Curating for Convergence: Data Stewardship for Interdisciplinary Inquiry

Carole L. Palmer, Melissa H. Cragin

Library Trends, Volume 71, Number 1, August 2022, pp. 113-131 (Article)

Published by Johns Hopkins University Press
DOI: *https://doi.org/10.1353/lib.2023.0007*

Curating for Convergence: Data Stewardship for Interdisciplinary Inquiry

Carole L. Palmer and Melissa H. Cragin

Abstract

Advances in data infrastructure are often led by disciplinary initiatives aimed at innovation in federation and sharing of data and related research materials. In library and information science (LIS), the data services area has focused on data curation and stewardship to support description and deposit of data for access, reuse, and preservation. At the same time, solutions to societal grand challenges are thought to lie in convergence research, characterized by a problem-focused orientation and deep cross-disciplinary integration, requiring access to highly varied data sources with differing resolutions or scales. We argue that data curation and stewardship work in LIS should expand to foster convergence research based on a robust understanding of the dynamics of disciplinary and interdisciplinary research methods and practices. Highlighting unique contributions by Dr. Linda C. Smith to the field of LIS, we outline how her work illuminates problems that are core to current directions in convergence research. Drawing on advances in data infrastructure in the earth and geosciences and trends in qualitative domains, we emphasize the importance of metastructures and the necessary influence of disciplinary practice on principles, standards, and provisions for ethical use across the evolving data ecosystem.

Introduction

Much of the progress in data infrastructure in the United States has been led by disciplinary initiatives to archive and share digital data and support the advancement of informatics and data science methods. The recent turn toward convergence research elevates an integrative, problem-focused approach to grand challenges and development of cyberinfrastructure to

LIBRARY TRENDS, Vol. 71, No. 1, 2022 ("Library and Information Science, Interdisciplinary Perspectives: A Festschrift in Honor of Linda C. Smith," edited by Anita S. Coleman and Martha Kyrillidou), pp. 113–131.

support interdisciplinary discovery and innovation. With the growing emphasis on cross-disciplinary research approaches, data professionals need to be prepared to support both disciplinary and interdisciplinary research practices and the interplay between the two. In library and information science (LIS) that expertise is grounded in foundational work on information systems for research and knowledge production, an area where Linda Smith has made unique and prescient contributions to the field.

With the occasion of this festschrift, we examine, and celebrate, the long-standing value of Dr. Smith's scholarship on the potential of information systems to support discovery and consider how to build on these roots in LIS for convergence curation—professional data curation work that actively supports convergence research. We discuss our perspectives on curation for convergence by drawing on current directions in data infrastructure in the earth and geosciences, and other more qualitative domains, to illustrate the importance of metastructures, and the necessary influence of disciplinary practice on principles, standards, and provisions for secure and ethical use of data. Our emphasis is on the intersection of expertise for understanding the differences that make a difference, across disciplines or data communities, for the effective curation of data for convergence research.

Potential of Information Systems

Smith's body of work is perhaps most notable for shining a spotlight on essential information science constructs, often long before their importance was recognized within the field. At the same time, from the beginning of her long, illustrious career, Dr. Smith's scholarship confronted a fundamental information science problem: *How do we realize the potential of information systems for scientific discovery and innovation*? Much has changed in the information environment of science and scholarship in recent decades, as methods and materials of research transitioned to digital infrastructure and LIS extended its purview to the curation and stewardship of digital data. This perennial question about the potential of information, however, remains of central importance and, we believe, serves as a touchstone for how LIS should conceive of its role in the research enterprise and the creation of new knowledge.

Within her scholarly repertoire, three of Dr. Smith's early papers stand out in their astute targeting of areas of critical significance for the potential of information systems for discovery. The topics covered—interdisciplinary search, representation in information retrieval, and the Memex concept—foretell an interrelated set of core concerns that have since been deeply implicated in the development of contemporary data infrastructure and the recent convergence turn in the research enterprise, which are discussed further below.

Smith's 1974 student paper on interdisciplinary search was far ahead of its time as one of the first empirical investigations of the problem of cross-disciplinary access to information. Her insights on the loss of precision with the increase in recall are at the core of interdisciplinary search dynamics, and her conclusions on suggestive versus prescriptive search have retained validity in relation to the function of high impact information in scientific discovery (Palmer, Cragin, and Hogan 2007). From our vantage point today, it is easy to see the foresight of Smith's attention to tools for semantic alignment across disciplinary vocabularies, given the rapid growth in the number and scale of data repositories and increasing prominence of boundary-crossing inquiry.

With her 1981 paper, "Representation Issues in Information Retrieval System Design," Smith emphasizes the centrality of knowledge representation for artificial intelligence (AI) and anticipates the introduction of "data banks" into the base of information resources for retrieval. At that point in time in LIS, this focus on AI is rare, as is her strong user orientation. However, Smith's balanced treatment of representation of both content and context is especially significant from a data infrastructure perspective. The lack of contextual metadata is a widely recognized weakness in representation practices for research data (Faniel, Frank, and Yakel 2019; Faniel et al. 2013). Moreover, for cross-disciplinary research, valid reuse often depends on rich description of methods of data collection, processing, and analysis (Palmer et al. 2017; Yan et al. 2020). To promote understanding of the intellectual foundations of new interdisciplinary knowledge, convergence curation will necessarily address representation of the paths of data access and provenance across different disciplinary information systems.

The concept of convergence resonates strongly in Dr. Smith's (1980) synthetic review of how information science research was influenced by Vannevar Bush's seminal paper, "As We May Think" (1945). Her analysis underscores how Bush's idea of the Memex, as a tool for managing a great mass of collected information, animated the theme of "potentiality" of large-scale information systems within the discourse of information science. Much of the inspirational power of Memex for the field was harbored in the promise of exploiting "association trails" that could extend and amplify memory (Kochen 1964; Perry, Kent, and Berry 1956; Tague 1969) to stimulate discovery and innovation for local, personal applications. Written in an era when search was still primarily mediated and confined to bibliographic information, Smith's treatment looks ahead to consider databases for full-text search. More surprising for the time, it boldly suggests the inclusion of data sources in retrieval systems and draws attention to risks of prioritizing computing at the expense of "social, cultural, and spiritual aspects of human communication" (Smith 1980, 353).

The perspectives and insights from Smith's scholarship outlined above affirm the value of LIS as a metascience, promoted by Bates (1999) as a distinct value proposition inherent in the field. The themes of interdisciplinary search, knowledge representation, and the potential of large systems of interconnected information all draw on a kind of landscape expertise that encompasses and can be applied across all disciplines. However, as we discuss below, effectively bridging disciplines and promoting true cross-disciplinary diffusion also requires robust understanding of the methods and materials within disciplines. Throughout our narrative, we use the term "discipline" as shorthand for cohesive areas of research practice that include the concept of "data communities" (e.g., Cooper and Springer 2019).

