



# A Critical Literature Review of Historic Scientific Analog Data: Uses, Successes, and Challenges

JULIA A. KELLY

SHANNON L. FARRELL

LOIS G. HENDRICKSON

JAMES LUBY

KRISTEN L. MASTEL

\*Author affiliations can be found in the back matter of this article

REVIEW

]u[ubiquity press

## ABSTRACT

For years scientists in fields from climate change to biodiversity to hydrology have used older data to address contemporary issues. Since the 1960s researchers, recognizing the value of this data, have expressed concern about its management and potential for loss. No widespread solutions have emerged to address the myriad issues around its storage, access, and findability. This paper summarizes observations and concerns of researchers in various disciplines who have articulated problems associated with analog data and highlights examples of projects that have used historical data. The authors also examined selected papers to discover how researchers located historical data and how they used it. While many researchers are not producing huge amounts of analog data today, there are still large volumes of it that are at risk. To address this concern, the authors recommend the development of best practices for managing historic data. This will take communication across disciplines and the involvement of researchers, departments, institutions, and associations in the process.

## CORRESPONDING AUTHOR:

**Julia A. Kelly**

University of Minnesota, USA

[jkelly@umn.edu](mailto:jkelly@umn.edu)

## KEYWORDS:

Analog data; Historic data;  
Data policies; Risk of loss;  
Data rescue; Dark data

## TO CITE THIS ARTICLE:

Kelly, JA, Farrell, SL,  
Hendrickson, LG, Luby, J and  
Mastel, KL. 2022. A Critical  
Literature Review of Historic  
Scientific Analog Data: Uses,  
Successes, and Challenges.  
*Data Science Journal*, 21: 14,  
pp. 1–11. DOI: [https://doi.  
org/10.5334/dsj-2022-014](https://doi.org/10.5334/dsj-2022-014)

## INTRODUCTION

Concerns about the management of data, including its preservation, findability, and reuse, are almost entirely focused on recently-generated data in electronic, machine-readable formats. While many of the principles of the management of electronic data such as proper description and good organization apply to data in any format, the discussions about applying those principles to older data in non-electronic formats have not received much attention.

In this paper we review publications in various scientific fields that discuss older data that is in analog or print format and the use or reuse of older data in general. By analog data we mean items in print format such as numeric data as well as field or lab notebooks, photographs, drawings, and maps. Analog data may also be called historic data, legacy data, heritage data, or dark data, although these and other phrases can include older data that is not necessarily in print format. Some authors also use the term 'data rescue' which has also been used to describe recent efforts to duplicate and secure electronic data that may be at risk of loss (see Data Refuge: <https://www.datarefuge.org/>).

Our interest in this topic began when a few senior faculty members approached the University library for assistance in organizing and possibly housing their analog data (Farrell et al. 2019). A survey of life sciences researchers on campus revealed that many held analog data and considered it valuable but were unsure of how to preserve it (Farrell et al. 2020). Nearly all were willing to share it. Given that most researchers now either collect data digitally or quickly transfer any analog data, this is a finite problem, but because many of the stewards of analog data are nearing retirement, it is timely. We undertook this literature review to learn how scientific researchers are dealing with the analog data in their possession and if any large scale efforts have been undertaken to address the issues.

## TYPES OF ANALOG DATA

Much of the analog data that exists in offices, labs, homes, archives, and other locations is numeric in nature. It was probably collected before electronic spreadsheets were commonly available for both capturing and analyzing data. The format could be loose notebook paper, index cards, large data sheets, or bound or unbound notebooks. It could also take the form of a log, possibly combining numeric and descriptive data in chronological order.

The data may also be descriptive in nature and contained in field notebooks or diaries. The tags associated with museum and herbarium specimens are often mined for the data that they note such as species, location, dates, and other parameters. Although they are inextricably tied, when we discuss analog data we are not including the specimens themselves but just the information on the tags.

Drawings and photographs may accompany other forms of data or may stand on their own, hopefully with enough description to make them useful to current researchers. The same is true of maps, which may be printed or hand drawn.

## HISTORY OF CONCERN ABOUT ANALOG DATA

A number of authors have written about analog data over the last 50+ years, often noting its potential value and lamenting the lack of procedures, funding, and best practices to help support its ongoing use and preservation. Psychologists in the 1960s and 1970s noted not only the importance of new observations coming from the re-examination of older data but also the practice of comparing newly-gathered data to historic data (Johnson 1964; Craig & Reese 1973). Speaking about data that authors have not retained, Wolins (1962) suggests a role for professional associations: 'If it were clearly set forth by the APA [American Psychological Association] that the responsibility for retaining raw data...this dilemma would not exist'. For a time the U.S. government played a role through the American Documentation Institute at the Library of Congress, which accepted some raw data to be preserved (Craig & Reese 1973). Recently, Buma made use of photographs in Glacier Bay to longitudinally track plant growth and establishment and noted that if it was easier to learn of the existence of older data and to obtain copies, its value would grow (Buma 2018; Buma et al. 2019).

In most cases authors limit their discussions to the situation in their own subspecialty although a few have taken a broader view. A notable example is the final report of the Ecological Society of America (ESA) committee on the Future of Long-term Ecological Data (Gross & Pake 1995). The lengthy report details the situation as well as offers numerous recommendations for the future. Although it does not exclusively focus on analog data, it states ‘[a]mong the least secure are data in the hands of an individual researcher who has made little or no provision for long-term curation’.

