IJDC | *Conference Paper*

Two Decades, Same Story? Insights and Future Directions in Long Tail Data Curation

Inna Kouper
Luddy School of Informatics, Computing, and Engineering
Indiana University

Gretchen R. Stahlman
School of Information, College of Communication & Information
Florida State University

Abstract

This paper examines the evolution of the concept of long tail research data in the scholarly literature. The "long tail" concept, originally used to describe "niche" digital products that have a significant market share when taken as an aggregate, was first applied to research data in 2007 to refer to a vast array of smaller, heterogeneous data collections that cumulatively represent a substantial portion of scientific knowledge. These datasets are frequently overlooked due to inadequate data management practices and institutional support. Bridging the discussions on data curation in library & information science (LIS) and domain-specific contexts, this paper identifies several themes in these discussions and offers insights, or provocations, that encourage researchers to rethink the existing frameworks and methods and find new approaches that would help both researchers and data professionals. This review seeks to enhance understanding of long tail data as both a concept and a field, while also informing current and future research and practice.

Submitted 30 January 2025 ~ *Accepted* 20 February 2025

Correspondence should be addressed to Inna Kouper, Email: inkouper@iu.edu

This paper was presented at the International Digital Curation Conference IDCC25, 17-19 February 2025

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

http://dx.doi.org/10.2218/ijdc.v19i1.1057
DOI: 10.2218/ijdc.v19i1.1057

Introduction

The concept of the "long tail" in business illustrates how online platforms and digital distribution have transformed sales and consumption patterns. It reveals that niche products—those with limited individual demand—can collectively comprise a significant share of the market when aggregated. This aggregation, which was once impractical due to physical limitations of traditional retail, can now generate economic value that is comparable or even surpasses that of mainstream blockbusters (Anderson, 2007). Platforms such as Amazon, Netflix, and Spotify exemplify this phenomenon as they leverage their vast digital catalogs to capture the cumulative demand for countless niche products and consumer preferences.

This concept has been applied in research data management and curation to describe trends where a few datasets are highly popular, but most "niche" datasets have low popularity and use (Palmer et al., 2007). Long tail data refers to smaller datasets generated by individual researchers or teams that, despite low individual usage, may hold significant value for future research and innovation when considered in aggregate. Such data has often remained inaccessible due to inadequate curation (Heidorn, 2008). Over the past two decades, scholars and curators have used the term "long tail data" to discuss policies, infrastructures, and training aimed at better managing these smaller individual data collections (PLAN-E, 2018).

This paper examines how the "long tail" concept and related frameworks in data management and curation have evolved over time. Building upon our previous work (Stahlman and Heidorn, 2020; Stahlman and Kouper, 2024), we searched Google Scholar for "long tail data" and, starting with a smaller sample of papers, used a snowballing technique to find more papers that provide a panoramic perspective on the history and evolution of long tail data. Additionally, we identified seven highly referenced papers that we consider to be seminal for this topic in LIS and collected papers that cite those seven papers, aiming at including at least one paper per each seminal paper per year published between 2007 and 2022. The LIS-oriented papers were added to our larger database of papers for the review, bringing the final set to over 300 papers. We closely examined the papers and identified several themes, which we discuss below. We then synthesize these themes into key insights and recommendations that can inform and guide future research and practice in this area.

Digitalization and Shifts in Librarianship

Even before the term "long tail data" became popular in the scholarly literature, the challenges of long tail data curation were discussed as part of the transition to digital information and collection-oriented thinking about data in the 1990s and 2000s. Digital libraries and curation initiatives have significantly impacted the management and preservation of long tail research data over the past decades, with discussions of more efficient processing of information assets going back to earlier visions of mechanized information storage and retrieval (Bush, 1945; Lesk, 2012). Libraries constantly adopt newer technologies to meet changing information needs and maintain relevance in the digital age (Borgman, 2000).