Rise of Convergence Research

Convergence research is characterized as a transformative approach to solving grand challenges for science, technology, and society by changing the dynamics of knowledge and material production. The National Science Foundation's (NSF) "Convergence Reports and References" webpage briefs the long history of conceptual development: Following early theoretical work, the National Research Council produced a series of reports articulating the prospects for "bio-inspired" technologies to benefit the nation's health and prosperity based on new efforts at the intersection of biology and mathematics, computing, the physical sciences, and engineering (National Research Council 2005a, 2005b, 2008, 2009, 2010, 2014). Aims of the convergence movement are to catalyze novel methods and even launch new disciplines (Sharp and Langer 2011). In Sharp's AAAS president's address (2014), he quoted Susan Hockfield, then president of MIT, to illustrate the changes needed for convergent knowledge work: "Physicists gave engineers the electron, and they created the IT revolution. Biologists gave engineers the gene, and together they will create the future" (1470). That is, "convergence encompasses engineers and physical scientists applying their knowledge and tools to problems . . . in their own professional domains" (1470).

In recent years, initiatives have given way to new centers designed to streamline convergence research, housing scientists, engineers, and technologists together, and offering courses for students organized around problems or complex systems questions rather than content from individual disciplines. Examples include the Wyss Institute at Harvard (https://wyss.harvard.edu/how-we-work), the Connecticut Convergence Institute for Translation in Regenerative Engineering (https://health.uconn.edu/connecticut-convergence-institute/areas-of-focus), and the Koch Institute for Integrative Cancer Research at MIT (https://ki.mit.edu/about). Convergence is not simply about speeding the pace of discovery; it is also concerned with changing the way problems are viewed, the ways information work is performed, and how knowledge flows from discovery through product conception and design to market, or for public good. The NSF identified growing convergence research as one of the 10 Big Ideas in 2016, launching a program to fund projects that would require deep interdisciplinary collaborations to address complex problems of societal concern.

Convergence in the Geosciences

For geosciences, mobilization for convergence is rooted, in part, in the rise of geoinformatics. Generally, geoinformatics is the use of information science, computer science, and related engineering and technologies to gather, represent, organize, integrate, visualize, analyze, and reuse data related to earth and space systems. The geoinformatics community brings together experts from multiple domains to utilize evolving technologies to transform data-to-knowledge infrastructure, improving use of increasing stores of heterogeneous data, and utilizing the semantic web for machine readability (Sinha et al. 2010). Significantly, leading voices described the need for convergence-oriented work to build data models, develop crosswalks for vocabulary interoperability, and generate "information products for the non-specialist" (Fox and Hendler 2009, 150). They noted that this work requires "a peer relationship . . . between physical scientists and computer scientists, and between software engineers and data managers and data providers" and that both project-specific and community organized efforts are necessary (150).

Building on these approaches, the geoscience and AI communities, along with information science and allied technical communities, have outlined a research agenda for intelligent systems in the geosciences (Gil et al. 2019), which would be designed to address scientific, societal, and global concerns. Science and policy leaders are also calling for the development of convergence research in the geosciences, to address questions posed by scientists, engineers, policy makers, landowners and managers working to understand and plan for local and societal concerns, including changes in land use, regulatory requirements and oversight, and natural hazard management (McNutt 2022).

As indicated in Gil et al. (2019), interconnected knowledge representation tools (e.g., metadata standards that use domain vocabularies or formal ontologies, OCLC's WorldCat Entities service; OCLC n.d.) are foundational components of the intelligent systems needed to address such scientific challenges. One of these tools is a knowledge graph (KG), "a graphical representation of structured knowledge from the real world" (Ma 2022, 1), "associating domain-specific meanings" to specify relationships between real-world entities (Chaudhri et al. 2022, 17–18). Knowledge graphs in the form of "data graphs, taxonomies, and ontologies" serve as "information structures that enable access, integration, and use of vast" stores of data, making the domain meanings machine readable and useful for AI and machine learning approaches (Chaudhri et al. 2022, 26). Another description of knowledge graphs particularly relevant to our practice perspective adds the concept "data cultures." Janowicz (2021) writes, "These graphs combine scalable technologies, semantics, *and data cultures* to represent densely interconnected statements derived from structured or unstructured sources across domains in a reasonable way that is readable by humans and machines" (16, emphasis added). Notably, Ma (2022) considers knowledge graphs in the context of the data lifecycle and science production, stating, "associated works of KG connect the upstream work of knowledge engineering and representation, the midstream work of data curation and integration, and the downstream work of data analysis and result communication" (1).

Another initiative now underway at the NSF is the Convergence Accelerator program that is designed to stimulate research by linking grand challenge problems, cross-sector collaborations, and training in entrepreneurship (Baru et al. 2022). A track in this program is focused on open knowledge networks (OKNs), which are organized around a knowledge graph to leverage evolving technologies that support accumulation and linkage among and across siloed data to improve multisource search and query (Janowicz et al. 2022). For the user, these graphs serve as metastructures that can tilt the balance of their attention from data acquisition and processing to analysis and interpretation. Geoscience applications include the KnowWhereGraph project, for example, which is building a cross-domain environmental intelligence service to integrate geospatial data and tools with industry and public sector data to improve situational awareness for decision-makers (Janowicz et al. 2022). The Urban Flooding OKN (UF-OKN) team (Johnson et al. 2022) is building an urban multiplex inventory (UMI) that will capture data on built environments and systems, including buildings, infrastructure, and movement and interactions among goods, people, and services. Urban planners, services providers, and emergency responders currently need to draw information from multiple, often siloed, sources that were designed for particular administrative entities or business purposes. The UF-OKN project is developing an approach and a model for constructing UMIs that would scale to the continental level.

Yet significant curation and stewardship work is necessary to address current and underlying challenges related to data collection, representation and organization, interoperability, and integration. For hydrologists studying the properties and movement of water through earth systems, for example, the use of needed data is constrained by limited access, lack of information on uncertainty, difficulties using small or local data for modeling, lack of a comprehensive data inventory, and gaps in coverage (Wagener et al. 2021, 9). Realizing the potential of OKNs and related technologies needed for convergence research will require a cooperative effort from a range of experts, including those trained in LIS.

