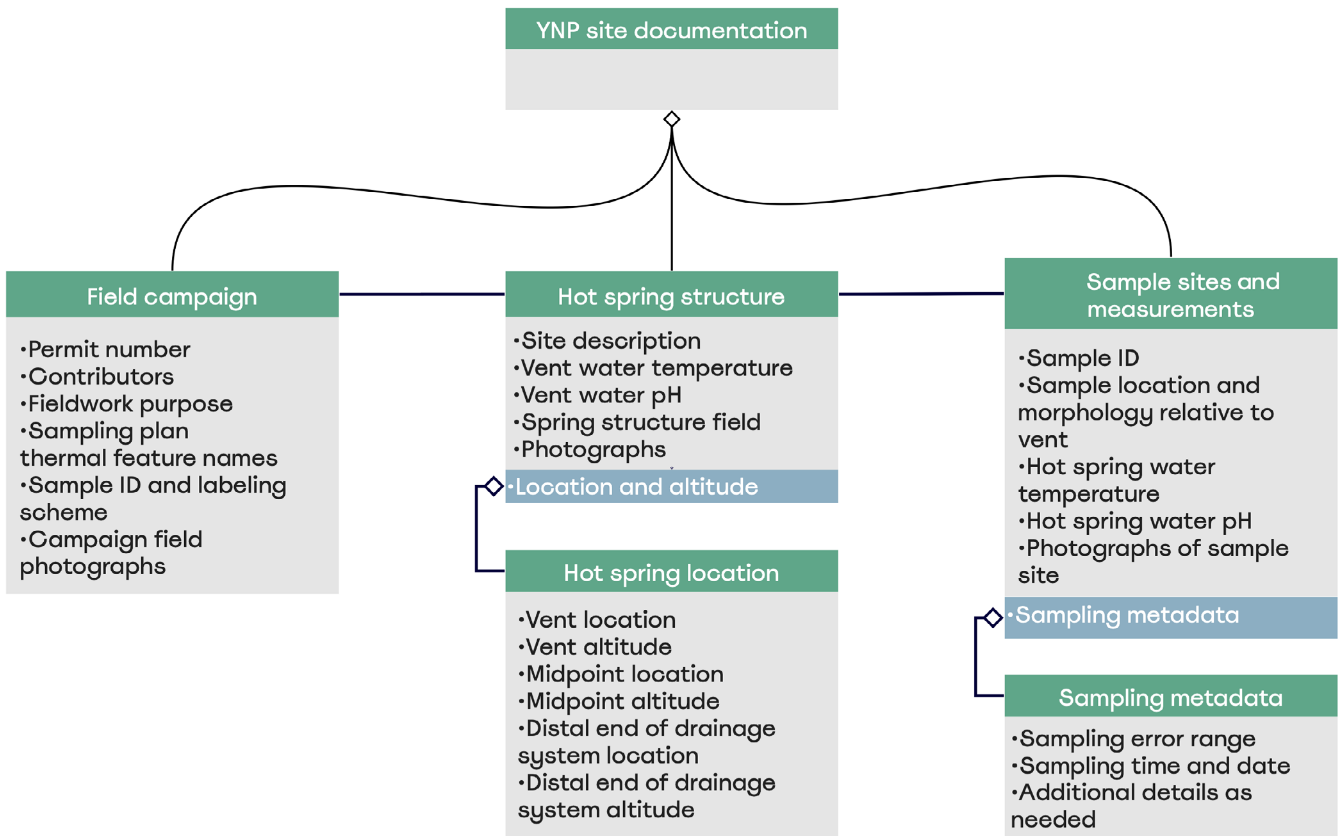

FIGURE 1 The YNP minimum information framework

oil and sediment at La Brea. An estimated 3–4 million fossils, ranging from microfossils to mammoth bones, have been recovered (La Brea Tar Pits FAQs, 2015). Excavation is ongoing at La Brea, making it a popular research locality for paleontologists, as well as a popular tourist destination; the excavation sites are all in a public park and viewable to passers-by. The fossil collections are housed at the La Brea Tar Pits Museum, which is located in the same city block as the fossil deposits. As one staff member described it: "[I]t's a site museum. It's all here. We're sitting on the environment; we're sitting on the fossils. … [W]e have this opportunity to look at our field site all the time" (RLB-Exc-1).

The resource managers at La Brea include the curatorial staff at the museum, a team of staff and volunteer excavators, and fossil preparators (who clean and curate the fossils to be added to the museum collection). As at Yellowstone, the researchers include the visiting researchers who come to La Brea to collect data. While La Brea resource managers must approve visiting researchers, the application process is less formal than at Yellowstone. Resource managers at La Brea play a much more involved role in data collection, as well, described in the next section.

4.2.1 | Resource manager roles: Research support and data curation

La Brea differs from Yellowstone in that the primary "data" from the site—the fossils—are excavated and curated by resource managers rather than visiting researchers. Researchers visiting the site are typically there to work with established specimen collections—not to collect their own physical samples. Consequently, resource managers at La Brea are much more involved in developing data collection and curation standards than the resource managers at Yellowstone. Of major concern is an ongoing curatorial backlog; the rate of excavation has long exceeded the lab's capacity to clean and curate bones, and there is a years-long backlog of fossils awaiting curation. Thus, resource managers are invested in developing cost- and time-efficient workflows. One collection manager summarized this ethic: "Work as you go, and don't create a backlog. That is my goal" (RLB-CM-1).

Like their counterparts at Yellowstone, La Brea resource managers were invested in making the site accessible to the public. But rather than trying to hide scientific work from view, the museum built exhibits around the research at the site—including a "fishbowl" lab in the middle of the museum, where visitors can watch fossil preparation as it is done, and interpretive signs around the excavation site. The resource managers saw this outreach and interpretive work as an important part of their job (while also expressing some exhaustion at always being on display).