The transition from physical to digital collections marked a transformative period in the history of libraries, as they adapted to new forms of information storage, retrieval, and dissemination. This evolution has been driven by the need to manage an ever-expanding volume of digital content, including research data, multimedia, and specialized collections. In the early 2000s, a range of tools emerged to address the evolving needs of digital libraries, focusing on preservation, metadata management, and accessibility and including repository software systems, digital collection storage and curation tools, and frameworks for preservation and interoperability (Candela, Castelli, & Pagano, 2011).

Digital curation emerged in the early 2000s as a profession and a growing subdiscipline of LIS that focused on the long-term management of digital information assets (Beagrie, 2006; Higgins, 2011; Kouper, 2016). While related to digital preservation and archiving, digital curation is about

maintaining and adding value to digital research data throughout its lifecycle, and many developments in this field are conceptually connected to how scientific communities work with data (Oliver & Harvey, 2016). That is why this field consistently advocates for a lifecycle perspective on data management (Pennock, 2007; Higgins, 2008), with various data lifecycle models guiding curation practices across research contexts (Weber and Kranzlmüller, 2019; Huang et al., 2020; Stahlman, 2022). By adopting data lifecycle models and leveraging new tools and standards, libraries aimed at positioning themselves as essential partners in the curation and stewardship of research data.

By the early 2010s, digital data curation and the libraries' role in it became a stable topic of research and practice (Hank & Davidson, 2009; Tibbo & Lee, 2012). Initially, only a few research libraries had dedicated units and staff for digital data curation (Cox & Pinfield, 2014). Over time, the capacity of libraries to provide data curation services grew, with many libraries in the US, EU, and globally developing capabilities for long tail data curation and management (Kaushik, 2017; Yoon & Donaldson, 2019). The field has also seen increasing focus on the challenges of incentivizing and supporting researchers in sharing and reusing data (Borgman, 2012; Borgman et al., 2016). In light of limited resources and risks, "small science" and long tail data producers and curators are encouraged to collaborate and share data to increase the visibility of their work (Wallis et al., 2013; Wallis, 2014).

Cyberinfrastructure for Open Science

The development of new computational tools and paradigms affected data management and sharing, and the topics of cyberinfrastructure and open science became another nexus of discussions in the 2000s, emphasizing the importance of data-intensive discovery and advocating for new tools for data capture, curation, and analysis (Hey, Tansley & Tolle, 2009). In response to growing digital data and computing needs, libraries engaged in *e-science* efforts, leading to the creation of institutional repositories (IRs) to support digital scholarship and data archiving (Lynch, 2003). These IRs became a common approach to preserving and disseminating scholarly work, including datasets (Cragin et al., 2010; Reilly, 2014). Both generalist and specialist data repositories have evolved to support long tail data, though they sometimes struggle to meet researchers' needs due to the rapid changes in research data management, sharing, and preservation (Murillo, 2020; Rodrigues & Rodrigues, 2012).

Efforts to integrate cyberinfrastructure (CI) and data repositories emerged in order to improve long tail data curation (Choudhury & Kunze, 2009; Mokrane & Parsons, 2014). By leveraging CI resources, repositories could handle larger and more diverse datasets, facilitate metadata standards, and support data and software preservation. Projects like the NSF DataNet programs SEAD and DataONE have aimed to support environmental sciences and create networks of interoperable scientific data repositories (Plale et al., 2013; Michener et al., 2011). These initiatives emphasized the importance of enabling open science, viewing data as valuable assets, and fostering a culture of openness and transparency in research (Ramachandran et al., 2021).

Disciplinary Approaches to Long Tail Data

Long tail data discussions have been occurring across many disciplines, including earth and social sciences and biological and medical research (Hanson et al., 2020; Sinha et al., 2013; Ramdeen & Poole, 2017; Stephenson et al., 2020). The approaches to long tail data vary depending on the nature of the discipline, the structure of the data, and the resources available for curation. Geoscientists, for example, have initiated efforts to standardize data formats and metadata descriptions to improve discovery, interoperability, and usability of long tail data (Cutcher-

Gershenfeld et al., 2016). In fields like earth sciences, where large datasets are often generated by instruments and sensors, curators have worked closely with researchers to develop infrastructure and standards for managing diverse and high-volume datasets (Schindler et al., 2012).