Positioning Curation for Convergence Research

As suggested by the knowledge graph case, the data curation and stewardship work needed to foster convergence research requires expertise that balances general approaches that support interoperability and federation with more specialized domain-specific knowledge related to the methods and cultures of research communities and disciplines. There are, however, many other areas in the development of data infrastructure and provision of data services that will need to respond to disciplinary practices and expectations, such as the application of the FAIR principles and standards for data integrity.

Discipline Responsive FAIR

Originally published in 2016, the FAIR Guiding Principles for scientific data management and stewardship (Wilkinson et al. 2016)—findability, accessibility, interoperability, and reusability—have been widely accepted as best practice across an array of fields and within data services in repositories and libraries. FAIR promotes access and reuse of data in compliance with the expectations of federal agencies and the emerging norms of many research communities. Adoption has been relatively rapid within a global research environment that increasingly values openness and interoperability, yet they are "ambiguous, domain respectful, and not synonymous with open"; they exist as a continuum for making discipline-aligned improvements toward FAIR (Goble 2020). More specifically, we assert that some modes of inquiry require customized interpretations of FAIR that respond to scholarly epistemologies and can reconcile competing data principles and governance priorities.

Discipline-responsive FAIR is particularly appropriate for qualitative modes of research, where evidence tends to be far from uniform and the interpretive approach requires deep and iterative exploration of data. Meaningful analysis in the qualitative social sciences, for example, is often based on the researcher's relationship to participants and the environment, and important contextual factors may introduce sensitivities related to research participants, sites, and other data sources and processes that emphasize the "relationality and constitutive character" of data (Feldman and Shaw 2019, 704). In response, stewardship of qualitative data needs to be guided by principles of *epistemically responsible reuse*, as articulated by Karcher et al. (2021), that prioritize contextualization of data and ensure ethical sharing over openness and interoperability.

In the humanities, there is growing acceptance of the FAIR principles, despite their genesis in the sciences. The 2020 All European Academies report tightly associates FAIR data with humanities sustainability (ALLEA

2020). However, as discussed by Tóth-Czifra (2020), the path to "FAIRification" of data in the arts and humanities is paved with complexities and challenges, including the deeply analogue past, legal liability concerns, and complicated provenance trails. Without the rich description that underlies data integrity in the humanities, "the FAIR principles of maximum reusability and interoperability cannot be achieved on an epistemic level, even if they can be achieved technically" (Tóth-Czifra 2020, 245). As shown with early efforts to federate digital cultural heritage materials at a national scale, "contextual mass" is more important than critical mass to ensure standardization and scaling does not lose contextual meaning vital to how scholars use research collections (Palmer, Zavalina, and Fenlon 2010). At the same time, the humanities have some inherent advantages in regard to the time-consuming work involved in the preparation of FAIR data. As discussed by Cremer et al. (2021), the data types created through historical methods—"validation, provenance, transcriptions, editions, entity extraction, contextualisation, registries, etc.—could directly improve the reuse potential of the sources in other research contexts" (170).

In some areas of research, disciplinary or cultural responsiveness will mean displacing FAIR as the primary guiding principles, especially in domains when protections for sensitive data, the well-being of research participants, or other ethical considerations prevail. In the case of Indigenous data stewardship, curation will need to adhere to the CARE Principles for Indigenous Data Governance (Research Data Alliance 2019), designed explicitly to support Indigenous research methods and data sovereignty. The CARE framework (collective benefit, authority to control, responsibility, and ethics) prioritizes local Indigenous values and decision making by Indigenous communities in the management of research processes and resulting data products (Carroll, Rodriguez-Lonebear, and Martinez 2019).

The international Indigenous data sovereignty community seeks to align the FAIR and CARE principles (e.g., "be fair and care"), acknowledging the importance of FAIR for reducing threats to Indigenous data being "left aside due to lack of identifiers such as provenance or attribution metadata" (Carroll et al. 2021). In practical application, however, synchronization of the two sets of principles will be challenging and require collaboration between Indigenous nations and communities and data infrastructure developers. The fundamental goals behind CARE range from direct data-related outcomes of improved data governance and increased data literacy to much broader outcomes of strengthening and revitalizing Indigenous languages and cultures and addressing historical barriers for underrepresented communities and knowledge systems (Carroll, Rodriguez-Lonebear, and Martinez 2019; Carroll et al. 2020). These ambitions can be achieved only through long-term partnerships and commitments among all stakeholders.

For convergence curation, emerging disciplinary variations and ethics-based alternatives to FAIR are likely to have significant implications for how data are represented and managed. Responsible stewardship will require retaining disciplinary data contexts and enforcing stakeholder values within those contexts. Responsible research data services, an emerging focus in academic libraries (ACRL 2020), will need to balance ethical approaches to rights and privacy with the "open by design" trends that are vital to scaling actionable digital collections (Padilla 2019; Whitehead et al. 2021).

Disciplinary Research Integrity

Much of the work conducted by data curators is in the service of research integrity. Data curation activities support access to data for verification and transparency of research processes, and they retain quality and add value to data to support valid reuse. In science discourse, concerns with research integrity have escalated in recent years, with considerable focus on reproducibility, a process understood to be integral to the integrity of the scientific method. Sometimes characterized as a reproducibility or replication crisis, the complicated issues surrounding reproducibility in the digital environment are now debated within many scientific disciplines (e.g., *Nature* 2013; Baker 2016; *Linguistics* 2021; Peng 2009; Russell 2013; Stodden 2010). The topic has also become fairly common in science journalism, where it has the potential to sow mistrust in scientific findings by policy makers and the public.

As Leonelli (2018) has observed, the high degree of awareness of reproducibility and its relationship to research rigor has resulted in inflated application of the concept, which now often serves as proxy for quality and reliability of research results more generally. Yet, as Leonelli argues, the term fails to reflect the many different methods and goals involved in successful science, especially in areas of research where results can be, and may need to be, obtained without reproduction as a goal. Understanding the methods and values that define research integrity and data quality within research communities is key to guarding against outsized assumptions about reproducibility.

In fields where cross-disciplinary integration and analysis are essential, such as ecology, climate science, and earth systems science, there is considerable variability in the relevance or applicability of reproducibility as an indicator of research quality or integrity. Specific reproducibility problems do exist, and techniques for improving reproducibility are shared and promoted (see, e.g., Fraser et al. 2018; Konkol, Kray, and Pfeiffer 2019; Stagge et al. 2019; Milcu et al. 2018; Powers and Hampton 2019; Santer, Wigley, and Taylor 2011). However, for some research areas and methods, investing in reproducibility would be a misplaced priority. Climate science, a field where convergence research hinges on data compatibility and integration, is a compelling example. As demonstrated in Mayernik's (2021) study of climate model intercomparison projects (MIPs), the work to collect, distribute, and compare data is a complex process that progresses not through reproducibility but through the efforts of organizations to standardize data and metadata and build a research community to develop common vocabularies and uniform methods for evaluating data.