4.2.2 | Resource manager data needs

As at Yellowstone, the resource managers at La Brea valued the site as a home for research, education, and tourism and are consequently invested in supporting the site's long-term sustainability and accessibility. However, because much of La Brea's data—the fossils themselves—are effectively managed as part of the site, La Brea resource managers additionally valued their ability to efficiently make that data usable to themselves and to others. They particularly valued the following data types:

Inventory. The resource managers said that simply "knowing what we have" is of high importance. Because of the current backlog of fossils awaiting curation, they have long been forced to estimate their holdings, which makes future space and work planning challenging.

Preparation process records. In the past, the lab used one consistent method in preparing fossils, and consequently, there was little need to keep individual preparation (prep) records. However, the prep lab recently changed its methods, so the resource managers needed to track what they were doing and create retrospective documentation for the existing collection. This is particularly important given recent interest in the collection by researchers doing proteomics work; certain prep methods—and even certain storage methods—might impact protein structures in the bones:

> When I showed [the proteomics researcher] the drawer of insects, he was like "Oh, are those gelatin capsules?" I said, "Yeah." He goes, "Pig protein." So, those are going to be problematic when we try and get proteins out of the insects … And, you know, he said, "Don't change your practice right now," and he said it's just important for us to know, and think about this, because [proteomics] is such a young field. (RLB-CM-1)

This same proteomics researcher also told the collection managers that some of the solvents being used might impact protein extraction, but it was unclear how. Thus, it remains important that these different treatments be documented clearly for future use.

Connections between data. Many of the La Brea data are distributed through different data systems and information structures: multiple catalog ledgers, multiple databases and spreadsheets, even multiple strategies for organizing physical specimens. Additionally, a great number of relevant data are published in academic journals or to repositories such as Dryad. The resource managers did not necessarily need all these data stored at La Brea, but (like the Yellowstone resource managers) they said that they needed to record the connections among these data points.

Geological context. Almost all resource managers said they wished they had more geological context about their current excavation or felt that they should be collecting more data about the geological stratigraphy, using best practices for geological fieldwork as they excavated.

Radiocarbon dates. Radiocarbon dates have long been used to date the age of the pits at La Brea. In the past, radiocarbon dates were generated by researchers using the fossils for their own projects and were published in journal articles. However, the museum is increasingly doing its own radiocarbon dating. Further collection of these dates, and correlation of the dates of the bones with their location within each pit, will be important for the La Brea resource managers going forward.

4.2.3 | Researcher roles and values

Whereas Yellowstone researchers expressed so/me frustration with the restrictions imposed by the NPS permitting process, the La Brea researchers were all quite understanding of the need for permitting and data management guidelines at the site. They all understood that the collections staff had to, at times, make strategic decisions about what to prioritize. Thus, the researchers saw themselves almost as guests at the site and wished to defer to the collections staff about how best to curate and manage the materials.

Researchers said that they valued La Brea because of the size, diversity, and breadth of its collections, as well as the range of studies that the site supports. They noted that at La Brea, many kinds of research are possible that simply are not possible at other fossil sites, particularly ecosystem- or population-scale studies, which are dependent on having access to a well-curated and complete collection of data related to a site. Several researchers also said that La Brea was unique for its human and curatorial infrastructure; the presence of a full-time data collection and curation staff makes the site much more appealing as a research site. As one researcher working on a pollen project at La Brea explained: “What's interesting is that you actually have a lot of people who are working on many different aspects of the system. And because it's associated with a museum, there can be some coordination with their different efforts, right” (RLB-Rsch-4).

4.2.4 | Researcher data needs

Researchers described the following as important for use:

Access to collections data. Researchers were satisfied with the level of access the collections staff provided for them, though several noted that they wished for an online database they could query on their own; researchers sometimes needed to check catalog information about the specimens they had worked with, and they disliked having to “bother” the collections managers for this, given that the collection managers were so busy.

Information about preparation history. Researchers echoed the resource managers here, saying that there is a growing need to understand how, exactly, fossils were cleaned, prepared, and repaired. This was particularly important for researchers doing molecular work (proteomics, DNA sequencing).

Consistently formatted and easily interpretable location data. All researchers said that knowing roughly where a fossil was taken from in a deposit was important (though some disagreed about how granular location data needed to be).

Radiocarbon dates correlated with locations. Researchers broadly agreed that radiocarbon dates tied to precise geolocations within each deposit are critical for their work. This is related to the need for precise location data, as described.

Ability to see other researchers' data. Similar to the research managers' desire to see connections among data, and the Yellowstone stakeholders' desire to see a big-picture view of research activities, La Brea researchers said they would generally like to see what data had been taken from individual specimens, and would like to be able to quickly look up publications in which a specimen had been mentioned. As one explained:

> There have been a lot of different researchers who have gone in and done different studies. ... but that data's held nowhere at La Brea. Also, when different people study different— like take measurements on fossils, a lot of times, those data, those individual specimen data aren't published ... Well how cool would it be if I looked at those same ones, that [another researcher] had done all these measurements of, and was able to, you know, use those data, right? (RLB-Rsch-9)

4.2.5 | La Brea MIF

As in Yellowstone, I worked with La Brea collection staff over the course of a year to develop an MIF in a similar vein as the MIF developed at Yellowstone, which models information about a site's structure as a core class (Figure 2). The three MIF classes here are:

1. The *Excavation Project* class. Just as the Yellowstone Field Campaign class includes information about a given field project, this class includes metadata about an excavation project overall, such as its motivations, contributors/investigators, and start and end dates. Information in this class supports later data integration by researchers, and ongoing management and monitoring by resource managers.
2. The *Fossil Deposit* class. This class documents the physical structure of the fossil deposits at La Brea. Curating data in this class would require a synthesis of the data collected by excavators as well as radiocarbon dates, which are typically generated by researchers. Participants thought that ongoing synthesis of this type would help make collections more usable in the long term.
3. The *Fossils and Samples* class. This class documents all information about individual fossils and samples collected at La Brea. Fossils and samples are split into three subclasses, depending on how much detailed positional and identification data need to be collected with each fossil in the field.