Social sciences and medical research often require researcher-curator collaborations to address ethical concerns, privacy issues, and the integration of heterogeneous data from multiple sources. These partnerships have seen varying degrees of success, often depending on the availability of funding, disciplinary norms, and the degree to which curators and researchers share common goals for data management and data sharing (Fegraus et al., 2005). Successful examples of long tail data sharing include the conversion of resting state electroencephalogram (EEG) measurements in Cuba into a large dataset for widespread use (Bosch-Bayard et al., 2020). Repositories like GenBank for genetic sequences, the Protein Data Bank for 3D structures, and clinicaltrials.gov for clinical study information support long tail data sharing in these fields (Costa et al., 2016).

Other research areas represented in the literature about long tail data include materials science (Akmon et al., 2011), hydrology (Yu et al., 2020), scientific ocean drilling (Collier et al., 2015), statistical research (Bahls & Tochtermann, 2013), and astronomy (Heidorn, Stahlman & Steffen, 2018; Stahlman & Heidorn, 2020). By contrast, humanities and social sciences disciplines are not widely represented in published literature referring to long tail data. These disciplines are nevertheless conducive to long tail data dynamics such as heterogeneous data formats, the use of digitization and computational tools, and data sharing challenges (Wang, 2018). International contexts are also gaining relevance for discussions about long tail data, particularly as developing countries struggle to participate in research data management (Van Deventer, 2015; Patterton et al., 2018; Stahlman, 2023).

Long Tail Data and LIS

The LIS literature has seen a stable flow of publications on the topics of long tail data, which were spurred by several initial highly cited papers that introduced and promoted this concept (Cragin & Shankar, 2006; Borgman et al., 2007; Palmer et al., 2007; Heidorn, 2008; Cragin et al., 2010; Akmon et al., 2011; Heidorn, 2011). Between 2006 and 2022 we observed the relative stability and persistence of themes in the LIS literature as most of them remained in the discussions throughout these years. Our review suggests a thematic expansion of the themes of digital curation, CI and openness, and disciplinary long tail data over time rather than a sequential development or a paradigmatic shift (see Figure 1 below).

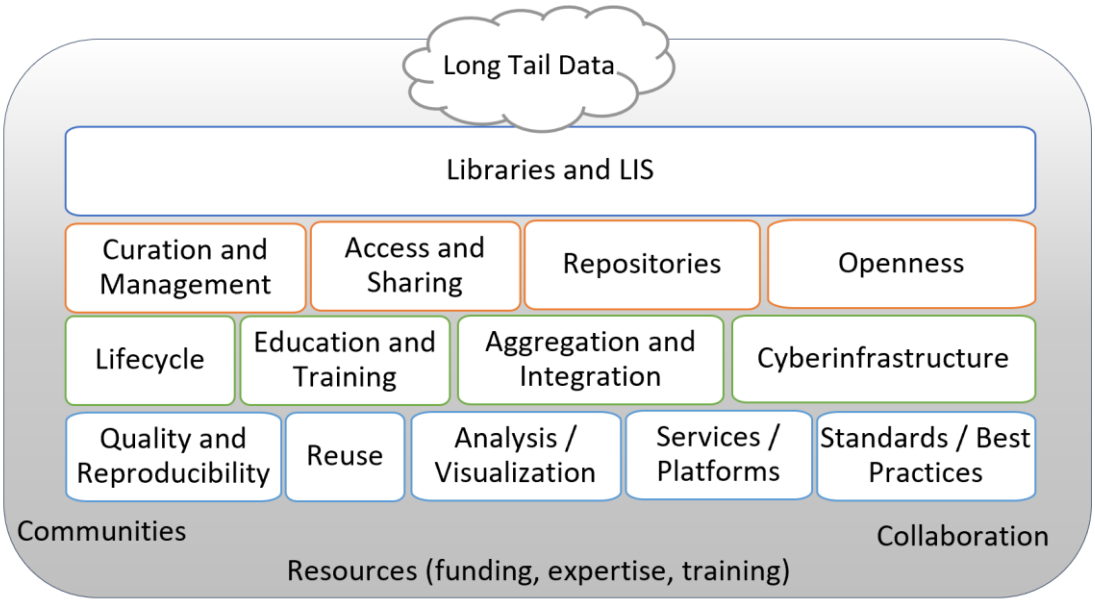

Figure 1. The evolution of the long tail data discussions in LIS.

