Check for updates

Issues and paths forward in the identification and reuse of historic analog records

Bethany G. Anderson[1], Erin Antognoli[2†], Sandi L. Caldrone[1], Justin D. Derner[3], Shannon L. Farrell[4], Katrina Fenlon[5], John R. Hendrickson[6], Lois G. Hendrickson[4], Holly A. Johnson[6], Nicole E. Kaplan[7], Julia A. Kelly[4]*, Kristen L. Mastel[4] and Sarah C. Williams[1]

[1]University Library, University of Illinois Urbana-Champaign, Urbana-Champaign, IL, United States, [2]Agricultural Research Service, United States Department of Agriculture, Beltsville, MD, United States, [3]Agricultural Research Service, United States Department of Agriculture, Cheyenne, WY, United States, [4]University Libraries, University of Minnesota, Minneapolis, MN, United States, [5]College of Information Studies, University of Maryland, College Park, MD, United States, [6]Agricultural Research Service, United States Department of Agriculture, Mandan, ND, United States, [7]Agricultural Research Service, United States Department of Agriculture, Fort Collins, CO, United States

Introduction: Historic data, often in analog format, is a valuable resource for assessing effects of directional changes in climate and climatic variability. However, historic data can be difficult to locate, interpret, and reformat into a useful state.

Methods: Teams of scientists, librarians, archivists, and data managers at four US institutions have undertaken various projects to gather, describe, and in some cases, transform historic data. They have also surveyed researchers who either possess historic data or have used it in their work.

Results: Historic data projects involved locating data, writing data descriptions, and connecting with individuals who had knowledge about the data's collection. The surveys and interviews found that researchers valued historic data and were worried that it was at risk of loss. They noted the lack of best practices.

Discussion: Each project attempting to rescue or enhance access to historic data has a unique path but being guided by FAIR principles should be at the core whether or not the end result is machine-readable data. Working with a team incorporating librarians, archivists, and data managers can aid individual researchers' in producing accessible, and reusable datasets. There is much work to be done in raising awareness about the value of historic data but motivating factors for doing so include its usefulness in environmental research and other disciplines and its risk of loss as researchers retire and are unsure of how to save historic data, both in analog and electronic formats.

1 Introduction

As the value of historic data becomes more widely recognized and more of this data becomes available for use by current scientists, it is important to think beyond individual data sets and consider the larger context. Although individual researchers in various scientific fields have made use of historic data over a number of decades, often comparing it to newly-collected data, little has been written about the broader issues around the use of previously-collected data.

Some of the basics about historic data, which may also be referred to as dark data, legacy data, analog data, or numerous other names, are known. Much historic data exists only in hard copy (e.g., raw data sheets, notebooks) and the broader research community is often unaware of the availability of the data. Some data may be kept in institutional archives but much of it is still in labs or offices of the researchers who generated or inherited it and thus it is neither easy to discover nor safe from harm. Verification of the data may be an issue as well as outdated methods and data capturing practices that could make direct comparisons unworkable. When historic data is described well, it can be invaluable in comparing the past to the present. The ever-increasing amount of research being done to assess the effects of directional climate change and increasing climatic variability only makes this historic data more valuable (Magurran et al., 2010). For example, since longitudinal studies can be expensive and difficult to conduct, using previously-collected data can give researchers the chance to make comparisons between the older data and current data without having to secure new multi-year funding. Both time and money may be saved.

As scientists, librarians, and archivists who have undertaken a variety of projects that have identified, described, and in some cases, reformatted historical data, we have also tried to take a broader view of this type of work. We will discuss nine recent projects in the life sciences undertaken by one or more of the authors that have helped inform our perspectives. See Table 1. The projects fall into two broad categories:

- Projects studying historic data
- Projects using historic data

The projects that study historic data were undertaken to gain an understanding of what researchers were thinking and doing in regard to historic data. These projects include two sets of interviews with individuals who either possess historic data or work on projects to recover it; an online survey of researchers who hold historic data; and a literature review highlighting past efforts and concerns about historic data.

The projects that use data include various activities such as identifying, locating, organizing, describing, and reformatting the data as well as adding it to a repository. The data sets vary in size and not all of them had a machine-readable format as their end product. All have potential relevance to current environmental research although that was not necessarily the prime motivation for all project teams when they organized and transformed these datasets.