As at Yellowstone, the data about the scientific project in general are primarily needed by the resource managers, whereas the information about the site and its observations is somewhat evenly needed by both resource managers and researchers alike. Again, the information from these different classes is centered around, and aggregated through, information about the site and its

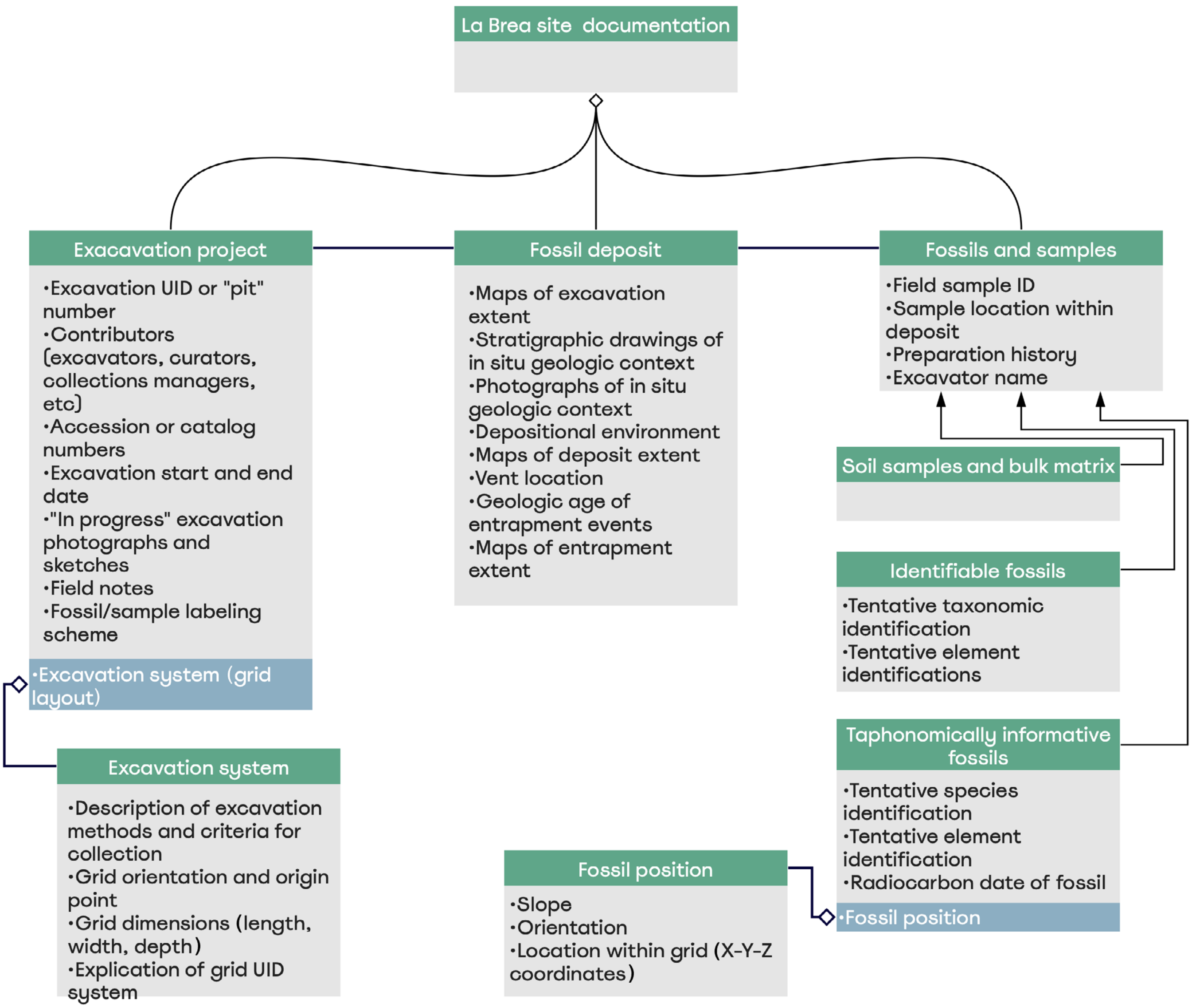

FIGURE 2 The La Brea information framework

TABLE 2 Comparison of role, values, and data needs across sites and stakeholder groups

	YNP resource managers	YNP researchers	La Brea resource managers	La Brea researchers
Roles	Administrative data collection, site mgmt., researcher mgmt., some data mgmt	Data collection, analysis, publishing, reporting	Data collection (fossils, field data), data mgmt., research mgmt., site mgmt.	Data collection (measurements, radiocarbon dates), analysis, publishing
Values and priorities	Ability to strategically plan for future research, awareness of research projects, efficiency, not having to police people	Accessibility of the site, ability to complete work independently without excessive oversight, precision, integrative reuse	Efficiency, avoiding backlog, protecting fossils, and other resources	Fossil collection breadth, integrative reuse, specimen diversity, and accessibility
Key data	"Larger context" of research activities (methods, data management plans), connections between silos	Genetic data, basic geochemistry about hot spring, photographs, methods metadata, access to other researchers' published data	Larger context of the collections overall (inventory, preparation history), geological context, radiocarbon dates	Radiocarbon dates, location information, information about prep history, access to other researchers' published data

changing structure over time. In developing this MIF, La Brea resource managers noted that the information about the site's structure was crucial to their work in managing and understanding the site, yet this is not currently curated in one consistent place.