Despite all the efforts, the difficulties in accessing, sharing, and reusing long tail data remain a persistent topic of discussions in the scholarly literature. Publications tend to combine "access and sharing" and "reuse" as they advocate for improved practices and a future research agenda within LIS. In 2020s the researchers and practitioners still reported reluctance to share and reuse data in some disciplines due to lack of trust, incentives, and user-friendly tools (Gries et al., 2023; Witting, 2023).

Earlier publications have studied the practices of individual researchers and specific communities, particularly in the US, in attempts to connect researcher needs to management and curation. Later publications continued discussing disciplinary curation and management, but they shifted their language to the broader challenges of making scientific research reproducible, recognizing interdisciplinary collaborations, and identifying gaps in existing cyberinfrastructure (Chawinga and Zinn, 2020; Lewis et al., 2018; Donaldson & Koepke, 2022). Management and curation discussions often co-occurred with the discussions of the changing technologies that support data and knowledge production (Atkins et al., 2003; Frischmann, 2012). The theme of management and curation gained a global perspective through publications about research data management in non-US and non-EU contexts.

The literature on long tail data is also marked by efforts to expand its conceptual repertoire. Many publications in our database anchored their research in discussions of varying scales and levels of data and the distinct needs of different domains and communities. As a result, long tail data has often been associated with "small" or "little" science, described as "dark" or "artisanal" data, and contrasted with big data, large-scale science, and web, digital, or crowdsourced data. These contrasting scales were frequently used to reiterate the value of long tail data and draw attention to its curation. At the same time, this discourse often blurred the lines between "scientific," "research," and "disciplinary" data across individuals, labs, broader communities, entire disciplines, and multidisciplinary collaborations.

The long tail data framework has also expanded into the topics of data quality and establishing standards and best practices as well as robust services and platforms. As more ideas have been introduced into the literature on long tail data, older themes have not disappeared but rather have been re-framed and incorporated into newer themes. Three cross-cutting themes—communities, collaboration, and resources—were frequently present as complementary, parallel, or sometimes orthogonal developments.

Key Insights from the Literature

To stimulate discussions on curating long tail data, we offer five *provocations* drawn from the literature.

P1. The concept of long tail data was a crucial rhetorical device that solidified ongoing work in archival and information science. Its greatest achievement has been advocacy for and establishment of the relevance of information professionals in the age of digital and data-intensive research.

P2. Despite efforts to advocate for openness and increased support in long tail data, this area has largely remained isolated, failing to connect researchers and data professionals. Researchers continue to face significant challenges in data management, including organizing and sharing data, selecting appropriate repositories, and curating their data for long-term preservation. Meanwhile, data professionals struggle with raising awareness about their services, embedding themselves into research teams, and establishing sustainable business models.

P3. Libraries may not be the most effective or efficient option for curating long tail research. Disciplinary and general-purpose repositories, such as Zenodo and Dryad, along with commercial solutions, such as Figshare, often compete with them. Many institutional and general repositories often fall short of meeting researchers' needs, as they lack the active data management support required for organizing data during the early stages of research.

P4. Successful examples of long tail data integration are the exception rather than the rule. There are many practical challenges of integrating long tail data, including lack of appropriate infrastructure, support, or incentives. Additionally, the complexity of coordinating data across disciplines further complicates data integration.