Also in 1995, the National Research Council published both ‘Finding the Forest in the Trees: The Challenge of Combining Diverse Environmental Data’ and ‘Preserving Scientific Data on Our Physical Universe’ (National Research Council (U.S.) 1995a; National Research Council (U.S.) 1995b). The former report highlights the variables, measures, and data management, and puts forward 18 recommendations. These call on professional societies, research institutions, funding agencies, and individual researchers to collaborate, plan carefully, focus on interoperability, create rich metadata, and make data more widely available. The latter report notes many problems and few solutions, stating ‘[t]he most important deficiencies are in the documentation, access, and long-term preservation of data in usable form.’ Again, analog data was not the focus of these works but it was covered.

Easterday et al. (2018) notes the ‘potential of historical dark data to contribute to the modern digital ecological data landscape’. She notes the importance of metadata and the need to promote the data and the best practices around it. In his book *Repurposing Legacy Data*, Berman (2015) states that ‘data repurposing creates value where none was expected’. It includes case studies from a variety of disciplines and has chapters on identifying data that might lend itself to repurposing and understanding the organization of older data.

Griffin (2015) advocates for the value of ‘heritage data’, noting that much of it is at risk and in order to secure it for future use, ‘certain priorities need to be re-ordered, new skills acquired and taught, resources redirected, and new networks constructed’. Griffin was active in the CODATA Data at Risk Task Group which, along with its successor, the Research Data Alliance’s Data Rescue Interest Group, worked to highlight the value of older data and promote projects that used or preserved it (<https://codata.org/blog/2015/07/02/data-at-risk-and-data-rescue/>). Patil and Siegel (2009) note that bringing more dark data to the forefront will require different incentives from all those involved: ‘journals, citation indexes, funding agencies, academic institutions and, not least, the researchers themselves’. Although they write from a health sciences perspective, this probably applies more broadly.

A number of authors have drawn attention to the use or potential use of analog data in their particular fields. In fisheries, Singer and his co-authors (Singer, Ellis & Page 2020) surveyed fellow researchers to get a better idea of how and why they used fish collections in order to inform both researchers and those who manage the collections. The value and possible reuse of data collected at biological field stations has been noted since at least the 1980s (Bowser 1986). Bowser emphasized the importance of data management and suggested that field station data might be deposited with libraries, historical societies or federal agencies. Easterday and colleagues (2018) make their observations about the use of data science principles by highlighting work from three California field stations and Michener and colleagues (2009) wrote an article entitled ‘Biological Field Stations: Research Legacies and Sites for Serendipity’.

Ecological researchers have long mined analog data and historical records in their work, according to Beans (Beans 2018). While she focuses on journal entries, maps, and photos, she highlights common challenges such as locating material and working with someone else’s organizational scheme. She highlights Loren McClenachan (2009, 2017, 2012), a marine ecologist who utilizes historical data in her research and also published a policy-oriented article on the benefits of using older data to set baselines in marine studies. Over 20 years ago Olson and McCord (1998, 2000) wrote two book chapters on data archiving in the ecological sciences. Although the emphasis is on digital data, they spell out recommendations on incentives, metadata, and components of an archive that apply to analog and digital materials.

Kwok (2017) reports on the use of older data in the fields of both ecology and climate science. In the area of climate science, Brönnimann et al. (2018) are mainly concerned with digital data but provide an overview of efforts to locate and digitize analog data, commenting that ‘the fraction of yet-to-be-digitized data is difficult to quantify’, implying that it is large indeed.

Geological researchers sometimes have an added reason to want to discover and use older data—it may have been collected using methods that are now difficult or impossible to employ due to stricter regulations. Diviacco et al. (2015) writes about a project where data was both analog and digital and had been obtained using dynamite. Vearncombe et al. (2016, 2017), using examples from the mining industry, note that ‘upcycling’ of data can mean cost savings as well as new insights from reexamination of data.

A number of disciplines have employed citizen science projects to assist in the analog data efforts. These take the form of both mining older citizen science projects for their data or initiating new projects that provide person-hours to reformat or otherwise transform or collate analog data. Clavero and co-authors (2014, 2017) examine species lists to study trout decline, Hof and Bright (2016) look at previous counts of hedgehogs, and Snall et al. (2011) consider the use of presence data from bird monitoring. A recent citizen science project on the Zooniverse platform involves identifying data in papers written by students at the University of Michigan Field Station (<https://www.zooniverse.org/projects/jmschell/unearthing-michigan-ecological-data/about/faq>).

While many authors bemoan the unfortunate state of older data in their subdisciplines, a few areas offer success stories. Researchers working in biodiversity, many of whom are connected with museums or herbaria which hold physical specimens and their metadata-rich identification tags, are an example. They have built networks and secured funding for several international biodiversity-related projects that address data tied to specimens as well as the objects themselves. Projects include Integrated Digitized Biocollections (iDigBio, <https://www.idigbio.org/>), Global Biodiversity Information Facility (GBIF, <https://www.gbif.org/>), and Distributed System of Scientific Collections (DiSSCo, <https://www.dissco.eu/>). The progress in digitization and dissemination of biodiversity data over the last 20 years is summarized by Nelson and Ellis (2019).

Climate researchers have also made great strides in gathering disparate data in analog and digital format and making it accessible to the global community of scientists. The EU-based Copernicus Climate Change Service (C3S, <https://datarescue.climate.copernicus.eu/>) and International Data Rescue Portal (I-DARE, <https://www.idare-portal.org/>) serve as examples.