4.3 | Comparison across sites and stakeholders

Across sites, stakeholder groups held some common values (summarized in Table 2). Resource managers valued their ability to protect the sites they manage and make them accessible for both research and non-research uses. They also valued the efficiency and effectiveness of resource management policies and procedures. At Yellowstone, this meant that they did not wish to adopt data sharing or archiving policies that they could not reasonably enforce or manage; at La Brea, this meant that they wished to avoid any collection or curation work that would increase an existing curatorial backlog. Finally, they all valued having a big-picture view of research activities over time; this would allow them to plan curatorial workflows, better support researchers, and monitor the impact of research activities on the site and related resources.

Researchers at both sites similarly valued this big-picture view of research activities, though largely to support new integrative studies or to help them position their own work within prior work. Yellowstone researchers were eager for systems that would make data integration easier and more streamlined. La Brea researchers also valued the accessibility of the site and its associated data collection. At Yellowstone, this largely entailed supporting access to the hot springs themselves, whereas at La Brea, this meant providing access to the specimen collections and their associated data.

5 | DISCUSSION

Through the cases above, I described a unique curatorial context with unique curatorial practices: the scientifically significant site. I found that the curatorial work of resource managers at these sites is quite distinct from that of data curators at other organizations. Additionally, the site-based researchers I spoke with have a strong interest in, and vision for, integrative reuse of data at these sites, even though these data are often not yet accessible. I discuss these findings further below. I also discuss how my research contributes to a framework to support curation and integrative reuse of site-based data. I conclude by reflecting on the need for further work to understand the scope and range of site-based data practices.

5.1 | Scientifically significant sites: Hubs for curation, research coordination, and recreation

I originally scoped scientifically significant sites as those with (a) special protection and/or administration by a governing body, and (b) the ongoing curation of associated specimen/data collections. I found an additional characteristic of scientifically significant sites: they are sites of curation *and* research coordination *and* recreation.

Both La Brea and Yellowstone were described as research hubs by participants, where the co-location of natural phenomena, research infrastructure, and resource management staff allowed for a much more iterative and effective scientific process. Yellowstone was regarded as a "living laboratory," and La Brea a place where researchers could access a wealth of resources and easily contextualize their work with others'. At both sites, though, the resource managers were viewed as almost as important as the natural sites themselves. They provided advice on what sites or specimens to use, provided context and institutional memory about the site, and made it possible for the researchers to make the most of their time. Dedicated resource management staff is critical to a site's usefulness—and therefore, significance—as a research locality.

Additionally, both Yellowstone and La Brea are popular tourist destinations. Resource managers at both sites struggled with how to manage their service to the public with their service to scientific researchers. At Yellowstone, this led to efforts to keep scientific research physically separate—and out of view—from public exhibits, whereas at La Brea, this led to absorption of the ongoing scientific work into a central feature of the site. Some of this reflects the different of each site. La Brea is geographically and organizationally much smaller, and it is much easier to build permanent fencing around the excavation sites to ensure that tourists do not attempt to join the excavators in their work and wind up stuck in the pits themselves. At Yellowstone, the thermal features are so abundant—and so dangerous—that it is understandable that resource managers do not want scientists to model the risky of behavior of getting up close and personal with a hot spring. Further work is needed to better understand how these different approaches to communicating (or obscuring) research to the public impacts the curatorial and research missions of the sites. Prior work studying the use of National Parks like Yellowstone has focused on their use by the general public, not researchers (e.g., Manning et al., 2017; Miller et al., 2017; Scott & Lee, 2018). Less research has focused on their use to researchers, let alone the tension between research use and recreational use. Understanding these varied uses will be critical to building effective research infrastructure.

5.2 | Facilitating research coordination and integrative reuse

Participants at my sites exhibited unique data practices, which complicate prior accounts of how researchers and data curators participate in the scientific process. First, I found that resource managers have their own data needs to facilitate the stewardship of the site, its collections, and its natural features. At Yellowstone, this "data" includes the annual reports and permits. At La Brea, this includes details about the fossils' excavation and preparation history. Several recent prior studies have shown the importance of studying the data practices of information professionals, and treating them as data users in their own right (e.g., Acker, 2020; Downey et al., 2019; Thomer, Weber, & Twidale, 2018). Here, understanding the data needs of resource managers is critical in designing effective infrastructures for them and their sites.

Second, I find that the resource managers at my sites were highly involved in research activities from their very inception—from the beginning of the "data lifecycle" (Ball, 2012; Higgins, 2008). Many have argued that the involvement of domain-specific information professionals early in a project is critical to producing data that are well-curated and ready to share (Heidorn, 2011; Johnston et al., 2018; Palmer et al., 2013). Though I initially thought the resource managers would be well-positioned to encourage data curation best practices from the beginning of a project, I found that they instead felt they did not have the right to tell researchers what to do, and that they needed to prioritize their time on the management of the natural resources as opposed to information resources.

Some of this hesitance to guide data curation is likely due to a lack of administrative support, clear data curation policies, and data infrastructure. However, I believe some is also due to the tensions inherent in the resource managers' research coordination work. By "research coordination," I mean the permitting and gatekeeping that resource managers do to manage access to the site. This also includes the coordination work they do to ensure that research visits are scheduled appropriately, and that resources are available for study yet not overtaxed. This is a distinct role in the research process, different from that of post hoc data curation at institutions such as libraries, or dedicated data repositories like those within the LTER. Where these more traditional curatorial units are focused primarily on coordinating their own curatorial work (Darch et al., 2021; Karasti et al., 2006; Karasti & Baker, 2008), site-based resource managers must coordinate the research enterprise more broadly.