The objective of this study is to use the variety of historic data projects that the authors have undertaken to inform the future use of historic data among scientists. Potential stumbling blocks and successful strategies for working with this kind of data will be identified.

2 Methods

2.1 Projects studying historic data

The semi-structured interview projects include one with a broad perspective, both geographically and in regard to topics around historical data recovery, and another that was conducted at one institution with a much more narrow focus around individual life

TABLE 1 Brief description of individual analog data projects included in analysis.

Title	Type	Authors	Description
Recovering and Reusing Archival Data for Science	Studying historic data	U of Maryland	Interviews with a broad range of professionals working with historic data to identify curation practices and processes as well as challenges
Historic Scientific Analog Data and Life Sciences Faculty	Studying historic data	U of Minnesota	Survey and follow-up interviews covering what kind of analog data faculty members held, how much, if it was being used, and their concerns
Literature Review of Historic Scientific Analog Data	Studying historic data	U of Minnesota	Review of over 70 papers from the 1960s to the present that cover the value of utilizing historic data as well as the challenges
Cultivating the Scientific Data of the Morrow Plots	Using historic data	U of Illinois	Data from the Morrow plots (1876-present) on corn crop rotation and yields was identified, collected and curated
Long-Term Livestock Production Data	Using historic data	USDA	Data to support rangeland management was gathered from research stations across five states
Locating Historic Scientific Data in a University Archive	Using historic data	U of Minnesota	Pilot study investigating whether university archives hold historic data and if so, is it well described and does it have potential for reuse
Resurfacing Historical Fruit Breeding Data	Using historic data	U of Minnesota	Over 100 years of fruit breeding data was inventoried, categorized, further described and in some cases digitized
Data in student reports at the Itasca Biological Station	Using historic data	U of Minnesota	Nearly 3,000 papers were checked for raw data and if it was found, what species were covered and were collection methods described
Bird censuses at the Itasca Biological Station	Using historic data	U of Minnesota	Data about individual species from several hundred bird censuses, 1940–2010, was compiled for use by current and future researchers

sciences researchers' analog data and their attitudes about its preservation and reuse. In both cases interviews were recorded and transcribed, and then qualitatively coded to identify themes raised by each participant.

Researchers at the University of Maryland College of Information Studies took a broad approach by undertaking a project, entitled "Recovering and Reusing Archival Data for Science" (RRAD-S), that studied the opportunities and challenges confronting the use of historic data to support active scientific work across many disciplines and organizational contexts (Sorensen et al., 2022). Through interviews with 23 practitioners, including research scientists, science librarians, curators, and archivists working in a wide variety of institutions—from scientific research labs to digital humanities centers, national libraries to natural history museums—this project explored the broad range of data curation practices deployed in the processes of historical data recovery, and the cross-cutting challenges they face. Topics included how institutional policies and local processes impact historical data recovery, how curators and scientists assess the value of data prior to recovery, the data curation and management practices they deploy in the process, how they collaborate, and how they evaluate data recovery and reuse.

In an effort to get an idea of the scope of historic analog data held in offices and labs at one institution, an online survey of agricultural and life sciences researchers at the University of Minnesota was conducted, with 108 responses from 772 surveys sent out (Farrell et al., 2020). Questions focused on the nature of the data (topic, format, date range, amount, condition, perceived value) and whether it had been or might be reused. A subset of the 17 respondents were interviewed to learn more about both the nature of the data and the researcher's attitudes and concerns about it.

To better understand past attitudes about historic scientific data, particularly in analog format, and to learn about earlier efforts to identify and preserve it, an in-depth literature review was undertaken by a group of librarians and archivists at the University of Minnesota (Kelly et al., 2022). With few if any subject terms about historic data utilized by any of the large life sciences and agricultural article indexes, much of the discovery of papers relied on using known authors, reference lists from relevant papers, and papers that cited relevant papers.