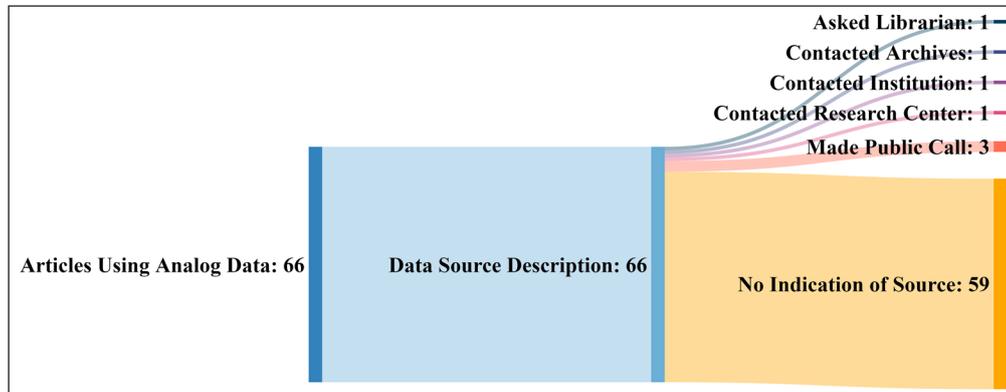
Some contemporary groups that rescue and reuse older analog data have very narrowly focused subject areas. The Living Data Project (<https://www.ciee-icee.ca/data.html>), sponsored by the Canadian Institute of Ecology and Evolution, funds new projects each year with topics such as ‘Species ranges, diversity and life history of Neotropical birds’ and ‘Responses of freshwater zooplankton to road salt pollution: A global perspective’. Another project, based at the USDA National Agricultural Library (Data Rescue Case Study: Long-Term Livestock Production Data), gathered older data from throughout the US, converted it to electronic formats and deposited it in AgData Commons (Patton et al. 2022).

Field and lab notebooks have been the focus of a number of digitization projects. They may be held in archives, museums, libraries, or research facilities as well as by individuals. The Biodiversity Heritage Library, in conjunction with several other institutions including the Smithsonian, includes nearly 3,000 scanned field books (<https://www.biodiversitylibrary.org/collection/FieldNotesProject>). On a smaller scale, Texas A&M Libraries has digitized the field notebooks and specimen catalogs of W. B. Davis (1930–1981) and they have been viewed over 1,000 times (<https://oaktrust.library.tamu.edu/handle/1969.1/129120>). Thomer et al. (2012) proposes a method for efficiently extracting species data from handwritten field notebooks.

## WAYS THAT OLDER ANALOG DATA IS UTILIZED

Researchers may use older data in a variety of ways. Some strive to repeat an earlier survey or experiment as closely as possible (Lannoo et al. 1994; Gent & Morgan 2007; Hédli, Petřík & Boublik 2011; Riddell et al. 2021). Others reexamine older data or incorporate portions of it into their current work (Trisurat et al. 2020; Azeria et al. 2006; Brodman, Cortwright & Resetar 2002; Fellers & Drost 1993). Authors may also have consulted earlier data as they developed their research plans. Mandates for the preservation of data that have emerged in the last 15 years have elevated the topic of data reuse, although most recent research has considered only digital data (Curty et al. 2017; Khan, Thelwall & Kousha 2021; Yoon & Kim 2017).

The methods that researchers use to obtain older data often remains a mystery. Large data collections such as iDigBio provide background, training, examples, and other resources for potential data users (<https://www.idigbio.org/research>) and authors are likely to mention or cite these collections. This is often not true for projects that use older data. In a preliminary investigation the authors conducted examining 66 scientific papers that used analog data, only seven spelled out how the authors located it (see [Figure 1](#)). None of the authors of this set of papers mention going back to the original authors of the publications to obtain more detailed information although it is hard to imagine that none of them took that step.



**Figure 1** Description of sources of historic data for scientific researchers who had re-used it in publications. Scientific papers (N = 66) that illustrated evidence of use of this data were examined to determine the source of the data and how it was identified and located by the researchers.

Obtaining data directly from the researchers is known to be problematic and a statement such as ‘data available on request’ in an article does not always lead to success. A 2014 study focused on 500+ articles from two to 22 years old and the authors state ‘[o]ur results reinforce the notion that, in the long term, research data cannot be reliably preserved by individual researchers’ (Vines et al. 2014). A new study suggests that all data associated with open data publishing needs to go into an open repository before publication. Of authors who indicated that data were available on request in publications, 1,670 (93%) did not respond to a request for data or chose not to share (Gabelica, Bojčić & Puljak 2022).

In addition to individual researchers, various types of organizations may be in possession of analog data. Government agencies hold weather data as well as the aforementioned museum and herbarium records. Fisheries and agricultural records have also been used along with conservation-related documents (Cardinale et al. 2015; Chauvel et al. 2012; Edwards & Contreras-Balderas 1991; Chuine et al. 2004; Smith & Jones 2007). Nonprofits may hold data from citizen science projects (Hof & Bright 2016). Archives can also be a source for analog data; this is sometimes where researchers discover field books and diaries (Llasat et al. 2005; Ledneva et al. 2004). They may also hold photographs used by those conducting repeat photography work (Lorenz et al. 1993; Rogers 1984; Webb, Boyer & Turner 2010).

There are numerous examples of authors reusing analog data that they located using less conventional sources. This includes literature, ship logs, tax records, newspapers, and church records (Primack & Miller-Rushing 2012; Lescauwat 2013; Brazdil et al. 2016; Van Der Veken, Verheyen & Hermy 2004; Martin, Brown & Young 2004; Sharma et al. 2016; Kelso & Vogel 2007).