Because of their role in coordinating and gatekeeping research, resource managers are acutely aware that they have the potential to fundamentally shape what kind of research is done at a site. They worry that dictating data sharing and reporting guidelines would only increase that influence. This points to an important area for future research: how do data curation and research access policies at scientifically significant sites influence the scholarly work that is done at those sites? How can site administrators, research coordinators, and curation units work together effectively while supporting researcher autonomy?

Finally, I found that stakeholders confirmed my initial proposition that site-based centers are powerful because their data collection has uniquely high potential for integrative reuse. Researchers described a clear vision for how they could do systems-level research with the right data. However, I found that much of this potential is unrealized. Data from each site were still distributed among researchers' individual data holdings, the site's databases, and publications. Existing curatorial infrastructures and workflows were not set up to holistically curate data from the site.

In the process of creating MIFs with my participants, I found that curating data about the site itself—its structure, characteristics, and changes over time—is important in supporting integrative reuse. Yet, I also found that site infrastructures and workflows were not set up to support this curatorial work. I discuss this in the next subsection.

5.3 | Supporting integrative data reuse through site-based information frameworks

A primary goal of this project was to contribute to the development of infrastructure for site-based data curation. It was out-of-scope to develop actual data systems; rather, I developed information frameworks to guide later system development. Developing semantic frameworks is key to the future development of data infrastructure, particularly to support integrative data reuse in sciences with complex and heterogenous data collections (Fox et al., 2009; Fox & Hendler, 2009; Narock & Fox, 2012). Here, I additionally, used framework development to reveal important commonalities between the sites I studied.

For both sites, data infrastructure must be able to record changing qualities of the site itself over time. A site-based information framework can be simplified into three classes:

- *Collecting Events*: the people, processes, methods, and strategies used to collect data;
- *Site Structure*: the key natural features at a site at the time of data collection, the relationships between those features, and their relationship to sampling locations; and
- *Samples and Measurements*: the individual observations, measurements, samples, and specimens collected from a site, and their orientation or location within a site.

Site-based data infrastructures must account for each of these information classes to be effective. The need for well-curated information about the *physical characteristics and structure of the sites themselves over time* is a key finding of this work. Rather than simply recording geocoordinates of a data point's origination, more detailed information about the site's structure is needed for context and later integrative reuse. At Yellowstone, information about the site structure includes information that makes it possible to reconstruct a hot spring's physical extent, flowpath, and water chemistry. At La Brea, site structure is represented through information that makes it possible to reconstruct the geological age and stratigraphy of a fossil deposit, as well as the position of the fossils.

Part of the goal in calling site structure data out in a separate class is to highlight its importance for the curation and sustainability of these collections. Often, this information is treated as metadata, and therefore peripheral to primary observation data. Treating site structure data as its own class might help ensure that it is not lost. Additionally, modeling site structure data as its own class is crucial in providing a mechanism for integrative reuse. Individual observations and samples can be rooted and therefore integrated by tracing their provenance back to the natural site from which they came. This is how systems-wide studies in earth and environmental studies are already conducted; my framework translates this approach into the curatorial process.

An information framework alone is not enough, however. Standards must be enacted to be effective (Millerand & Bowker, 2009) and infrastructure sustainability is the product of ongoing maintenance processes (Eschenfelder & Shankar, 2017) rather than one-time system implementations. My site-based information framework should be viewed as a starting point for policy and infrastructure development. Curation workflows and data sharing policies need to be substantially revised to facilitate site-based data curation. At both sites, information about site structure was not always collected in a consistent way, not always stored at the site, and not always curated. At La Brea, curation has primarily focused on the fossils themselves, and not the pits as a whole; at Yellowstone, they still lack the personnel and cyberinfrastructure to collect all the data desired. At both sites, some important data are collected by researchers visiting the sites, and there are no mechanisms in place to get that data back to the site. Thus, supporting site-based data curation at both sites would require organizational changes.

6 | CONCLUSION AND FUTURE WORK

In this paper, I have described two cases of scientific research and data curation at scientifically significant sites. Through participatory action research, I detailed the roles, values, and priorities of the stakeholders at these sites. I also developed MIFs of the information needed to support integrative curation and reuse of the data at these sites. I found that scientifically significant sites are valued for their human infrastructure as well as their data infrastructure and natural features. I found that to support data curation at scientifically significant sites, more information about the site's structure and characteristics must be curated. However, to facilitate this curation, substantial changes to existing curatorial workflows are likely required. Future work is needed to better understand how to build these site-based curatorial infrastructures and to understand other features of scientifically significant sites.

The case studies presented in this paper contain the usual limitations of case studies; though rich, their findings are not necessarily predictive of curatorial or research arrangements of other sites. Further work is needed to explore the applicability of these findings to other scientifically significant sites, perhaps with a less geological focus (for instance, ecological sites), which might rely on very different data collection strategies and ways of knowing. An important first step in this direction could be better identifying the range and number of comparable sites of site-based data curation. Within the United States, they may include:

- *National and state parks, particularly those with robust research programs and scientific collections*. Within the United States, about half of the 432 NPS units (those with an ecological, geological, or biodiversity focus rather than historical or cultural) likely fit the criteria of scientifically significant sites (National Park System [U.S. National Park Service], n.d.). There are over 6,600 state parks in the United States (Walls, n.d.), and while every state supports scientific research in its parks, it is unclear how many have associated data collections, and where those collections are managed.
- *Federal, state, county, and university-sponsored ecological/biological research stations or observatories*. The LTER sites are the most well-known examples of these; there are 28 of such sites (Sites Archive, n.d.). The Organization of Biological Field Stations lists 183 sites in their directory (OBFS—What Is a Field Station?, n.d.). The United States Geological Survey also runs dozens of site-based science centers, observatories, and field stations. Again, it is hard to know how many of these have data or specimen collections. Also, idiosyncratic sites like the La Brea Tar Pits, which exist independently of a broader network, are harder to identify and account for.
- *Botanical gardens, and arboretums*. Many gardens have research programs as well as there are over 400 accredited arboreta around the world, at least 64 of which feature curators and research staff (Arbnet|Morton Register, n.d.).