2.2 Projects using historic data

The projects related to using individual historic data sets include two large undertakings. One focused on the Morrow Plots on the University of Illinois campus, which were established in 1876 and are the longest-running continuous experimental agricultural fields in the Americas. The Morrow Plots data project was initiated at the administrative level. Its goals were to curate the data and make it available for the 150th anniversary of the Morrow Plots and also to serve as an exemplar of good data management practices (Anderson et al., 2022). The other project is a U.S. Department of Agriculture (USDA) effort focused on historical livestock production data from 1916 to the present from five Western states (Sanderson et al., 2016; Long-Term Livestock Production data sets, 2023). Also noted are several smaller-scale projects conducted at the University of Minnesota including an inventory of selected collections in the

University Archives, an inventory/digitization initiative at the Horticultural Research Center, and two efforts to organize and enhance data collected at the Itasca Biological Station and Laboratories in northern Minnesota (Farrell et al., 2019; Farrell and Kelly, 2022; Farrell et al., 2023; Farrell and Kelly, 2023). These projects were initiated to locate potentially valuable climate-change adjacent analog research.

The nature of the data depended on the project, and in the two large efforts it was a mix of analog and a variety of digital formats. The University of Illinois project looked to locate any data related to the Morrow Plots. The USDA livestock effort had specific inclusion criteria: records 10 years or greater; data from experiments on rangelands or pasturelands containing key information on management; and livestock performance measured as weight gains and/or body condition scores. Data were preferred to be associated with individual animals. The University of Minnesota projects were smaller and focused on data in analog format. In the case of the Horticultural Research Center there was a small amount of proprietary data that was inventoried and described but not put in the pipeline for digitization.

Amongst the projects, input from data producers varied. It is not always possible to locate someone who knows the history of older data, details of the data collection or even to find a thorough written description of the methods. Projects fell along a spectrum from no input from data authors to working closely with researchers who knew the details of data collection. The large USDA livestock project required coordination with a key contact person from each of the original locations of the data, who would be familiar with the livestock management practices, current and past. In the Morrow Plots project, the group reached out to a variety of stakeholders, including plot managers, active and retired researchers, and administrators in a quest to locate as much background information as they could. The first University of Minnesota project that was conducted in the University Archives found quite a few data sets with no explanation of the methods, and the second, that tried to locate raw data in student papers from the Itasca Biological Station found some had detailed methods descriptions while others did not. It was not possible to consult the majority of researchers in the latter. Finally the University of Minnesota work at the Horticultural Research Center was done in close conjunction with two long-time fruit breeders.

In all of the projects involving rescuing historical data, the goal was to abide by the FAIR principles (Wilkinson et al., 2016), striving to make the data findable, accessible, interoperable, and reusable, as much as possible. Data was identified, described, organized, cleaned, and in most cases, reformatted. In the cases of the two large data projects (i.e., USDA livestock and Morrow Plots), the teams employed spreadsheets, R, and other tools to convert the data into tidy data (Wickham, 2014). This ensured that each row contained an observation and each column a variable, and that all variables were defined in a data dictionary.

3 Results

3.1 Projects studying historic data

Topics repeatedly mentioned by many of the survey or interview participants who either had historic data or worked with it included:

FIGURE 1
Year ranges for data projects.
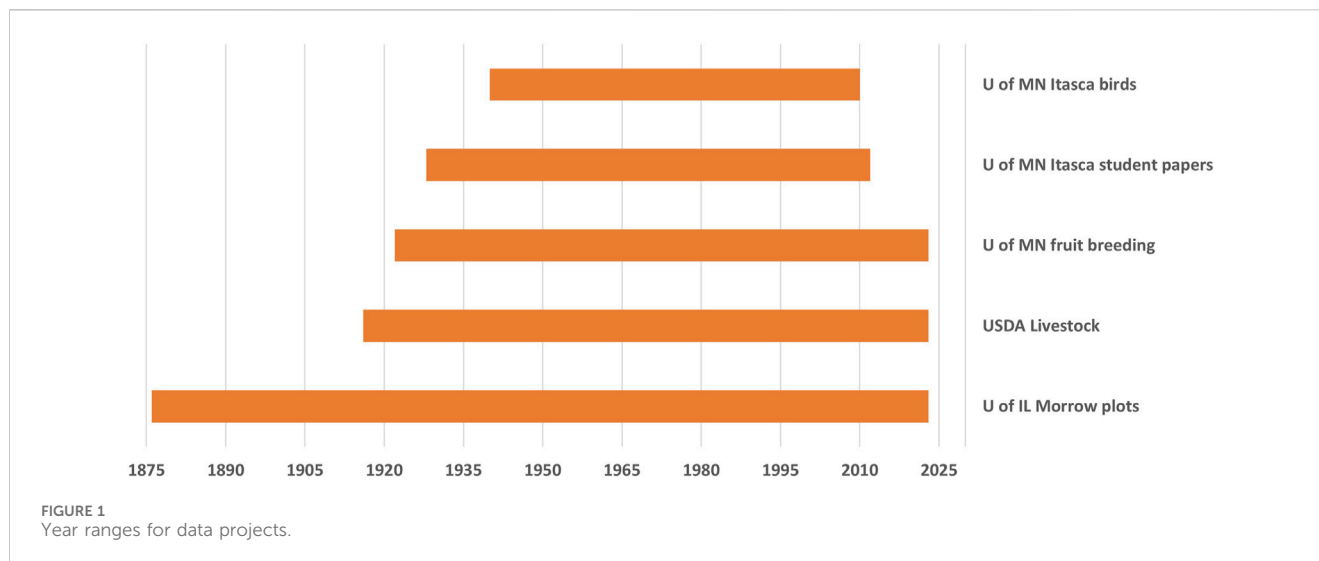
- The value of historic data
- The lack of solutions for long term preservation, and risk of loss
- The importance of best practices for transforming historic data
- The added challenges of working with older data as opposed to recently-generated data

Respondents and interviewees let the authors know that they and their colleagues held varying amounts of historic data ranging from one shelf to multiple file cabinets in a wide variety of formats. These included notebooks, data sheets, photographs, and a number of outdated electronic media.

Their concerns included ethical issues, such as the provenance of and access to data that potentially represent or reflect on marginalized communities, their knowledge or traditions, or sensitive sites and artifacts, as well as worries about who would serve as stewards of this older data once current researchers retire. They were aware of the many risk factors threatening the safety and stewardship of the data but few had ideas about how to move forward.

The literature review gave a picture of decades-long concern among scientists about the loss of historic data but few success stories of large scale data digitization and rescue to report. With a few noteworthy exceptions, papers were mostly focused on the situation in a narrow subdiscipline, often directed at other researchers in the same field. While both individual authors and a few professional associations took up the cause, there seemed to be little interest among universities, large research institutes, or funding agencies.

3.2 Projects using historic data

The data access and rescue-related projects covered a wide range of years (Figure 1). The end result in most cases is data that was dramatically transformed and much more likely to be understood and reused by future researchers. This is true in particular for the

Morrow Plots and USDA efforts. Data in multiple analog and digital formats was gathered together, sometimes from disparate geographic areas (Anderson et al., 2022). After extensive efforts to reformat and organize them, the end products were deposited in a freely accessible data repository, in machine-readable format with a Digital Object Identifier (DOI) for citation (Morrow Plots Data Curation Working Group, 2022). Descriptions of the data are quite detailed, including a data dictionary and extensive descriptions of methodologies, data collection locations, and any relevant extreme conditions or deviations in land management practices.

Another way that data access and rescue was approached for the three smaller University of Minnesota projects was through depositing the results in a data repository or institutional repository, but not transforming them to machine-readable data. For the Horticultural Research Center, logbooks of fruit breeding activities have been scanned and deposited in the institutional repository along with descriptive readme files. The Itasca Biological Station projects resulted in spreadsheets that direct users to research papers on either 1) bird species studied in the area (1940–2010) organized by species or family and 2) data contained in student papers written at the station (1928–2012) with descriptions including a brief summary of the study, amount and type of data, species studied, and locations. The inventory of the University Archives has not yet resulted in a publicly available data set. The project was an experiment to discover whether raw data existed in University Archives collections and if so, was it potentially reusable.

4 Discussion

Time is of the essence to locate and preserve these historical datasets as knowledge and responsibility for these baseline datasets changes. Project data in these studies frequently intersect with climatic events and have the potential to contribute to a greater understanding of how the past impacts the present and future and potential mitigation strategies in environmental science. Both the temporal and longitudinal data in these projects are unique and

TABLE 2 Necessary information to provide along with tidy data to make it FAIR.