## CHALLENGES WITH REUSING OLDER DATA

Scientists note the challenges and potential pitfalls when combining or comparing old and new data. There are few standards or best practices (Thomer et al. 2012; Berman 2015; Gross & Pake 1995). Individual authors provide rich insights from their experiences but finding general guidance is mostly lacking, unlike the situation with digital data (Bowser 1986; Weibe & Allison 2015). Also unlike the digital data landscape, ownership and stewardship responsibilities are often unclear. Costs for reformatting and preservation of analog data can be high with few options for funding (Gross & Pake 1995; Griffin 2015). Institutions have few incentives to save data (Pullin & Salafsky 2010).

Although not much is known about how individual researchers find the analog data that they reuse, several authors note the difficulty in locating it. It languishes in labs, gets redistributed to multiple locations, or disappears (Easterday et al. 2018; Duckworth, Grayce & Thornhill 2018; Curry 2011; Wicherts et al. 2006; Downs & Chen 2017). Many data repositories, especially those

housed at academic institutions, require data to be in machine-readable formats. Some such as AgData Commons (<https://data.nal.usda.gov/>) will accept scanned data. Zenodo (<https://zenodo.org/>), the European Community repository that includes data as well as software and documents, welcomes data in any format although they are currently working on guidelines for deposit. Data registries, where metadata about analog data could reside, have not materialized as predicted (Sheffield et al. 2011).

There are numerous challenges as researchers bring together data gathered years apart. Combining old and new data sets can be complex (National Research Council 1995b; Loehle & Weatherford 2017; Magurran et al. 2010). Metadata is a concern, as the original description may lack elements and they may have been defined differently in the earlier work (Bowser 1986; Reznick, Baxter & Endler 1994; Knapp, Bates & Barkstrom 2007; Löffler et al. 2021; Wiebe & Allison 2015; Sprague, Oelsner & Argue 2017).

Interpreting historic data may involve assumptions and comparison methods that need to be selected carefully (Kery et al. 2006; Pollock 2006; Rivadeneira, Hunt & Roy 2009; Huisman & Millar 2013; Kullman 2010). Engelhard, Righton, and Pinnegar (2014), studying the distribution of North Sea cod, noted ‘the well-known problems with fisheries data such as discarding and misreporting practices by fishers’. Beans (2018) notes that this underreporting was often due to attempts to minimize taxes on a boat’s catch. Historical records may have biases that must be dealt with when comparing with current surveys (Schulte & Mladenoff 2001; Delisle et al. 2003; Smith & Jones 2007; van Bavel et al. 2019).

## POSSIBLE PATHS FORWARD

While the individuals who are the current stewards of analog data and the organizations where they work have major parts to play in the solution to this issue, other entities can also take a role in developing solutions. Although few professional societies are in a position to host a data repository, there are other important roles that they can play. They could investigate and report on the status of analog data availability, use, and status in their realm, like the ESA. If they publish journals they could encourage authors to cite data papers and accept data papers where appropriate. Societies could call on their members to describe and preserve their own analog data. They could also endorse standards for metadata. If they have the financial means, they could fund the digitization of selected data.

Funding agencies already play a large role in the preservation and reuse of recently-produced electronic data through their mandates and they could also play a role with older data. Agencies could encourage pre-mandate grant recipients to make their data available or follow the lead of the USDA which welcomes scanned as well as machine-readable data from pre-mandate grant recipients in its AgData Commons repository. Agencies could promote the idea of data papers for material from earlier grants and endorse particular repositories in their subject areas. Funders could also award grants to projects that preserve analog data or make it more easily findable.

What can individual researchers do? They can organize, inventory, and describe any analog data in their purview and document details about how it was generated (Faniel, Frank & Yakel 2019). Many standards exist for doing this with digital data and those can be used for analog data as well. In a survey of those holding analog data, many reported that there was a person who could describe the origins of the data but fewer had documented that information (Farrell et al. 2020). If you have used historic data, think about how you found it and how you wish you might have been able to find it. Explore the concept and content of data papers and think about whether you might have some older data that you could describe in that same way. Talk to others about the topic and look for commonalities, especially across disciplines in your organization. Consult with the science librarians at your institution to see how you might work together. Think about your professional societies and how they might play a role.

## CONCLUSIONS

Researchers across the sciences use older data in analog format but little is known about how they learn of its existence or locate it. Over the last 50+ years authors have expressed concern about its fate and noted challenges with its use. With the exception of the community of

biodiversity researchers, there have been few large projects to address the preservation and findability of analog data and little interest expressed by government agencies, professional associations, and academic and research institutions that would be in a position to act on a broader scale. The best practices (including selection of metadata schema, developing a data dictionary, describing data collection methods) and policies developed to govern the preservation and dissemination of digital data could serve as an example for developments concerning analog data. In the digital realm best practices are often developed by professional associations, both disciplinary and data-focused, as well as those who manage data repositories.

## COMPETING INTERESTS

The authors have no competing interests to declare.

## AUTHOR CONTRIBUTIONS

All authors made significant contributions to the design of this review as well as drafting and revising the manuscript. All have approved this final version, agreed to be accountable, and have approved of the inclusion of those in the list of authors.