Thus, even conservative estimates of scientifically significant sites within the US alone could be in the thousands—and exponentially more globally.

The scientifically significant sites at the heart of this study are hugely important to scientific research and popular culture. Many of these sites are the home of unique long-term data about our planet's ecology, biodiversity, and geology. Site-based data curation could be crucial to creating more robust data collections for these fields—and it could be necessary to facilitate integrative, longitudinal study of how our planet is changing in response to anthropogenic forces and warming climates. Finally, I hope that better describing the work of site-based resource managers can act as a first step toward better supporting this work—whether through better infrastructure, more staffing, or more recognition. Their stewardship of these sites is critical, yet too often invisible.

ACKNOWLEDGMENTS

This research was funded by IMLS grant LG-06-12-0706-12. My work would not have been possible without the guidance, brilliance, and patience of my PhD advisor Carole Palmer, with whom I originally framed this study. Thanks to Bruce Fouke, and my committee members Allen Renear, Mike Twidale, and Bryan Heidorn for support throughout this research and my dissertation. Thank you to the many participants whose insights informed and directed to this work, and to my collaborators on the SBDC project, particularly Karen Wickett. Thank you to the JASIST reviewers and editors for their helpful comments. Finally: this paper is dedicated to Dr. Peter Fox, who encouraged me to push this project forward, and whose wit and wisdom are missed.

ORCID

Andrea K. Thomer https://orcid.org/0000-0001-6238-3498

REFERENCES

Acker, A. (2020). Emulation encounters: Software preservation in libraries, archives, and museums. *Proceedings of the Association for Information Science and Technology*, *57*(1), e279. https://doi.org/10.1002/pra2.279

Akmon, D., Zimmerman, A., Daniels, M., & Hedstrom, M. (2011). The application of archival concepts to a data-intensive environment: Working with scientists to understand data management and preservation needs. *Archival Science*, *11*(3), 329–348. https://doi.org/10.1007/s10502-011-9151-4

Arbnet|Morton Register. (n.d.). http://arbnet.org/morton-register-arboreta

Ball, A. (2012). Review of data management lifecycle models. University of Bath. http://opus.bath.ac.uk/28587/

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, *63*(6), 1059–1078. https://doi.org/10.1002/asi.22634

Borgman, C. L., Wallis, J. C., & Enyedy, N. (2007). Little science confronts the data deluge: Habitat ecology, embedded sensor networks, and digital libraries. *International Journal on Digital Libraries*, *7*(1–2), 17–30. https://doi.org/10.1007/s00799-007-0022-9

Brunt, J. W., & Michener, W. K. (2009). The resource discovery initiative for field stations: Enhancing data management at North American Biological Field Stations. *Bioscience*, *59*(6), 482–487. https://doi.org/10.1025/bio.2009.59.6.6

Buttigieg, P., Morrison, N., Smith, B., Mungall, C. J., Lewis, S. E., & the ENVO Consortium. (2013). The environment ontology: Contextualising biological and biomedical entities. *Journal of Biomedical Semantics*, *4*(1), 43. https://doi.org/10.1186/2041-1480-4-43

Carlson, S. (2006). Lost in a sea of science data. *The Chronicle of Higher Education*, *52*(42), A35. https://www.chronicle.com/article/lost-in-a-sea-of-science-data/

Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *368*(1926), 4023–4038. https://doi.org/10.1098/rsta.2010.0165

de Solla Price, D. J. (1986). *Little science, big science—And beyond*. Columbia University Press.

Darch, P. T., & Borgman, C. L. (2016). Ship space to database: Emerging infrastructures for studies of the deep subseafloor biosphere. *PeerJ Computer Science*, *2*, e97. https://doi.org/10.7717/peerj-cs.97

Darch, P. T., Sands, A. E., Borgman, C. L., & Golshan, M. S. (2021). Do the stars align?: Stakeholders and strategies in libraries' curation of an astronomy dataset. *Journal of the Association for Information Science and Technology*, *72*(2), 239–252. https://doi.org/10.1002/asi.24392

Dennis, A., Wixom, B. H., Tegarden, D. P., & Seeman, E. (2015). *System analysis & design: An object-oriented approach with UML* (5th ed.). Wiley.

Downey, G., Eschenfelder, K. R., & Shankar, K. (2019). Talking about metadata labor: Social science data archives, professional data librarians, and the founding of IASSIST. In W. Aspray (Ed.), *Historical studies in computing, information, and society: Insights from the flatiron lectures* (pp. 83–113). Springer International Publishing. https://doi.org/10.1007/978-3-030-18955-6_5

Eschenfelder, K., & Shankar, K. (2017). Organizational resilience in data archives: Three case studies in social science data archives. *Data Science Journal*, *16*, 12. https://doi.org/10.5334/dsj-2017-012

Faniel, I. M., & Jacobsen, T. E. (2010). Reusing scientific data: How earthquake engineering researchers assess the reusability of Colleagues' data. *Computer Supported Cooperative Work (CSCW)*, *19*(3), 355–375. https://doi.org/10.1007/s10606-010-9117-8

Fouke, B. W. (2011). Hot-spring systems geobiology: Abiotic and biotic influences on travertine formation at mammoth hot springs, Yellowstone National Park, USA.Sedimentology, *58*, 170–219. https://doi.org/10.1111/j.1365-3091.2010.01209.x

Fox, P., & Hendler, J. (2009). Semantic eScience: Encoding meaning in next-generation digitally enhanced science. In T. Hey, S. Tansley, & K. Tolle (Eds.), *The fourth paradigm: Data-intensive scientific discovery* (pp. 147–152). Microsoft Research.