Earth systems science (ESS) is another informative case where the need for convergence capabilities is high and some of the primary methods do not align with reproducibility goals. ESS draws on heterogeneous data from geology, climatology, oceanography, ecology, geoinformatics, and computer science to understand and predict phenomena at different scales through deep integration of biophysical processes and human dynamics (Steffen et al. 2020). Our study of data sharing, reuse, and reproducibility in ESS (Yan et al. 2020), representing researchers across ten ESS disciplines, indicated that reproducibility is not a concern for researchers who generate and use data gathered in the field or who conduct large-scale computational modeling. Nearly all respondents reported reusing data generated by others but primarily for new and comparative analyses—only 18.5% reused data for reproducing existing studies, and 54% had never tried to reproduce research. Reproducibility was considered inapplicable due to the uniqueness of samples and variability inherent in natural systems and site conditions across time periods. Documentation of data and methods was one of the most pronounced practical problems inhibiting data reuse, consistent with previous investigations of earth and environmental science (Murillo 2022), geophysics (Tenopir et al. 2018), and geobiology (Palmer et al. 2017). Consistent with the research integrity orientation in MIP discussed above, ESS researchers suggested that computational work should aim for "consensus results" and ways to verify research credibility, rather than reproducibility.

Curation for ESS and other fields where convergence research is growing will need to be prudent in its investment in reproducibility, guided by understanding of research methods, the ways that data are reused in practice, and what constitutes data quality for specific research purposes. For many research communities, effort needs to be directed toward retention of the methods and context of data production to preserve the meaning of data over time as well as the potential for aggregation, integration, and comparison in future applications. For field sciences, for example, capturing the irreproducible aspects of the research process is particularly important, such as the spatial and temporal context of immediate conditions where and when samples are taken (Palmer et al. 2017; Powers and Hampton 2019). Moreover, for domains reliant on ever-growing layers of complex data and high-performance computing, not only is reproduction impractical (Bai et al. 2017), but the financial resources required for reproducibility are "disproportionately large in comparison to the added value" (Nüst et al. 2018).

Ecosystems and Metastructures

The divergent applications of FAIR and reproducibility are part of the functioning of the distributed system of diverse and interconnected data repositories and services, described by Parsons and Fox (2013) and many others as an "ecosystem." Diversity and interrelations are central to an ecosystem by definition, and based on the discussion above, we can expect that the variability and connections will continue to introduce complexity and dependencies into the ecosystem. More disciplines, sub-disciplines, and research communities will be evolving practices according to their orientation to the FAIR principles and prioritizing their own standards of research integrity. The varying organizational models of data management they identified illustrate how data infrastructure is evolving in response to disciplinary differences: Within the ecosystem, "big iron" approaches for instrument driven big science coexist with artisanal "science support" for the more customized curation typical in lab or field site settings (Parsons and Fox 2013, WDS37–38). For the humanities, Anderson and Blanke (2012) make the case for an independent ecosystem, separate from the sciences, to represent what they see as a distinct and flexible combination of humans, machines, and content to achieve research goals.

While the utility of the ecosystem as a metaphor for data infrastructure has its limits, the foregrounding of the fundamental nature of the interconnections is apt and underscores the centrality of metastructures. In keeping with Smith's Memex perspective, metastructures are the essence of a functioning system that supports convergence, elevating associative trails and providing scaffolding that reduces scatter. The scatter problem is a long-standing area of research within LIS aimed at improving access to topics with high distribution across literatures and databases (Bates 1996). It also serves as a unifying concept for the field's body of research on interdisciplinary information (Palmer and Fenlon 2017) that includes Smith's study of interdisciplinary search (1974), Don Swanson's (1986) pioneering work on literature-based discovery, and many others addressing cross-disciplinary access for interdisciplinary synthesis (e.g., Vakkari and Talja 2005; Weisgerber 1993; White 1996). There is now a need for renewed attention to operationalizing and synergizing data-related metastructures for convergence purposes.

Registries of data repositories currently offer a significant structural layer across the scatter of global data platforms. Re3data (https://www.re3data.org), for example, provides access to thousands of disciplinary and institutional repositories, large and small, across the world. Another important resource, FAIRsharing (https://fairsharing.org), brings together metadata and semantic standards, schema, and guidelines, interrelated to databases and data policies. The fast growth of this collection of standards makes evident the abundance and complexity of the different disciplinary representation structures within the ecosystem. An array of minimum information frameworks (MIFs), for example, demonstrates the commitment within research communities to developing standards for their specific research methods, but with a broader, overarching meta-aim of facilitating integration of data across disciplines (Taylor et al. 2008).

Generalist multidisciplinary repositories, such as Zenodo (https://zenodo.org) and Figshare (https://figshare.com), also play an important role within the data ecosystem, as readily accessible platforms that host data as well as other research outputs. They have been a valuable outlet for authors of journal publications that require access to associated data. Often considered supplemental to the paper, the data deposited in generalist repositories retain a tight coupling with the corresponding publication. The trend, however, also contributes to the scatter of individual datasets, compared to placement of data in platforms designed for disciplines or research communities, and is exacerbated by a general lack of machine readability of supplementary journal material. The newer, distinct genre of data journals, however, is making a unique contribution to the ecosystem through dissemination of papers expressly about datasets. Data papers have singular value as perhaps the richest source of contextual metadata for a dataset. They can fully document provenance and processes involved in the complex subsetting, repurposing, and integration of data that will become more common in convergence research.

Repository registries are valuable as an entry point into data resources, but they also serve as a vantage point. They provide a unique landscape view of the ecosystem from which we can discern the range of data resources and the density of particular subject areas or kinds of data. They may also be able to seed certain kinds of data-driven inquiry. However, they are not yet optimized for cross-disciplinary inquiry or processes such as identifying data about a key event, location, or phenomenon that may serve as anchors for interdisciplinary investigation. Even as the data registries, generalist platforms, and modes of data publication mature, scatter remains pronounced. To continue the metaphor, a thriving ecosystem will require serious advances in representation of interconnections within and across disciplines. These metastructures are how the individual data outputs in their many forms, guided by many standards and managed in a range of repositories and platforms, will evolve into a more complete, functioning ecosystem.