## AUTHOR AFFILIATIONS

**Julia A. Kelly**  [orcid.org/0000-0003-0796-0461](https://orcid.org/0000-0003-0796-0461)  
University of Minnesota, USA

**Shannon L. Farrell**  [orcid.org/0000-0002-9452-8345](https://orcid.org/0000-0002-9452-8345)  
Natural Resources Library, University of Minnesota, USA

**Lois G. Hendrickson**  [orcid.org/0000-0003-2896-7136](https://orcid.org/0000-0003-2896-7136)  
Wangensteen Historical Library, University of Minnesota, USA

**James Luby**  [orcid.org/0000-0002-9667-7379](https://orcid.org/0000-0002-9667-7379)  
Department of Horticultural Science, University of Minnesota, USA

**Kristen L. Mastel**  [orcid.org/0000-0001-8128-3089](https://orcid.org/0000-0001-8128-3089)  
Magrath Library, University of Minnesota, USA

## REFERENCES

- Azeria, ET**, et al. 2006. Temporal dynamics and nestedness of an oceanic island bird fauna. *Global Ecology and Biogeography*, 15(4): 328–338. DOI: <https://doi.org/10.1111/j.1466-822X.2006.00227.x>
- Beans, C.** 2018. Journal entries, maps, and photos help ecologists reconstruct ecosystems of the past. *Proceedings of the National Academy of Sciences of the United States of America*, 115(52): 13138–13141. DOI: <https://doi.org/10.1073/pnas.1819526115>
- Berman, JJ.** 2015. *Repurposing legacy data: Innovative case studies*. Amsterdam; Boston: Elsevier. Available at: <https://www.sciencedirect.com/book/9780128028827/repurposing-legacy-data>.
- Bowser, CJ.** 1986. Historic data sets: Lessons from the past, lessons for the future. In: Michener, WK (ed.), *Research Data Management in the Ecological Sciences*, 155–179. Columbia: University of South Carolina Press.
- Brazdil, R.**, et al. 2016. Damaging hailstorms in South Moravia, Czech Republic, in the seventeenth to twentieth centuries as derived from taxation records. *Theoretical and Applied Climatology*, 123(1–2): 185–198. DOI: <https://doi.org/10.1007/s00704-014-1338-1>
- Brodman, R, Cortwright, S and Resetar, A.** 2002. Historical changes of reptiles and amphibians of northwest Indiana fish and wildlife properties. *The American Midland Naturalist*, 147(1): 135–144. DOI: [https://doi.org/10.1674/0003-0031\(2002\)147\[0135:HCORAA\]2.0.CO;2](https://doi.org/10.1674/0003-0031(2002)147[0135:HCORAA]2.0.CO;2)
- Brönnimann, S**, et al. 2018. A roadmap to climate data rescue services. *Geoscience Data Journal*, 5(1): 28–39. DOI: <https://doi.org/10.1002/gdj3.56>
- Buma, B.** 2018. The hidden value of paper records. *Science*, 360(6389): 613. DOI: <https://doi.org/10.1126/science.aat5382>
- Buma, B**, et al. 2019. 100 yr of primary succession highlights stochasticity and competition driving community establishment and stability. *Ecology*, 100(12): e02885. DOI: <https://doi.org/10.1002/ecy.2885>
- Cardinale, M**, et al. 2015. A centurial development of the North Sea fish megafauna as reflected by the historical Swedish longlining fisheries. *Fish and Fisheries*, 16(3): 522–533. DOI: <https://doi.org/10.1111/faf.12074>