Fox, P., McGuinness, D. L., Cinquini, L., West, P., Garcia, J., Benedict, J. L., & Middleton, D. (2009). Ontology-supported scientific data frameworks: The virtual solar-terrestrial observatory experience. *Computers & Geosciences*, *35*(4), 724–738. https://doi.org/10.1016/j.cageo.2007.12.019

Frank, R. D., Kriesberg, A., Yakel, E., & Faniel, I. M. (2015). Looting hoards of gold and poaching spotted owls: Data confidentiality among archaeologists & zoologists. *Proceedings of the Association for Information Science and Technology*, *52*(1), 1–10. https://doi.org/10.1002/pra2.2015.145052010037

Hayes, G. R. (2011). The relationship of action research to human-computer interaction. *ACM Transactions on Computer-Human Interaction*, *18*(3), 1–20. https://doi.org/10.1145/1993060.1993065

Heidorn, P. B. (2008). Shedding light on the dark data in the long tail of science. *Library Trends*, *57*(2), 280–299. https://doi.org/10.1353/lib.0.0036

Heidorn, P. B. (2011). The emerging role of libraries in data curation and E-science. *Journal of Library Administration*, *51*(7–8), 662–672. https://doi.org/10.1080/01930826.2011.601269

Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation*, *3*, 453. https://doi.org/10.1145/1378889.1378998

Johnston, L. R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., Stewart, C., Blake, M., Herndon, J., McGeary, T. M., & Hull, E. (2018). Data curation network: A cross-institutional staffing model for curating research data. *International Journal of Digital Curation*, *13*(1), 125–140. https://doi.org/10.2218/ijdc.v13i1.616

Karasti, H. (2014). Infrastructuring in participatory design. *Proceedings of the 13th Participatory Design Conference: Research Papers*, *1*, 141–150. https://doi.org/10.1145/2661435.2661450

Karasti, H., & Baker, K. S. (2008). Digital data practices and the long term ecological research program growing global. *International Journal of Digital Curation*, *3*(2), 42–58. https://doi.org/10.2218/ijdc.v3i2.57

Karasti, H., Baker, K. S., & Halkola, E. (2006). Enriching the notion of data curation in E-science: Data managing and information Infrastructuring in the long term ecological research (LTER) network. *Computer Supported Cooperative Work (CSCW)*, *15*(4), 321–358. https://doi.org/10.1007/s10606-006-9023-2

Kent, W. (1978). *Data and reality: Basic assumptions in data processing reconsidered*. North-Holland Pub. Co.; sole distributors for the U.S.A. and Canada Elsevier/North-Holland.

Khan, N., Thelwall, M., & Kousha, K. (2019). Data citation and reuse practice in biodiversity—Challenges of adopting a standard citation model. In G. Catalano, C. Daraio, M. Gregori, H. F. Moed, & G. Ruocco (Eds.), *17th International Conference on Scientometrics & Infometrics, ISSI2019: Proceedings, volume I* (pp. 1220–1225). International Society for Scientometrics and Informetrics. https://wlv.openrepository.com/handle/2436/623005

Kowalczyk, S., & Shankar, K. (2011). Data sharing in the sciences. *Annual Review of Information Science and Technology*, *45*(1), 247–294. https://doi.org/10.1002/aris.2011.1440450113

La Brea Tar Pits FAQs. (2015, June 24). La Brea Tarpits and Museum. https://tarpits.org/la-brea-tar-pits/faqs

Lewin, K. (1943). Forces behind food habits and methods of change. *Bulletin of the National Resource Council*, *108*, 35–65.

Manning, R. E., Anderson, L. E., & Pettengill, P. (2017). *Managing outdoor recreation* (2nd ed.). Case Studies in the National Parks. CABI.

Michener, W. K., Brunt, J. W., Helly, J. J., Kirchner, T. B., & Stafford, S. G. (1997). Nongeospatial metadata for the ecological sciences. *Ecological Applications*, *7*(1), 330–342. https://doi.org/10.1890/1051-0761(1997)007[0330:NMFTES]2.0.CO;2

Miller, Z. D., Fefer, J. P., Kraja, A., Lash, B., & Freimund, W. (2017). Perspectives on visitor use management in the national parks. *The George Wright Forum*, *34*(1), 37–44.

Millerand, F., & Baker, K. S. (2010). Who are the users? Who are the developers? Webs of users and developers in the development process of a technical standard. *Information Systems Journal*, *20*(2), 137–161. https://doi.org/10.1111/j.1365-2575.2009.00338.x

Millerand, F., & Bowker, G. C. (2009). Trajectories and enactment in the life of an ontology. In S. L. Star & M. Lampland (Eds.), *Standards and their stories* (pp. 149–165). Cornell University Press.