FAIR principle	Information to provide
Find data	Title
	Description (Abstract)
	Author or Creator Name(s)
	Contact Name
	Publisher
	Publication Date
	Keywords
	Spatial Description
Provide Access	DOI for distribution
Enable Interoperability	Attribute name, description, relevant codes, and units in a data dictionary
Facilitate Reuse	Narrative descriptions of methods and any extreme conditions
	Links to code repositories for working with data

valuable resources that can be applied to questions, for example, about resource management and ecological changes. Themes of risk and urgency surfaced in the semi-structured interviews conducted by the University of Maryland. Format, workflow and policy changes were mentioned as reasons for the loss of historical data. The literature search confirmed that retirements and shifting priorities, along with the lack of institutional support and funding were also factors, and the University of Minnesota archives project demonstrated that locating data in institutional archives is challenging for a variety of reasons.

As the need for environmental data and information collected under various climatic conditions increases, major changes must be made to increase access to this historic data on a global scale. Two recent papers provide an example of how this can work. In a 2022 paper, a team reported on compiling a digitized data set with 61 years of daily temperatures and precipitation in Yangambi in the central Congo Basin (Yakusu et al., 2022). In a 2023 paper, another team used that data set in their work to characterize changes over time in woody understory species in the region (Hatangi et al., 2023).

Data curation strategies in the projects differed. However, two key themes emerged: approaches need to be adaptive and diverse and collaborative teams are needed. The USDA study underlined the critical role and value of domain specialists, while the Illinois project identified that making this kind of work sustainable requires collaborative efforts. Identifying large-scale solutions and creating structures will require the collaborative efforts of researchers, institutions, professional societies, data managers, librarians and archivists. Ultimately this can relieve researchers from being the sole caretakers or gatekeepers of historic climate-related datasets. Possible directions include embedding data managers in projects or at research locations (Baker and Karasti, 2018; Kaplan et al., 2021) to help bridge gaps between research programs/projects and institutional repositories and adopting metadata and citation standards.

FAIR principles guided all of these projects, and despite the fact that not all resulted in machine-readable data, the outputs were made discoverable and accessible to a broader audience.

Table 2 summarizes necessary information to provide along with tidy data to make it FAIR. Reasons for choosing not to transform data were often related to lack of resources. Looking to the future, what the studies revealed is there is no one right way to curate the data. Instead, creating good documentation, organizing the data, or going as far as making the data tidy, well defined, and configured in such a way to be open and machine readable, is what will allow for others to build on it in the future. A minimum set of information can facilitate reuse when inevitable technology changes happen. A data dictionary could be developed to organize relevant information. Suggested elements include creator, description, publication date, variable definitions, and other items listed in Table 2.

However, the projects all revealed that this work is a heavy lift. Making historic data discoverable and usable not only requires significant effort but also a variety of expertise. For example, the Morrow Plots project involved using R to clean and package the data, GitHub, and the cooperation of institutional repositories. In addition, the Morrow Plots, USDA and horticultural projects required working directly with scientific researchers or having extensive scientific knowledge to manage and organize the data. Methods used to transform the data should be documented in a readme file. This could also include, for example, a change log, paying particular attention to any changes made, including deletions.

Formatting and gathering data and information as presented in Table 1 for ingest can be undertaken by librarians, data curators and embedded data managers. However, the interviews revealed this task often falls to graduate research assistants who encounter inconsistent data gathering, lack good procedural documentation, and struggle serving as an intermediary between the person(s) who collect data and the person(s) who deposit it. The Morrow Plots data curators attempted to address this by providing the R code used to clean and compile disparate data sets, which can be adapted to other projects and includes a narrative description of each step of the process.

Policy changes and open data mandates that require researchers to spend time organizing and making their data available apply only

to contemporary data, not historic analog data. In a classic chicken or egg scenario, data in an analog form is not discoverable so cannot be used. When made discoverable through descriptions and indexes, even if data are not fully FAIR, they are more accessible and can lead to reuse. In any case, new mandates for sharing data have inspired researchers to consider the value and utility of historic data.