- Chauvel, B**, et al. 2012. History of chemical weeding from 1944 to 2011 in France: Changes and evolution of herbicide molecules. *Crop Protection*, 42: 320–326. DOI: <https://doi.org/10.1016/j.cropro.2012.07.011>
- Chuine, I**, et al. 2004. Grape ripening as a past climate indicator. *Nature*, 432(7015): 289–290. DOI: <https://doi.org/10.1038/432289a>
- Clavero, M**, et al. 2017. Historical citizen science to understand and predict climate-driven trout decline. *Proceedings of the Royal Society B: Biological Sciences*, 284(1846): 20161979. DOI: <https://doi.org/10.1098/rspb.2016.1979>
- Clavero, M** and **Revilla, E**. 2014. Biodiversity data: Mine centuries-old citizen science. *Nature*, 510(7503): 35. DOI: <https://doi.org/10.1038/510035c>
- Craig, JR** and **Reese, SC**. 1973. Retention of raw data: A problem revisited. *American Psychologist*, 28(8): 723–723. DOI: <https://doi.org/10.1037/h0035667>
- Curry, A**. 2011. Rescue of old data offers lesson for particle physicists. *Science*, 331(6018): 694–695. DOI: <https://doi.org/10.1126/science.331.6018.694>
- Curty, RG**, et al. 2017. Attitudes and norms affecting scientists' data reuse. In: Sugimoto, CR (ed.), *ONE*, 12(12): e0189288. DOI: <https://doi.org/10.1371/journal.pone.0189288>
- Delisle, F**, et al. 2003. Reconstructing the spread of invasive plants: Taking into account biases associated with herbarium specimens: Invasive plants and herbarium specimens. *Journal of Biogeography*, 30(7): 1033–1042. DOI: <https://doi.org/10.1046/j.1365-2699.2003.00897.x>
- Diviacco, P**, et al. 2015. Data rescue to extend the value of vintage seismic data: The OGS-SNAP experience. *GeoResJ*, 6: 44–52. DOI: <https://doi.org/10.1016/j.grj.2015.01.006>
- Downs, RR** and **Chen, RS**. 2017. Curation of scientific data at risk of loss: Data rescue and dissemination. In: Johnston, L (ed.), *Curating Research Data: Volume One: Practical Strategies for Your Digital Repository*, 275–294. Chicago, IL: Association of College and Research Libraries.
- Duckworth, S**, **Grayce, M** and **Thornhill, K**. 2018. The trouble with legacy public health data. *Research Data Access & Preservation Summit*, Chicago, IL. Available at: <https://osf.io/ujhn2/> (Accessed: 5 July 2019).
- Easterday, K**, et al. 2018. From the field to the cloud: A review of three approaches to sharing historical data from field stations using principles from data science. *Frontiers in Environmental Science*, 6: 88. DOI: <https://doi.org/10.3389/fenvs.2018.00088>
- Edwards, RJ** and **Contreras-Balderas, S**. 1991. Historical changes in the ichthyofauna of the Lower Rio Grande (Rio Bravo del Norte), Texas and Mexico. *The Southwestern Naturalist*, 36(2): 201. DOI: <https://doi.org/10.2307/3671922>
- Engelhard, GH**, **Righton, DA** and **Pinnegar, JK**. 2014. Climate change and fishing: A century of shifting distribution in North Sea cod. *Global Change Biology*, 20(8): 2473–2483. DOI: <https://doi.org/10.1111/gcb.12513>
- Faniel, IM**, **Frank, RD** and **Yakel, E**. 2019. Context from the data reuser's point of view. *Journal of Documentation*, 75(6): 1274–1297. DOI: <https://doi.org/10.1108/JD-08-2018-0133>
- Farrell, S**, et al. 2019. Resurfacing historical scientific data: A case study involving fruit breeding data. *Journal of eScience Librarianship*, 8(2). DOI: <https://doi.org/10.7191/jeslib.2019.1171>
- Farrell, SL**, et al. 2020. Historical scientific analog data: Life sciences faculty's perspectives on management, reuse and preservation. *Data Science Journal*, 19(1): 51. DOI: <https://doi.org/10.5334/dsj-2020-051>
- Fellers, GM** and **Drost, CA**. 1993. Disappearance of the cascades frog *rana cascadae* at the southern end of its range, California, USA. *Biological Conservation*, 65(2): 177–181. DOI: [https://doi.org/10.1016/0006-3207\(93\)90447-9](https://doi.org/10.1016/0006-3207(93)90447-9)
- Gabelica, R**, **Bojčić, M** and **Puljak, L**. 2022. Many researchers were not compliant with their published data sharing statement: Mixed-methods study. *Journal of Clinical Epidemiology*, in press. Accepted 24 May, 2022. DOI: <https://doi.org/10.1016/j.jclinepi.2022.05.019>
- Gent, ML** and **Morgan, JW**. 2007. Changes in the stand structure (1975–2000) of coastal Banksia Forest in the long absence of fire. *Austral Ecology*, 32(3): 239–244. DOI: <https://doi.org/10.1111/j.1442-9993.2007.01667.x>
- Griffin, ER**. 2015. When are old data new data? *GeoResJ*, 6: 92–97. DOI: <https://doi.org/10.1016/j.grj.2015.02.004>
- Gross, KL** and **Pake, CE**. 1995. *Final report of the ecological society of America committee on the future of long-term ecological data (FLED)*. Washington, DC: Ecological Society of America, p. 122.
- Hédál, R**, **Petřík, P** and **Boublík, K**. 2011. Long-term patterns in soil acidification due to pollution in forests of the eastern Sudetes mountains. *Environmental Pollution*, 159(10): 2586–2593. DOI: <https://doi.org/10.1016/j.envpol.2011.06.014>
- Hof, AR** and **Bright, PW**. 2016. Quantifying the long-term decline of the west European hedgehog in England by subsampling citizen-science datasets. *European Journal of Wildlife Research*, 62(4): 407–413. DOI: <https://doi.org/10.1007/s10344-016-1013-1>