Narock, T., & Fox, P. (2012). From science to e-science to semantic e-science: A heliophysics case study. *Computers & Geosciences*, *46*, 248–254. https://doi.org/10.1016/j.cageo.2011.11.018

National Park System (U.S. National Park Service). (n.d.). https://www.nps.gov/aboutus/national-park-system.htm

OBFS—What is a Field Station? (n.d.). https://www.obfs.org/what-s-a-field-station

Palmer, C. L., Thomer, A. K., Baker, K. S., Wickett, K. M., Hendrix, C. L., Rodman, A., Sigler, S., & Fouke, B. W. (2017). Site-based data curation based on hot spring geobiology. *PLoS One*, *12*(3), e0172090. https://doi.org/10.1371/journal.pone.0172090

Palmer, C. L., Weber, N. M., & Cragin, M. H. (2011). The analytic potential of scientific data: Understanding re-use value. *Proceedings of the American Society for Information Science and Technology*, *48*, 1–10. https://doi.org/10.1002/meet.2011.14504801174/full

Palmer, C. L., Weber, N. M., Renear, A., & Muñoz, T. (2013). Foundations of data curation: The pedagogy and practice of "purposeful work" with research data. *Archive Journal*. https://www.archivejournal.net/essays/foundations-of-data-curation-the-pedagogy-and-practice-of-purposeful-work-with-research-data/

Pasquetto, I., Randles, B., & Borgman, C. (2017). On the reuse of scientific data. *Data Science Journal*, *16*, 8. https://doi.org/10.5334/dsj-2017-008

Peters, M., & Robinson, V. (1984). The origins and status of action research. *The Journal of Applied Behavioral Science*, *20*(2), 113–124. https://doi.org/10.1177/002188638402000203

Ribes, D., & Finholt, T. (2009). The long now of technology infrastructure: Articulating tensions in development. *Journal of the Association for Information Systems*, *10*(5), 375–398.

Sacchi, S. & Wickett, K. M. (2012). Taking modeling seriously [in digital curation]. IPres: Research Challenges in Digital Preservation, 14–16.

Scott, D., & Lee, K. J. J. (2018). People of color and their constraints to National Parks Visitation. *The George Wright Forum*, *35*(1), 73–82.

Sites Archive. (n.d.). LTER. https://lternet.edu/site/

Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., Parsons, M., Robinson, E., & Wyborn, L. (2019). Make scientific data FAIR. *Nature*, *570*(7759), 27–29. https://doi.org/10.1038/d41586-019-01720-7

Strasser, B. J. (2012). Data-driven sciences: From wonder cabinets to electronic databases. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(1), 85–87. https://doi.org/10.1016/j.shpsc.2011.10.009

Tenopir, C., Christian, L., Allard, S., & Borycz, J. (2018). Research data sharing: Practices and attitudes of geophysicists. *Earth and Space Science*, *5*(12), 891–902. https://doi.org/10.1029/2018EA000461

Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorf, R., & Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLoS One*, *15*(3), e0229003. https://doi.org/10.1371/journal.pone.0229003

Thomer, A. K. (2017). *Site-based data curation: Bridging data collection protocols and curatorial processes at scientifically significant sites*. (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign. http://hdl.handle.net/2142/98372

Thomer, A. K., Palmer, C. L., Wickett, K. M., Baker, K. S., Jett, J. G., Dilauro, T., Fouke, B. W., Asangba, A. E., Rodman, A., & Choudhury, G. S. (2014). *Data curation for Geobiology at Yellowstone National Park: Report from Workshop Held April 16-17, 2013* (p. 41). Center for Informatics Research in Science and Scholarship. http://hdl.handle.net/2142/47070

Thomer, A. K., Weber, N. M., & Twidale, M. B. (2018). Supporting the long-term curation and migration of natural history museum collections databases. *Proceedings of the Association for Information Science and Technology*, *55*(1), 504–513. https://doi.org/10.1002/pra2.2018.14505501055

Thomer, A. K., Wickett, K. M., Baker, K. S., Fouke, B. W., & Palmer, C. L. (2018). Documenting provenance in non-computational workflows: Research process models based on geobiology fieldwork in Yellowstone National Park. *Journal of the Association for Information Science and Technology*, *69*(10), 1234–1245. https://doi.org/10.1002/asi.24039

Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS One*, *8*(7), e67332. https://doi.org/10.1371/journal.pone.0067332

Walls, M. (n.d.). Parks and Recreation in the United States: State Park Systems. 14.

Walls, R. L., Deck, J., Guralnick, R., Baskauf, S., Beaman, R., Blum, S., Bowers, S., Buttigieg, P. L., Davies, N., Endresen, D., Gandolfo, M. A., Hanner, R., Janning, A., Krishtalka, L., Matsunaga, A., Midford, P., Morrison, N., Tuama, É. Ó., Schildhauer, M., … Wooley, J. (2014). Semantics in support of biodiversity knowledge discovery: An introduction to the biological collections ontology and related ontologies. *PLoS One*, *9*(3), e89606. https://doi.org/10.1371/journal.pone.0089606

Weber, N. M., Baker, K. S., Thomer, A. K., Chao, T. C., & Palmer, C. L. (2013). Value and context in data use: Domain analysis revisited. *Proceedings of the American Society for Information Science and Technology*, *49*, 1–10. https://doi.org/10.1002/meet.14504901168

Wickett, K. M., Sacchi, S., Dubin, D., & Renear, A. H. (2012). Identifying content and levels of representation in scientific data. *Proceedings of the American Society for Information Science and Technology*, *49*(1), 1–10. https://doi.org/10.1002/meet.14504901199

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., & Vieglais, D. (2012). Darwin core: An evolving community-developed biodiversity data standard. *PLoS One*, *7*(1), e29715. https://doi.org/10.1371/journal.pone.0029715

Yan, A., Huang, C., Lee, J.-S., & Palmer, C. L. (2020). Cross-disciplinary data practices in earth system science: Aligning services with reuse and reproducibility priorities. *Proceedings of the Association for Information Science and Technology*, *57*(1), e218. https://doi.org/10.1002/pra2.218

Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). SAGE Publications Ltd.

Zimmerman, A. S. (2008). New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology & Human Values*, *33*(5), 631–652. https://doi.org/10.1177/0162243907306704

How to cite this article: Thomer, A. K. (2022). Integrative data reuse at scientifically significant sites: Case studies at Yellowstone National Park and the La Brea Tar Pits. *Journal of the Association for Information Science and Technology*, 1–16. https://doi.org/10.1002/asi.24620