Conclusion

With significant investment now flowing to science and engineering to spur convergence research for transdisciplinary innovation, it is more important than ever for LIS to continue to forge Linda Smith's path toward realizing the potential of information systems for discovery. Development of expertise in curation for convergence will be essential to LIS roles supporting the many disciplinary and interdisciplinary communities we serve in our libraries and repositories and in ongoing development of data infrastructure. Facilitating convergence research will require growth and maturation of curatorial approaches that illuminate and exploit the potential of interrelationships across disciplines, through new semantic tools and other metastructures, while also investing in the stability and progress of foundational disciplines.

This intersection of disciplinary and cross-disciplinary expertise will be vital to the profession's contributions to convergence systems and services. Metastructural solutions will be successful only if they are leveraging the methods and materials of working research communities, while also accounting for the kinds of critical differences in priorities and practices illustrated in our discussion of FAIR and reproducibility. At the same time, there is still much that needs to be determined about the ratios of disciplinary and convergence investment needed for different research problems and purposes. The fact that there is no one-size-fits-all solution is at the heart of this dual perspective of convergence curation.

Ultimately, the field's ambitions need to reach further, beyond access and use of data and information for research convergence. LIS is also uniquely positioned to help build the bridges necessary for public access to research outputs, especially in areas of urgency for public awareness and planning, such as climate change. There has been a trend in intent to open up data resources for public access and use. For example, data centers may state that they hold "data for everyone," such as at the National Snow and Ice Data Center, or acknowledge that their data are of interest to the "general public," such as at the IPCC Data Distribution Centre. However, at present, high levels of intermediation are required for the public to benefit from these vast and complex resources. An interpretive interface of services and products is needed to put evidence, and its significance, into context for understanding and use by the public and other intermediating professionals, such as journalists and teachers. Basic tenets of data curation that call for transparency of the "who," "how," and "why" of data production are clearly fundamental. The emerging panoply of city, state, and federal open data resources are also within scope for this broader conception of the data ecosystem. There are many dimensions of public understanding of science and data literacy that need to be factored into long-term work of building robust data infrastructure.

Effective convergence curation that promotes true cross-disciplinary integration will come from expertise that understands and responds to the variations that make a difference across working research communities. While some empirical work has examined how to assess differences in curation requirements across disciplines (Chao, Cragin, and Palmer 2015), we still need a sound theory-driven framework for decision making on how to treat and maintain different types and domains of data, where, and for how long. However, among the decades of studies on how researchers collect, distribute, compare, and reuse data, Mayernik's (2021) conclusions about the climate model intercomparison research hold strong. Advances come through organizations that work to standardize data and metadata, while also building communities that generate metastructures and evolve principled approaches to determining the integrity and value of research products, for their methods and applications.

We continue to be informed and inspired by Dr. Smith's legacy as a scholar and educator. As vividly illustrated in the White and McCain's 1998 cocitation analysis of the discipline of information science, Linda was one of a small set of "integrative forces at work" (347), serving as a synthetic voice across the two primary subdisciplines—information retrieval and domain analysis—at the time. We offer this reflection on how her work underpins our scholarship as a tribute to the cohesion she brought to the field and the abundance of wisdom and warmth she contributed to our own academic and professional pursuits.

Acknowledgments

Linda's generosity and insight as a doctoral advisor seem to have no bounds, especially in terms of opening pathways to opportunity. For just one example, Linda was instrumental in connecting Melissa to the international community working on scientific data stewardship, supporting her participation in the ERPANET/CODATA meeting in 2003 on Selection, Appraisal and Retention of Digital Scientific Data. Connections from that meeting were the starting point for international collaborations and service to the field in data curation, stewardship, and policy.

We also wish to acknowledge Peter Fox, a generous friend and colleague, who served as a research fellow from 2012 to 2014 at the School of Information Sciences at the University of Illinois. Peter's mastery and collaborative spirit were invaluable to our work at the Center for Informatics Research in Science and Scholarship (CIRSS) and left an indelible imprint on the many faculty and students he engaged with at CIRSS.

We are grateful to the editorial team whose suggestions and corrections have improved this work; any remaining errors reside with the authors.

This work was supported in part by NSF grant 1928208 and draws on previous work from NSF grant 0830976 and IMLS grant LG-06-12-0706-12.

References

ACRL Research Planning and Review Committee. 2020. “2020 Top Trends in Academic Libraries: A Review of the Trends and Issues Affecting Academic Libraries in Higher Education.” *College & Research Libraries News* 81 (6): 270.

ALLEA. 2020. “Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities.” Edited by Natalie Harrower, Maciej Maryl, Beat Immenhauser, and Timea Biro. Berlin: ALLEA—All European Academies. https://doi.org/10.7486/DRI.tq582c863.

Anderson, Sheila, and Tobias Blanke. 2012. “Taking the Long View: From e-Science Humanities to Humanities Digital Ecosystems.” *Historical Social Research* 37 (3): 147–64.

Bai, Yuqi, Clifford A. Jacobs, Mei-Po Kwan, and Christoph Waldmann. 2017. “Geoscience and the Technological Revolution [Perspectives].” *IEEE Geoscience and Remote Sensing Magazine* 5 (3): 72–75.

Baker, Monya. 2016. “1,500 Scientists Lift the Lid on Reproducibility.” *Nature* 533 (26): 452–54.

Baru, Chaitanya, Lara Campbell, Aurali Dade, Pradeep Fulay, Alex Loewi, Douglas Maughan, Ibrahim Mohedas, Linda Molnar, Michael Pozmantier, Michael Reksulak, et al. 2022. “The NSF Convergence Accelerator Program.” *AI Magazine* 43 (1): 6–16. https://doi.org/10.1609/aimag.v43i1.19118.

Bates, Marcia J. 1996. “Learning about the Information Seeking of Interdisciplinary Scholars and Students.” *Library Trends* 45 (2): 155–64.

———. 1999. “The Invisible Substrate of Information Science.” *Journal of the American Society for Information Science* 50 (12): 1043–50. https://doi.org/10.1002/(SICI)1097-4571(1999)50:12<1043::AID-ASI1>3.0.CO;2-X.

Bush, Vannevar. 1945. “As We May Think.” *Atlantic Monthly* 176 (1): 101–8. https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881.

Carroll, Stephanie Russo, Ibrahim Garba, Oscar L. Figueroa-Rodríguez, Jarita Holbrook, Raymond Lovett, Simeon Materechera, Mark Parsons, Kay Raseroka, Desi Rodriguez-Lonebear, Robyn Rowe, et al. 2020. “The CARE Principles for Indigenous Data Governance.” *Data Science Journal* 19 (1): 43. http://doi.org/10.5334/dsj-2020-043.

Carroll, Stephanie Russo, Edit Herczog, Maui Hudson, Keith Russell, and Shelley Stall. 2021. “Operationalizing the CARE and FAIR Principles for Indigenous Data Futures.” *Scientific Data* 8: 108. https://10.1038/s41597-021-00892-0.