P5. Persistent challenges in long tail data curation are addressed through small-scale studies with no high-impact solutions. Long tail data curation faces a double bind: in some research communities, effective data management practices were already in place before curation efforts began, making additional interventions redundant. In contrast, other communities showed indifference or resistance to adopting new curation practices. The curation community's reliance on open science mandates proved insufficient in overcoming these challenges. As a result, comprehensive solutions for long tail data curation remain elusive.

Recommendations for the Future of Data Curation

Considering these insights, we propose a series of *recommendations* that could help to reshape the agenda and future research and practice in long tail data curation.

R1. Critique historical narratives and identify new questions. While the long tail data framework has been influential, its current relevance is uncertain. Do we still need it? Can we move beyond the concepts of "digital curation" and "data curation lifecycle"? While informative for LIS and for caring for digital data beyond its original use, these frameworks may have deepened the divide between the needs of science and research and the goals of archival and library communities. Unpacking these questions can help to pave the way for new frameworks and approaches.

R2. Explore new methods. Data curation research has relied upon surveys, interviews, and ethnography to understand and map data related needs, practices, challenges, and opportunities. Experimental studies and case studies that are tied directly to scientific outcomes can lead to more robust insights.

R3. Innovate in theory and practice. Long tail data curation relies on a mix of library-supported services and commercial tools that institutions purchase to support their research and education; it also faces resistance toward cross-unit and cross-institutional collaborations. We need new models that acknowledge this complexity and empower librarians and curators to navigate it. Data professionals may need to reimagine their roles and how they can help overcome limitations imposed by this mixture of community and commercial services. For example, instead of focusing on technology, they can shift toward advocating for supporting cross-institutional data management and upholding values of openness, privacy, and integrity in scholarship.

Conclusion

In summary, to address the challenges outlined above, we recommend critically evaluating the value of the long tail data framework for the future of data curation. The persistence of recurring themes over two decades suggests that key issues remain unresolved. Despite the proliferation of tools, models, platforms, and recommendations, it may be time to reconsider current approaches and return to the drawing board.

References

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. doi: 10.1007/s10502-011-9151-4

Anderson, C. (2007). *The long tail: How endless choice is creating unlimited demand*. Random House.

Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., Messina, P., Ostriker, J. P., & Wright, M. H. (2003). *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure* (No. 8; pp. 1562–1567). National Science Foundation. https://www.nsf.gov/cise/sci/reports/atkins.pdf

Bahls, D., & Tochtermann, K. (2013). Semantic retrieval interface for statistical research data. *Proceedings of the 3rd International Workshop on Semantic Digital Archives (SDA 2013)* (pp. 93-103). https://ceur-ws.org/Vol-1091/paper9.pdf

Beagrie, N. (2006). Digital curation for science, digital libraries, and individuals. *International Journal of Digital Curation, 1*. doi: 10.2218/ijdc.v1i1.2

Borgman, C. L. (2000). Whither, or wither, libraries? In: Borgman, C.L. *From Gutenberg to the Global Information Infrastructure: Access to information in the networked world*. (pp.169-208). Cambridge, MA: MIT Press.

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*, 17-30. doi: 10.1007/s00799-007-0022-9

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

Borgman, C.L., Golshan, M.S., Sands, A.E., Wallis, J.C., Cummings, R.L., & Randles, B.M. (2016). Data management in the Long Tail: Science, software and service. *International Journal of Digital Curation, 11*(1), 128-149. doi: 10.2218/ijdc.v11i1.428

Bosch-Bayard, J., Galan, L., Aubert Vazquez, E., Virues Alba, T., & Valdes-Sosa, P. A. (2020). Resting state healthy EEG: the first wave of the Cuban normative database. *Frontiers in Neuroscience, 14*, 555119. doi: 10.3389/fnins.2020.555119

Bush, V. (1945). As we may think. *The Atlantic Monthly, 176*(1), 101-108. https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/

Candela, L., Castelli, D., & Pagano, P. (2011). History, Evolution, and Impact of Digital Libraries. In I. Iglezakis, T. Synodinou, & S. Kapidakis (Eds.), *E-Publishing and Digital Libraries: Legal and Organizational Issues* (pp. 1-30). IGI Global Scientific Publishing. doi: 10.4018/978-1-60960-031-0.ch001