- Huisman, JM** and **Millar, AJK**. 2013. Australian seaweed collections: Use and misuse. *Phycologia*, 52(1): 2–5. DOI: <https://doi.org/10.2216/12-089.1>
- Johnson, RW**. 1964. Retain the original data! comment. *American Psychologist*, 19(5): 350–351. DOI: <https://doi.org/10.1037/h0039238>
- Kelso, C** and **Vogel, C**. 2007. The climate of Namaqualand in the nineteenth century. *Climatic Change*, 83(3): 357–380. DOI: <https://doi.org/10.1007/s10584-007-9264-1>
- Kery, M**, et al. 2006. How biased are estimates of extinction probability in revisitation studies? *Journal of Ecology*, 94(5): 980–986. DOI: <https://doi.org/10.1111/j.1365-2745.2006.01151.x>
- Khan, N**, **Thelwall, M** and **Kousha, K**. 2021. Measuring the impact of biodiversity datasets: Data reuse, citations and altmetrics. *Scientometrics*, 126(4): 3621–3639. DOI: <https://doi.org/10.1007/s11192-021-03890-6>
- Knapp, KR**, **Bates, JJ** and **Barkstrom, B**. 2007. Scientific data stewardship: Lessons learned from a satellite-data rescue effort. *Bulletin of the American Meteorological Society*, 88(9): 1359–1362. DOI: <https://doi.org/10.1175/BAMS-88-9-1359>
- Kullman, L**. 2010. Alpine flora dynamics—A critical review of responses to climate change in the Swedish Scandes since the early 1950s. *Nordic Journal of Botany*, 28(4): 398–408. DOI: <https://doi.org/10.1111/j.1756-1051.2010.00812.x>
- Kwok, R**. 2017. Historical data: Hidden in the past. *Nature*, 549(7672): 419–421. DOI: <https://doi.org/10.1038/nj7672-419>
- Lannoo, MJ**, et al. 1994. An altered amphibian assemblage: Dickinson County, Iowa, 70 years after Frank Blanchard's survey. *American Midland Naturalist*, 131(2): 311. DOI: <https://doi.org/10.2307/2426257>
- Ledneva, A**, et al. 2004. Climate change as reflected in a naturalist's diary, middleborough, massachusetts. *The Wilson Journal of Ornithology*, 116(3): 224–231. DOI: <https://doi.org/10.1676/04-016>
- Lescrauwaet, A-K**. 2013. *Belgian fisheries: Ten decades, seven seas, forty species*. PhD. Ghent University.
- Llasat, M-C**, et al. 2005. Floods in Catalonia (NE Spain) since the 14th century. Climatological and meteorological aspects from historical documentary sources and old instrumental records. *Journal of Hydrology*, 313(1–2): 32–47. DOI: <https://doi.org/10.1016/j.jhydrol.2005.02.004>
- Loehle, C** and **Weatherford, P**. 2017. Detecting population trends with historical data: Contributions of volatility, low detectability, and metapopulation turnover to potential sampling bias. *Ecological Modelling*, 362: 13–18. DOI: <https://doi.org/10.1016/j.ecolmodel.2017.08.021>
- Löffler, F**, et al. 2021. Dataset search in biodiversity research: Do metadata in data repositories reflect scholarly information needs? In: Suleman, H (ed.), *ONE*, 16(3): e0246099. DOI: <https://doi.org/10.1371/journal.pone.0246099>
- Lorenz, DC**, **Boise National Forest** and **United States Forest Service Intermountain Region**. 1993. Snapshot in time: Repeat photography on the Boise National Forest, 1870–1992. US Dept of Agriculture, Boise National Forest, Intermountain Region.
- Magurran, AE**, et al. 2010. Long-term datasets in biodiversity research and monitoring: Assessing change in ecological communities through time. *Trends in Ecology & Evolution*, 25(10): 574–582. DOI: <https://doi.org/10.1016/j.tree.2010.06.016>
- Martin, K**, **Brown, GA** and **Young, JR**. 2004. The historic and current distribution of the Vancouver Island White-tailed Ptarmigan (*Lagopus leucurus saxatilis*). *Journal of Field Ornithology*, 75(3): 239–256. DOI: <https://doi.org/10.1648/0273-8570-75.3.239>
- McClenachan, L**. 2009. Documenting loss of large trophy fish from the Florida Keys with historical photographs. *Conservation Biology*, 23(3): 636–643. DOI: <https://doi.org/10.1111/j.1523-1739.2008.01152.x>
- McClenachan, L**, et al. 2017. Ghost reefs: Nautical charts document large spatial scale of coral reef loss over 240 years. *Science Advances*, 3(9): e1603155. DOI: <https://doi.org/10.1126/sciadv.1603155>
- McClenachan, L**, **Ferretti, F** and **Baum, JK**. 2012. From archives to conservation: Why historical data are needed to set baselines for marine animals and ecosystems. *Conservation Letters*, 5(5): 349–359. DOI: <https://doi.org/10.1111/j.1755-263X.2012.00253.x>
- Michener, WK**, et al. 2009. Biological field stations: Research legacies and sites for serendipity. *BioScience*, 59(4): 300–310. DOI: <https://doi.org/10.1525/bio.2009.59.4.8>
- National Research Council**. 1995a. *Finding the forest in the trees: The challenge of combining diverse environmental data*. DOI: <https://doi.org/10.17226/4896>
- National Research Council (U S) Steering committee for the study on the long-term retention of selected scientific and technical records of the federal government**. (1995b). *Preserving scientific data on our physical universe: a new strategy for archiving the nation's scientific information resources*. Washington, DC: National Academy Press.
- Nelson, G** and **Ellis, S**. 2019. The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Transactions of the Royal Society B*, 374(1763): 20170391. DOI: <https://doi.org/10.1098/rstb.2017.0391>