Carroll, Stephanie Russo, Desi Rodriguez-Lonebear, and Andrew Martinez. 2019. “Indigenous Data Governance: Strategies from United States Native Nations.” *Data Science Journal* 18: 1–15. https://doi.org/10.5334/dsj-2019-031.

Chao, Tiffany C., Melissa H. Cragin, and Carole L. Palmer. 2015. “Data Practices and Curation Vocabulary (DPCVocab): An Empirically Derived Framework of Scientific Data Practices and Curatorial Processes” *Journal of the Association for Information Science and Technology* 66 (3): 616–33. https://doi.org/10.1002/asi.23184.

Chaudhri, Vinay, Chaitanya Baru, Naren Chittar, Xin Dong, Michael Genesereth, James Hendler, Aditya Kalyanpur, Douglas Lenat, Juan Sequeda, Denny Vrandečić, and Kuansan Wang. 2022. “Knowledge Graphs: Introduction, History, and Perspectives.” *AI Magazine* 43 (1): 17–29. https://doi.org/10.1609/aimag.v43i1.19119.

Cooper, Danielle, and Rebecca Springer. 2019. “Data Communities: A New Model for Supporting STEM Data Sharing.” *Ithaka S+R*, May 13. https://doi.org/10.18665/sr.311396.

Cremer, Fabian, Silvia Daniel, Marina Lemaire, Katrin Moeller, Matthias Razum, and Arnošt Štanzel. 2021. “Data Meets History: A Research Data Management Strategy for the Historically Oriented Humanities.” In *Band 21 Cultural Sovereignty beyond the Modern State: Space, Objects, and Media*, edited by Gregor Feindt, Bernhard Gissibl, and Johannes Paulmann, 155–78. Berlin: De Gruyter Oldenbourg. https://doi.org/10.1515/9783110679151-009.

Faniel, Ixchel M., Rebecca D. Frank, and Elizabeth Yakel. 2019. “Context from the Data Reuser’s Point of View.” *Journal of Documentation* 75 (6): 1274–97.

Faniel, Ixchel M., Eric Kansa, Sarah Whitcher Kansa, Julianna Barrera-Gomez, and Elizabeth Yakel. 2013. “The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse.” In *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, 295–304. New York: ACM. https://doi.org/10.1145/2467696.2467712.

Feldman, Shelley, and Linda Shaw. 2019. "The Epistemological and Ethical Challenges of Archiving and Sharing Qualitative Data." *American Behavioral Scientist* 63 (6): 699–721. https://doi.org/10.1177/0002764218796084.

Fox, Peter, and James A. Hendler. 2009. "Semantic eScience: Encoding Meaning in Next-Generation Digitally Enhanced Science." In *The Fourth Paradigm: Data-Intensive Scientific Discovery*, edited by Tony Hey, Stewart Tansley, and Kristin Tolle, 147–52. Seattle: Microsoft Research.

Fraser, Hannah, Tim Parker, Shinichi Nakagawa, Ashley Barnett, and Fiona Fidler. 2018. "Questionable Research Practices in Ecology and Evolution." *PLoS ONE* 13 (7): e0200303. https://doi.org/10.1371/journal.pone.0200303.

Gil, Yolanda, Suzanne A. Pierce, Hassan Babaie, Arindam Banerjee, Kirk Borne, Gary Bust, Michelle Cheatham, Imme Ebert-Uphoff, Carla Gomes, Mary Hill, et al. 2019. "Intelligent Systems for Geosciences: An Essential Research Agenda." *Communications of the ACM* 62 (1): 76–84. https://doi.org/10.1145/3192335.

Goble, Carole. 2020. "FAIR History and the Future." Talk at the FAIRplus Innovation and SME Forum: Implementing FAIR Data Principles in Industrial Life Science Research, January 29, 2020. https://www.youtube.com/watch?v=59f1s0-VfSE.

Janowicz, Krzysztof. 2021. "KnowWhereGraph Drives Analytics and Cross-Domain Knowledge." *ArcUser Spring* 2021:16–19. https://www.esri.com/about/newsroom/arcuser/knowwheregraph-drives-cross-domain-knowledge.

Janowicz, Krzysztof, Pascal Hitzler, Wenwen Li, Dean Rehberger, Mark Schildhauer, Rui Zhu, Cogan Shimizu, Colby K. Fisher, Ling Cai, Gengchen Mai, et al. 2022. "Know, Know Where, KnowWhereGraph: A Densely Connected, Cross-Domain Knowledge Graph and Geo-Enrichment Service Stack for Applications in Environmental Intelligence." *AI Magazine* 43:30–39. https://doi.org/10.1002/aaai.12043.

Johnson, J. Michael, Tom Narock, Justin Singh-Mohudpur, Doug Fils, Keith Clarke, Siddharth Saksena, Adam Shepherd, Sankar Arumugam, and Lilit Yeghiazarian. 2022. "Knowledge Graphs to Support Real-Time Flood Impact Evaluation." *AI Magazine* 43 (1): 40–45. https://ojs.aaai.org/index.php/aimagazine/article/view/19121.

Karcher, Sebastian, Dessislava (Dessi) Kirilova, Christiane Pagé, and Nic Weber. 2021. "How Data Curation Enables Epistemically Responsible Reuse of Qualitative Data." *Qualitative Report* 26 (6): 1996–2010. https://doi.org/10.46743/2160-3715/2021.5012.

Kochen, Manfred. 1964. "Adaptive Mechanism in Digital Concept Processing." *IEEE Transactions on Applications and Industry* 83 (74): 305–14. https://10.1109/TAI.1964.5407761.

Konkol, Markus, Christian Kray, and Max Pfeiffer. 2019. "Computational Reproducibility in Geoscientific Papers: Insights from a Series of Studies with Geoscientists and a Reproduction Study." *International Journal of Geographical Information Science* 33 (2): 408–29.

Leonelli, Sabina. 2018. "Rethinking Reproducibility as a Criterion for Research Quality." In *Including a Symposium on Mary Morgan: Curiosity, Imagination, and Surprise (Research on the History of Economic Thought and Methodology, Vol. 36B)*, 129–46. Bingley, UK: Emerald. https://doi.org/10.1108/s0743-41542018000036b009.

Linguistics. 2021. "Special Issue: The Replication Crisis: Implications for Linguistics." *Linguistics* 59 (5). https://www.degruyter.com/journal/key/ling/59/5/html.

Ma, Xiaogang. 2022. "Knowledge Graph Construction and Application in Geosciences: A Review." *Computers & Geosciences* 161: 1–15. https://doi.org/10.1016/j.cageo.2022.105082.