There were a few key questions that arose from the projects and discussion amongst the authors regarding perceptions of what motivates researchers to organize and preserve historic data, as well as engage in place-based data rescue projects. Data rescue efforts are very time-consuming and challenging and participants are busy professionals with impact measured as outcomes other than well-described and available legacy data. Given this, what factors might motivate or incentivize more work on analog data? In the case of the USDA project, serendipity played a role. The rangeland ecology and livestock production scientific communities identified a need for historic data. Connections among various research locations helped identify potential longitudinal data sets, while embedded data managers working with curators at the National Agricultural Library were able to facilitate the data work. Researchers knew of other locations with longitudinal data, and had worked with a data manager. The USDA team was able to share some approaches and have gathered stories from data providers. Their stories demonstrated the importance of public sector information available to decision-makers and land managers developing management strategies for grazing lands. This has been considered a successful team approach, and other USDA locations and centers are now considering formation of data teams to help identify and publish data from other research locations. The motivation for data from the Morrow Plots at the University of Illinois emerged in conjunction with a celebration of milestone or anniversary of the establishment of the plots. A Data Curation Group recognized this milestone and potential reuse of the data and formed a working group to preserve its legacy.

How might Digital Object Identifiers (DOIs) for datasets further motivate this work? Attribution matters to researchers and as more research journals require a data DOI, will this increase motivation to receive credit for other research outputs, including analog datasets? DOIs with attendant figures on use and downloads can demonstrate the value of research to administrators, impact tenure and promotion criteria, and allow for continued support for research programs. With the advent of new types of data set metrics, beyond downloads, could this historic data evolve to become an institutional asset, where every data provider can be attributed and data set published can be cited with its own (DOI) from an institutional repository. With increased sharing of data, researchers may want their older legacy data to be available along with their new datasets, perhaps so that they and others know where to find it in the future.

5 Conclusion

Findable and shareable data would increase the value proposition for analog data. Requests for this data would no longer need to be mediated by scientists or data curators, allowing for easier and more efficient access. New

perspectives on the advantage of historic data that are organized and readily available could attract new post-docs and researchers. Data that is organized and easily accessible from the start is an asset to researchers, allowing for increased productivity and innovation. For example, analog data sets could be used to inform decision support tools for questions related to making new adaptations to climate change. If models were developed for curation of historic analog data, more datasets would be made available, potentially leading to reuse. As scientists receive stronger encouragement to do this work, recognizing that historic analog data is valuable and can provide critical information about the effects resulting from climate change among other things, perspectives on the stewardship, management, and reuse of historic data may shift.

Awareness of the issue is one part of the solution, but what else can researchers do to help preserve these valuable assets? It may not be possible to digitize or reformat all historic analog data, and some level of rescue or preservation is better than none. Working with experts to identify and prioritize historic datasets will be necessary as not everything can be saved. Although the FAIR principles should of course guide all efforts, it will not always be achievable or feasible. There is no prescriptive solution, but there are steps that can be taken to make these datasets more discoverable and accessible. There are a range of options.

- Keep the data analog but make sure it is well organized and described.
- Scan the data to remove the need for or risk to the analog copy.
- Scan the data and provide an adequate description.;
- Scan, describe, and transform the data to a machine-readable, interoperable format and provide a data dictionary and an adequate description.

Some researchers may still want to consult the original analog format, so determining where and how to house that data may also be important. Every situation, and every dataset, is unique. Building relationships with data managers, librarians, archivists, experts and others who have gone through the process will be inherently beneficial.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: Our paper uses a variety of earlier work by the authors to discuss issues relating to historic data. The data from the earlier works may be found in these repository entries: https://doi.org/10.13020/K81F-Q625, https://doi.org/10.13020/EENG-X538, https://agdatacommons.nal.usda.gov/Long_Term_Livestock_Production, https://doi.org/10.13012/B2IDB-7865141_V1.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and

institutional requirements. Written informed consent from the [patients/participants OR patients/participants legal guardian/next of kin] was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

BA: Writing–review and editing. EA: Writing–review and editing. SC: Writing–review and editing. JD: Writing–review and editing. SF: Writing–original draft. KF: Writing–review and editing. JH: Writing–review and editing. LH: Writing–original draft. HJ: Writing–review and editing. NK: Writing–review and editing. JK: Writing–original draft. KM: Writing–review and editing. SW: Writing–review and editing.

Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author Disclaimer

EA is currently employed as a data specialist with the Congressional Research Service (CRS). This article was prepared prior to her CRS employment. The views expressed herein do not represent those of CRS or the Library of Congress.

References

Anderson, B., Caldrone, S., Henry, J., Imker, H., Luong, H., Trei, K., et al. (2022). *Cultivating the scientific data of the Morrow plots: visualization and data curation for a long-term agricultural Experiment_20220915.* doi:10.17605/OSF.IO/X8YMS

Baker, K. S., and Karasti, H. (2018). Data care and its politics: designing for local collective data management as a neglected thing. *Proc. 15th Participatory Des. Conf. Full Pap.* 1 1–12. doi:10.1145/3210586.3210587

Farrell, S., and Kelly, J. (2022). *Bird species included in bird censuses conducted by students and faculty and staff researchers at the University of Minnesota Itasca Biological Station and Laboratories 1940-2010.* doi:10.13020/EENG-X538

Farrell, S., and Kelly, J. (2023). *Data contained in written reports of research conducted by students at the University of Minnesota Itasca Biological Station and Laboratories, 1928-2012, with links to the related reports.* doi:10.13020/K81F-Q625

Farrell, S., Kelly, J., Hendrickson, L., and Mastel, K. (2023). *A pilot study to locate historic scientific data in a university archive.* ISTL. doi:10.29173/istl2728

Farrell, S. L., Hendrickson, L. G., Mastel, K. L., Allen, K. A., and Kelly, J. A. (2019). Resurfacing historical scientific data: a case study involving fruit breeding data. *JeSLIB* 8, e1171. doi:10.7191/jeslib.2019.1171

Farrell, S. L., Hendrickson, L. G., Mastel, K. L., and Kelly, J. A. (2020). Historical scientific analog data: life sciences faculty's perspectives on management, reuse and preservation. *Data Sci. J.* 19, 51. doi:10.5334/dsj-2020-051

Hatangi, Y., Nshimba, H., Stoffelen, P., Dhed'a, B., Depecker, J., Lassois, L., et al. (2023). Leaf traits of understory woody species in the Congo Basin forests changed over a 60-year period. *plecevo* 156, 339–351. doi:10.5091/plecevo.104593

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

Kelly, J. A., Farrell, S. L., Hendrickson, L. G., Luby, J., and Mastel, K. L. (2022). A critical literature review of historic scientific analog data: uses, successes, and challenges. *Data Sci. J.* 21, 14. doi:10.5334/dsj-2022-014

Long-Term Livestock Production data sets (2023). Available at: https://agdatacommons.nal.usda.gov/Long_Term_Livestock_Production.

Magurran, A. E., Baillie, S. R., Buckland, S. T., Dick, J. M. P., Elston, D. A., Marian Scott, E., et al. (2010). Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. *Trends Ecol. Evol.* 25 (10), 574–582. doi:10.1016/j.tree.2010.06.016

Morrow Plots Data Curation Working Group (2022). *Morrow plots treatment and yield data.* doi:10.13012/B2IDB-7865141_V1

Sanderson, M. A., Liebig, M. A., Hendrickson, J. R., Kronberg, S. L., Toledo, D., Derner, J. D., et al. (2016). A century of grazing: the value of long-term research. *J. Soil Water Conservation* 71, 5A–8A. doi:10.2489/jswc.71.1.5A

Sorensen, A. H., Fenlon, K., Escobar-Vredevoogd, C., and Wagner, T. L. (2022). "Recovering and reusing scientific data: investigating data curation practices across disciplines," in *Proceedings of the ALISE annual conference.* doi:10.21900/j.alise.2022.1073

Wickham, H. (2014). Tidy data. *J. Stat. Soft.* 59. doi:10.18637/jss.v059.i10

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 160018. doi:10.1038/sdata.2016.18

Yakusu, E. K., Acker, J. V., De Vyver, H. V., Bourland, N., Ndiapo, J. M., Likwela, T. B., et al. (2022). Six decades of ground-based climate monitoring indicate warming and increasing precipitation seasonality and intensity in Yangambi (central Congo basin) (preprint). *Review.* doi:10.21203/rs.3.rs-1968285/v1