- Olson, RJ and McCord, RA.** 1998. Data Archival. In: Michener, WK, Porter, JH and Stafford, SG (eds.), *Data and Information Management in the Ecological Sciences: A Resource Guide*, 53–58. LTER Network Office, University of New Mexico.
- Olson, RJ and McCord, RA.** 2000. Archiving ecological data and information. In: Michener, WK and Brunt, JW (eds.), *Ecological Data—Design, Management and Processing*, 117–141. Blackwell Science.
- Patil, C and Siegel, V.** 2009. Shining a light on dark data. *Disease Models & Mechanisms*, 2(11–12): 521–525. DOI: <https://doi.org/10.1242/dmm.004630>
- Patton, B,** et al. 2022. 27 years of livestock production data under different stocking rate levels at the Central Grasslands Research Extension Center near Streeter, North Dakota. *Ag Data Commons*. DOI: <https://doi.org/10.15482/USDA.ADC/1524719> [Last accessed 2022 06 27].
- Pollock, JF,** 2006. Detecting population declines over large areas with presence-absence, time-to-encounter, and count survey methods. *Conservation Biology*, 20(3): 882–892. DOI: <https://doi.org/10.1111/j.1523-1739.2006.00342.x>
- Primack, RB and Miller-Rushing, AJ.** 2012. Uncovering, collecting, and analyzing records to investigate the ecological impacts of climate change: A template from Thoreau’s Concord. *BioScience*, 62(2): 170–181. DOI: <https://doi.org/10.1525/bio.2012.62.2.10>
- Pullin, AS and Salafsky, N.** 2010. Save the whales? Save the rainforest? Save the sata!: Editorial. *Conservation Biology*, 24(4): 915–917. DOI: <https://doi.org/10.1111/j.1523-1739.2010.01537.x>
- Reznick, D, Baxter, RJ and Endler, J.** 1994. Long-term studies of tropical stream fish communities: The use of field notes and museum collections to reconstruct communities of the past. *American Zoologist*, 34(3): 452–462. DOI: <https://doi.org/10.1093/icb/34.3.452>
- Riddell, EA,** et al. 2021. Exposure to climate change drives stability or collapse of desert mammal and bird communities. *Science*, 371(6529): 633–636. DOI: <https://doi.org/10.1126/science.abd4605>
- Rivadeneira, MM, Hunt, G and Roy, K.** 2009. The use of sighting records to infer species extinctions: An evaluation of different methods. *Ecology*, 90(5): 1291–1300. DOI: <https://doi.org/10.1890/08-0316.1>
- Rogers, GF.** 1984. *Bibliography of repeat photography for evaluating landscape change*. Salt Lake City: University of Utah Press.
- Schulte, LA and Mladenoff, DJ.** 2001. The original US public survey land use records: Their use and limitations in reconstructing presettlement vegetation. *Journal of Forestry*, 99(10): 5–10.
- Sharma, S,** et al. 2016. Direct observations of ice seasonality reveal changes in climate over the past 320–570 years. *Scientific Reports*, 6: 25061. DOI: <https://doi.org/10.1038/srep25061>
- Sheffield, C,** et al. 2011. Merging metadata: Building on existing standards to create a field book registry. *LIBREAS. Library Ideas*, 18.
- Singer, RA, Ellis, S and Page, LM.** 2020. Awareness and use of biodiversity collections by fish biologists. *Journal of Fish Biology*, 96(2): 297–306. DOI: <https://doi.org/10.1111/jfb.14167>
- Smith, KL and Jones, ML.** 2007. When are historical data sufficient for making watershed-level stream fish management and conservation decisions? *Environmental Monitoring and Assessment*, 135(1–3): 291–311. DOI: <https://doi.org/10.1007/s10661-007-9650-1>
- Snall, T,** et al. 2011. Evaluating citizen-based presence data for bird monitoring. *Biological Conservation*, 144(2): 804–810. DOI: <https://doi.org/10.1016/j.biocon.2010.11.010>
- Sprague, LA, Oelsner, GP and Argue, DM.** 2017. Challenges with secondary use of multi-source water-quality data in the United States. *Water Research*, 110: 252–261. DOI: <https://doi.org/10.1016/j.watres.2016.12.024>
- Thomer, A,** et al. 2012. From documents to datasets: A mediawiki-based method of annotating and extracting species observations in century-old field notebooks. *ZooKeys*, 209: 235–253. DOI: <https://doi.org/10.3897/zookeys.209.3247>
- Trisurat, Y,** et al. 2020. Systematic forest inventory plots and their contribution to plant distribution and climate change impact studies in Thailand. *Ecological Research*, 35(5): 724–732. DOI: <https://doi.org/10.1111/1440-1703.12105>
- van Bavel, BJP,** et al. 2019. Climate and society in long-term perspective: Opportunities and pitfalls in the use of historical datasets. *Wiley Interdisciplinary Reviews-Climate Change*, 10(6): e611. DOI: <https://doi.org/10.1002/wcc.611>
- Van der Veken, S, Verheyen, K and Hermy, M.** 2004. Plant species loss in an urban area (turnhout, Belgium) from 1880 to 1999 and its environmental determinants. *Flora*, 199(6): 516–523. DOI: <https://doi.org/10.1078/0367-2530-00180>
- Vearncombe, J,** et al. 2017. Data upcycling. *Ore Geology Reviews*, 89: 887–893. DOI: <https://doi.org/10.1016/j.oregeorev.2017.07.009>
- Vearncombe, J, Conner, G and Bright, S.** 2016. Value from legacy data. *Applied Earth Science*, 125(4): 231–246. DOI: <https://doi.org/10.1080/03717453.2016.1190442>
- Vines, TH,** et al. 2014. The availability of research data declines rapidly with article age. *Current Biology*, 24(1): 94–97. DOI: <https://doi.org/10.1016/j.cub.2013.11.014>
- Webb, RH, Boyer, DE and Turner, RM.** 2010. *Repeat photography: Methods and applications in the natural sciences*. Washington, DC: Island Press.

- Wicherts, JM**, et al. 2006. The poor availability of psychological research data for reanalysis. *American Psychologist*, 61(7): 726. DOI: <https://doi.org/10.1037/0003-066X.61.7.726>
- Wiebe, PH** and **Allison, MD**. 2015. Bringing dark data into the light: A case study of the recovery of Northwestern Atlantic zooplankton data collected in the 1970s and 1980s. *GeoResJ*, 6: 195–201. DOI: <https://doi.org/10.1016/j.grj.2015.03.001>
- Wolins, L**. 1962. Responsibility for raw data. *American Psychologist*, 17: 657–658. DOI: <https://doi.org/10.1037/h0038819>
- Yoon, A** and **Kim, Y**. 2017. Social scientists' data reuse behaviors: Exploring the roles of attitudinal beliefs, attitudes, norms, and data repositories. *Library & Information Science Research*, 39(3): 224–233. DOI: <https://doi.org/10.1016/j.lisr.2017.07.008>

Kelly et al.  
*Data Science Journal*  
DOI: 10.5334/dsj-2022-014

11

#### TO CITE THIS ARTICLE:

Kelly, JA, Farrell, SL, Hendrickson, LG, Luby, J and Mastel, KL. 2022. A Critical Literature Review of Historic Scientific Analog Data: Uses, Successes, and Challenges. *Data Science Journal*, 21: 14, pp. 1–11. DOI: <https://doi.org/10.5334/dsj-2022-014>

**Submitted:** 31 March 2022

**Accepted:** 05 July 2022

**Published:** 28 July 2022

#### COPYRIGHT:

© 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

*Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.