Mayernik, Matthew S. 2021. "Credibility via Coupling: Institutions and Infrastructures in Climate Model Intercomparisons." *Engaging Science, Technology, & Society* 7 (2): 10–32. https://doi.org/10.17351/ests2021.769.

McNutt, Marcia K. 2022. "Civilization-Saving Science for the Twenty-First Century." *Annual Review of Earth and Planetary Sciences* 50 (1): 1–12. https://doi.org/10.1146/annurev-earth-033021-081125.

Milcu, Alexandru, Ruben Puga-Freitas, Aaron M. Ellison, Manuel Blouin, Stefan Scheu, Grégoire T. Freschet, Laura Rose, Sebastien Barot, Simone Cesarz, Nico Eisenhauer, et al. 2018. "Genotypic Variability Enhances the Reproducibility of an Ecological Study." *Nature Ecology & Evolution* 2 (2): 279–87.

Murillo, Angela P. 2022. "Data Matters: How Earth and Environmental Scientists Determine Data Relevance and Reusability." *Collection and Curation* 41 (3): 77–86. https://doi.org/10.1108/CC-11-2018-0023.

National Research Council. 2005a. *Catalyzing Inquiry at the Interface of Computing and Biology*. Washington, DC: National Academies Press.

———. 2005b. *Mathematics and 21st Century Biology*. Washington, DC: National Academies Press. https://doi.org/10.17226/11315.

———. 2008. *Inspired by Biology: From Molecules to Materials to Machines*. Washington, DC: National Academies Press.

———. 2009. *A New Biology for the 21st Century*. Washington, DC: National Academies Press.

———. 2010. *Research at the Intersection of the Physical and Life Sciences*. Washington, DC: National Academies Press.

———. 2014. *Convergence: Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond*. Washington, DC: National Academies Press. https://doi.org/10.17226/18722.

National Science Foundation. n.d.-a. "Convergence Reports and References." https://www.nsf.gov/od/oia/convergence/reports.jsp.

———. n.d.-b. "Convergence Research at NSF." https://www.nsf.gov/od/oia/convergence/index.jsp.

———. n.d.-c. "NSF's 10 Big Ideas." https://www.nsf.gov/news/special_reports/big_ideas/index.jsp.

Nature. 2013. "Announcement: Reducing Our Irreproducibility." *Nature* 496 (7446): 398. https://doi.org/10.1038/496398a.

Nüst, Daniel, Carlos Granell, Barbara Hofer, Markus Konkol, Frank O. Ostermann, Rusne Sileryte, and Valentina Cerutti. 2018. "Reproducible Research and GIScience: An Evaluation Using AGILE Conference Papers." *PeerJ* 6:e5072. https://doi.org/10.7717/peerj.5072.

OCLC. n.d. "Find WorldCat Entities to Improve Resource Discoverability through Connections." https://www.oclc.org/en/worldcat/oclc-and-linked-data/worldcat-entities.html.

Padilla, Thomas. 2019. *Responsible Operations: Data Science, Machine Learning, and AI in Libraries*. Dublin, OH: OCLC Research. https://doi.org/10.25333/xk7z-9g97.

Palmer, Carole L., Melissa H. Cragin, and Timothy P. Hogan. 2007. "Weak Information Work in Scientific Discovery." *Information Processing & Management* 43 (3): 808–20.

Palmer, Carole L., and Katrina Fenlon. 2017. "Information Research on Interdisciplinarity." In *The Oxford Handbook of Interdisciplinarity*, 2nd ed., edited by Robert Frodeman, Julie Thompson Klein, and Roberto Carlos Dos Santos Pacheco, 429–42. Oxford: Oxford University Press. https://10.1093/oxfordhb/9780198733522.013.35.

Palmer, Carole L., Andrea K. Thomer, Karen S. Baker, Karen M. Wickett, Christie L. Hendrix, Ann Rodman, Stacey Sigler, and Bruce W. Fouke. 2017. "Site-Based Data Curation Based on Hot Spring Geobiology." *PLoS ONE* 12 (3): e0172090. https://doi.org/10.1371/journal.pone.0172090.

Palmer, Carole L., Oksana L. Zavalina, and Katrina Fenlon. 2010. "Beyond Size and Search: Building Contextual Mass in Aggregations for Scholarly Use." In *ASIS&T '10: Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem* 47, 1–10. https://doi.org/10.1002/meet.14504701213.

Parsons, Mark A., and Peter A. Fox. 2013. "Is Data Publication the Right Metaphor?" *Data Science Journal* 12: WDS32–46. http://doi.org/10.2481/dsj.WDS-042.

Peng, Roger D. 2009. "Reproducible Research and *Biostatistics*." *Biostatistics* 10 (3): 405–8.

Perry, James W., Allen Kent, and Madeline M. Berry. 1956. "Searching Strategy and Association Trails." In *Machine Literature Searching*, 109–22. New York: Interscience.

Powers, Stephen M., and Stephanie E. Hampton. 2019. "Open Science, Reproducibility, and Transparency in Ecology." *Ecological Applications* 29 (1): e01822.

Research Data Alliance, International Indigenous Data Sovereignty Interest Group. 2019. "CARE Principles for Indigenous Data Governance." Global Indigenous Data Alliance. https://www.gida-global.org/care.

Russell, Jonathan F. 2013. "If a Job Is Worth Doing, It Is Worth Doing Twice." *Nature* 496 (7443): 7.

Santer, Brian D., Tom M. L. Wigley, and Karl E. Taylor. 2011. "The Reproducibility of Observational Estimates of Surface and Atmospheric Temperature Change." *Science* 334 (6060): 1232–33. https://doi.org/10.1126/science.1216273.

Sharp, Phillip A. 2014. "Meeting Global Challenges: Discovery and Innovation through Convergence." *Science* 346 (6216): 1468–71. https://doi.org/10.1126/science.aaa3192.

Sharp, Phillip A., and Robert Langer. 2011. "Promoting Convergence in Biomedical Science." *Science* 333 (6042): 527. https://doi.org/10.1126/science.1205008.

Sinha, A. Krishna, Zaki Malik, Abdelmounaam Rezgui, David L. Fox, Calvin G. Barnes, Kai Lin, Grant Heiken, William A. Thomas, Linda C. Gundersen, Robert Raskin, et al. 2010. "Geoinformatics: Transforming Data to Knowledge for Geosciences." *GSA Today* 20 (12): 4–10. doir.org/10.1130/GSATG85A.1.