Chawinga, W. D., & Zinn, S. (2020). Research data management at an African medical university: Implications for academic librarianship. *The Journal of Academic Librarianship*, *46*(4). doi: 10.1016/j.acalib.2020.102161

Choudhury, S., & Kunze, J. (2009, May 18). *NSF DataNet: Curating Scientific Data*. 4th International Conference on Open Repositories. https://jscholarship.library.jhu.edu/handle/1774.2/34022

Collier, J., Schumacher, S., Behrens, C., Driemel, A., Diepenbroek, M., Grobe, H., et al. (2015). Rescued from the deep: Publishing scientific ocean drilling long tail data. *GeoResJ*, *6*, 17-20. doi: 10.1016/j.grj.2015.01.003

Costa, M. R., Qin, J., & Bratt, S. (2016). Emergence of collaboration networks around large scale data repositories: A study of the genomics community using GenBank. *Scientometrics*, *108*, 21-40. doi: 10.1007/s11192-016-1954-x

Cox, A. M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, *46*(4), 299–316. doi: 10.1177/0961000613492542

Cragin, M. H., & Shankar, K. (2006). Scientific data collections and distributed collective practice. *Computer Supported Cooperative Work (CSCW)*, *15*, 185-204. doi: 10.1007/s10606-006-9018-z

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. doi: 10.1098/rsta.2010.0165

Cutcher-Gershenfeld, J., Baker, K. S., Berente, N., Carter, D. R., DeChurch, L. A., Flint, C. C., ... & Zaslavsky, I. (2016). Build it, but will they come? A geoscience cyberinfrastructure baseline analysis. *Data Science Journal*, *15*, 8-8. doi: 10.5334/dsj-2016-008

Donaldson, D. R., & Koepke, J. W. (2022). A focus groups study on data sharing and research data management. *Scientific Data*, *9*(1), 345. doi: 10.1038/s41597-022-01428-w

Fegraus, E. H., Andelman, S., Jones, M. B., & Schildhauer, M. (2005). Maximizing the value of ecological data with structured metadata: an introduction to ecological metadata language (EML) and principles for metadata creation. *Bulletin of the Ecological Society of America*, *86*(3), 158-168. doi: 10.1890/0012-9623(2005)86[158:MTVOED]2.0.CO;2

Frischmann, B. M. (2012). *Infrastructure: The social value of shared resources*. Oxford University Press.

Gries, C., Hanson, P. C., O'Brien, M., Servilla, M., Vanderbilt, K., & Waide, R. (2023). The Environmental Data Initiative: Connecting the past to the future through data reuse. *Ecology and Evolution*, *13*(1). doi: 10.1002/ece3.9592

Hank, C., & Davidson, J. (2009). International Data curation Education Action (IDEA) Working Group: A report from the second workshop of the IDEA. *D-Lib Magazine*, *15*(3/4). doi: 10.1045/march2009-hank

Hanson, K. A., Almeida, N., Traylor, J. I., Rajagopalan, D., & Johnson, J. (2020). Profile of data sharing in the clinical neurosciences. *Cureus*, *12*(8). doi: 10.7759/cureus.9927

Heidorn, P.B. (2008). Shedding light on the dark data in the long tail of science. *Library Trends 57*(2), 280-299. doi: 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. doi: 10.1080/01930826.2011.601269

Heidorn, P. B., Stahlman, G. R., & Steffen, J. (2018). Astrolabe: curating, linking, and computing astronomy's dark data. *The Astrophysical Journal Supplement Series, 236*(1), 3. doi: 10.3847/1538-4365/aab77e

Hey, T., Tansley, S., & Tolle, Kristin (Eds.). (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery.* Microsoft Research. Retrieved from http://research.microsoft.com/en-us/collaboration/fourthparadigm/default.aspx

Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation, 2*(1). doi: 10.2218/ijdc.v3i1.48

Higgins, S. (2011). Digital curation: the emergence of a new discipline. *International Journal of Digital Curation, 6*(2). doi: 10.2218/ijdc.v6i2.191

Huang, C., Lee, J. S., & Palmer, C. L. (2020). *DCC curation lifecycle model 2.0: Literature review and comparative analysis.* University of Washington. http://hdl.handle.net/1773/45392

Kaushik, A. (2017). Perceptions of LIS professionals about the data curation. *World Digital Libraries, 10*(2). doi: 10.18329/09757597/2017/10207

Kouper, I. (2016). Professional participation in digital curation. *Library & Information Science Research, 38*(3), 212-223. doi: 10.1016/j.lisr.2016.08.009

Lesk, M. (2012). A personal history of digital libraries. *Library Hi Tech, 30*(4), 592-603. doi: 10.1108/07378831211285077

Lewis, K. P., Vander Wal, E., & Fifield, D. A. (2018). Wildlife biology, big data, and reproducible research. *Wildlife Society Bulletin, 42*(1), 172-179. doi: 10.1002/wsb.847

Lynch, C. A. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. *Portal: Libraries and the Academy, 3*(2), 327-336. doi: 10.1353/pla.2003.0039

Michener, W., Vieglais, D., Vision, T., Kunze, J., Cruse, P., & Janée, G. (2011). DataONE: Data Observation Network for Earth—Preserving data and enabling innovation in the biological and environmental sciences. *D-Lib Magazine, 17*(1/2), 12. doi: 10.1045/january2011-michener

Mokrane, M., & Parsons, M. A. (2014). Learning from the International Polar Year to Build the Future of Polar Data Management. *Data Science Journal, 13*. doi: 10.2481/dsj.IFPDA-15

Murillo, A. P. (2020). An examination of scientific data repositories, data reusability, and the incorporation of FAIR. *Proceedings of the Association for Information Science and Technology, 57*(1), e386. doi: 10.1002/pra2.386

Oliver, G., & Harvey, R. (2016). *Digital curation.* American Library Association.

Palmer, C. L., Cragin, M. H., Heidorn, P. B., & Smith, L. C. (2007). Data curation for the long tail of science: The case of environmental sciences. In *Third International Digital Curation Conference* (pp. 11-13). Retrieved from https://www.dcc.ac.uk/sites/default/files/documents/events/dcc-2007/posters_demos/long_tail_of_science.pdf.

Patterton, L., Bothma, T. J., & Van Deventer, M. J. (2018). From planning to practice: an action plan for the implementation of research data management services in resource-constrained institutions. *South African Journal of Libraries and Information Science*, *84*(2), 14-26. doi: 10.7553/84-2-1761

Pennock, M. (2007). Digital curation: A life-cycle approach to managing and preserving usable digital information. *Library & Archives*, *1*(1), 1-3. Retrieved from https://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf.

Plale, B., McDonald, R. H., Chandrasekar, K., Kouper, I., Konkiel, S. R., Hedstrom, M. L., Myers, J., & Kumar, P. (2013). SEAD Virtual Archive: Building a federation of institutional repositories for long-term data preservation in sustainability science. *International Journal of Digital Curation*, *8*, 172–180. doi: 10.2218/ijdc.v8i2.281

PLAN-E. (2018). *The long tail of Science and Data*. Platform of National eScience Centers in Europe. Retrieved from https://planeurope.files.wordpress.com/2018/10/report-plan-e-workshop-the-long-tail-of-science-and-data-version-1-0.pdf.

Ramachandran, R., Bugbee, K., & Murphy, K. (2021). From open data to open science. *Earth and Space Science*, *8*(5), e2020EA001562. doi: 10.1029/2020EA001562

Ramdeen, S., & Poole, A. H. (2017). Using grounded theory to understand the archival needs of geologists. In Gilliland, A. & McKemmish, S. *Archival Multiverse*. Monash University Press.