Smith, Linda C. 1974. "Student Paper Award 1974. Systematic Searching of Abstracts and Indexes in Interdisciplinary Areas." *Journal of the American Society for Information Science* 25 (6): 343–53.

———. 1980. "'Memex' as an Image of Potentiality in Information Retrieval Research and Development." In *SIGIR '80: Proceedings of the 3rd Annual ACM Conference on Research and Development in Information Retrieval*, 345–69. New York: ACM. https://dl.acm.org/doi/10.5555/636669.636692.

———. 1981. "Representation Issues in Information Retrieval System Design." *ACM SIGIR Forum* 16 (1): 100–105. https://dl.acm.org/doi/abs/10.1145/1013228.511770.

Stagge, James H., David E. Rosenberg, Adel M. Abdallah, Hadia Akbar, Nour A. Attallah, and Ryan James. 2019. "Assessing Data Availability and Research Reproducibility in Hydrology and Water Resources." *Scientific Data* 6:190030.

Steffen, Will, Katherine Richardson, Johan Rockström, Hans Joachim Schellnhuber, Opha Pauline Dube, Sébastien Dutreuil, Timothy M. Lenton, and Jane Lubchenco. 2020. "The Emergence and Evolution of Earth System Science." *Nature Reviews Earth & Environment* 1 (1): 54–63.

Stodden, Victoria C. 2010. "Reproducible Research: Addressing the Need for Data and Code Sharing in Computational Science." *Computing in Science & Engineering* 12 (5): 8–13.

Swanson, Don R. 1986. "Undiscovered Public Knowledge." *Library Quarterly* 56 (2): 103–18.

Tague, Jean. 1969. "Association Trails." In *Encyclopedia of Library and Information Science*, vol. 2, edited by Allen Kent and Harold Lancour, 55–80. New York: Marcel Dekker.

Taylor, Chris F., Dawn Field, Susanna-Assunta Sansone, Jan Aerts, Rolf Apweiler, Michael Ashburner, Catherine A. Ball, Pierre-Alain Binz, Molly Bogue, Tim Booth, et al. 2008. "Promoting Coherent Minimum Reporting Guidelines for Biological and Biomedical Investigations: The MIBBI Project." *Nature Biotechnology* 26 (8): 889–896. https://doi.org/10.1038/nbt.1411.

Tenopir, Carol, Lisa Christian, Suzie Allard, and Joshua Borycz. 2018. "Research Data Sharing: Practices and Attitudes of Geophysicists." *Earth and Space Science* 5 (12): 891–902.

Tóth-Czifra, Erzsébet. 2020. "The Risk of Losing Thick Description: Data Management Challenges Faced by the Arts and Humanities in the Evolving FAIR Data Ecosystem." In *Digital Technology and the Practices of Humanities Research*, edited by Jennifer Edmond, 235–266. Cambridge: Open Book. https://doi.org/10.11647/OBP.0192.

Vakkari, Pertti, and Sanna Talja. 2005. "The Influence of the Scatter of Literature on the Use of Electronic Resources across Disciplines: A Case Study of FinELib." ECDL 2005. In *Research and Advanced Technology for Digital Libraries: ECDL 2005,* edited by Andreas Rauber, Stavros Christodoulakis, and A. Min Tjoa. *Lecture Notes in Computer Science* 3652: 207–17. https://doi.org/10.1007/11551362_19.

Wagener, Thorsten, Simon J. Dadson, David M. Hannah, Gemma Coxon, Keith Beven, John P. Bloomfield, Wouter Buytaert, Hannah Cloke, Paul Bates, Joseph Holden, et al. 2021. "Knowledge Gaps in Our Perceptual Model of Great Britain's Hydrology." *Hydrological Processes* 35 (7): e14288. https://doi.org/10.1002/hyp.14288.

Weisgerber, David W. 1993. "Interdisciplinary Searching: Problems and Suggested Remedies: A Report from the ICSTI Group on Interdisciplinary Searching." *Journal of Documentation* 49 (3): 231–54. https://doi.org/10.1108/eb026914.

White, Howard D. 1996. "Literature Retrieval for Interdisciplinary Syntheses." *Library Trends* 45 (2): 239–64.

White, Howard D., and Katherine W. McCain. 1998. "Visualizing a Discipline: An Author Co-citation Analysis of Information Science, 1972–1995." *Journal of the American Society for Information Science* 49 (4): 327–55. https://doi.org/10.1002/(SICI)1097-4571(19980401)49:4<327::AID-ASI4>3.0.CO;2-4.

Whitehead, Martha, Dale Askey, Donna Bourne-Tyson, Karen Estlund, Susan Haigh, Claire Stewart, Kornelia Tancheva, Tyler Walters, Jennifer Muilenburg, and Judy Ruttenberg. 2021.

"ARL/CARL Joint Task Force on Research Data Services: Final Report." Washington, DC, and Ottawa, ON: Association of Research Libraries and Canadian Association of Research Libraries. https://doi.org/10.29242/report.arlcarlrdstaskforce2021.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, et al. 2016. "The Fair Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1). https://doi.org/10.1038/sdata.2016.18.

Yan, An, Caihong Huang, Jian-Sin Lee, and Carole L. Palmer. 2020. "Cross-Disciplinary Data Practices in Earth System Science: Aligning Services with Reuse and Reproducibility Priorities." *Proceedings of the American Society for Information Science & Technology,* 57 (1). https://doi.org/10.1002/pra2.218.

Carole L. Palmer is a professor and associate dean for research at the Information School at the University of Washington. She works in the areas of data curation and digital research collections, specializing in cross-disciplinary access and use and emerging best practices in data services for researchers and the public. She is PI on the Data Services for Indigenous Scholarship and Sovereignty project (DSISS), funded by the Mellon Foundation, is co-PI on the NSF Qualitative Data Repository initiative, and led the recent IMLS Open Data Literacy project, aimed at advancing open data for the public good through education, engagement, and research. She holds a PhD in library and information science from the University of Illinois Urbana-Champaign. Before joining the Information School at the University of Washington, she was director of the Center for Informatics Research in Science & Scholarship at the School of Information Sciences at the University of Illinois.

Melissa H. Cragin is associate vice president for research computing at Rice University. Prior to joining Rice, she was chief strategist for data initiatives in the Research Data Services division at the San Diego Supercomputer Center, University of California, San Diego. Previously, she was executive director of the Midwest Big Data Hub, based at the National Center for Supercomputing Applications at the University of Illinois Urbana-Champaign (UIUC). She has served as staff associate and an AAAS science & technology policy fellow in the Office of the Assistant Director, Directorate of Biological Sciences at the National Science Foundation. She has a PhD in library and information science from UIUC, and a MLIS degree from Rutgers University.