Reilly, S. K. (2014). Rounding up the data: Libraries pushing new frontiers. *Learned Publishing*, *27*(5), S33-S34. doi: 10.1087/20140506

Rodrigues, M. E., & Rodrigues, A. M. (2012). Analyzing the performance of an institutional scientific repository – A case study. *LIBER Quarterly: The Journal of the Association of European Research Libraries*, *22*(2), Article 2. doi: 10.18352/lq.8047

Schindler, U., Diepenbroek, M., & Grobe, H. (2012, April). PANGAEA®-Research Data enters Scholarly Communication. In *EGU General Assembly Conference Abstracts* (p. 13378). Retrieved from https://meetingorganizer.copernicus.org/EGU2012/EGU2012-13378-1.pdf.

Sinha, A. K., Thessen, A. E., & Barnes, C. G. (2013). Geoinformatics: Toward an integrative view of Earth as a system. In *The Web of Geological Sciences: Advances, Impacts, and Interactions: Geological Society of America Special Paper, 500*, 591-604.

Stahlman, G. R. (2022). From nostalgia to knowledge: Considering the personal dimensions of data lifecycles. *Journal of the Association for Information Science and Technology*, *73*(12), 1692-1705. doi: 10.1002/asi.24687

Stahlman, G. R. (2023, March). Is there a scientific digital divide? Information seeking in the international context of astronomy research. In *International Conference on Information* (pp. 514-523). doi: 10.1007/978-3-031-28032-0_39

Stahlman, G. R., & Heidorn, P. B. (2020). Mapping the "long tail" of research funding: A topic analysis of NSF grant proposals in the division of astronomical sciences. *Proceedings of the Association for Information Science and Technology*, *57*(1), e276. doi: 10.1002/pra2.276

Stahlman, G. R., & Kouper, I. (2024). Evolution of the "long-tail" concept for scientific data. *Journal of the Association for Information Science and Technology*. doi: 10.1002/asi.24967

Stephenson, M. H., Cheng, Q., Wang, C., Fan, J., & Oberhänsli, R. (2020). Progress towards the establishment of the IUGS Deep-time Digital Earth (DDE) programme. *Episodes Journal of International Geoscience*, *43*(4), 1057-1062. doi: 10.18814/epiiugs/2020/020057

Tibbo, H. R., & Lee, C. A. (2012). Closing the digital curation gap: A grounded framework for providing guidance and education in digital curation. In *Proceedings of Archiving 2012*. https://ils.unc.edu/callee/p57-tibbo.pdf

Van Deventer, M., & Pienaar, H. (2015). Research data management in a developing country: A personal journey. *International Journal of Digital Curation, 10*(2). doi: 10.2218/ijdc.v10i2.380

Wallis, J. (2014). Data producers courting data reusers: Two cases from modeling communities. *International Journal of Digital Curation, 9*(4). doi: 10.2218/ijdc.v9i1.304

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. doi: 10.1371/journal.pone.0067332

Wang, Q. (2018). Distribution features and intellectual structures of digital humanities: A bibliometric analysis. *Journal of Documentation, 74*(1). doi: 10.1108/jd-05-2017-0076

Weber, T., & Kranzlmüller, D. (2019). Methods to evaluate lifecycle models for research data management. *Bibliothek Forschung und Praxis*, 43(1), 75–81. doi: 10.1515/bfp-2019-2016

Witting, M. (2023). (Re-)use and (re-)analysis of publicly available metabolomics data. *PROTEOMICS*, *23*(23–24). doi: 10.1002/pmic.202300032

Yoon, A., & Donaldson, D. R. (2019). *Library capacity for data curation services: A US national survey*. Indiana University. https://scholarworks.iupui.edu/handle/1805/19849

Yu, X., Lamačová, A., Shu, L., Duffy, C., Krám, P., Hruška, J., ... & Lin, K. (2020). Data rescue in manuscripts: A hydrological modelling study example. *Hydrological Sciences Journal*, *65*(5), 763-769. doi: 10.1080/02626667.2019.1614593