Journal of eScience Librarianship
putting the pieces together: theory and practice

Full-Length Paper

Resurfacing Historical Scientific Data: A Case Study Involving Fruit Breeding Data

Shannon L. Farrell, Lois G. Hendrickson, Kristen L. Mastel, Katherine Adina Allen, and Julia A. Kelly

University of Minnesota - Twin Cities, Minneapolis, MN, USA

Abstract

Objective: The objective of this paper is to illustrate the importance and complexities of working with historical analog data that exists on university campuses. Using a case study of fruit breeding data, we highlight issues and opportunities for librarians to help preserve and increase access to potentially valuable data sets.

Methods: We worked in conjunction with researchers to inventory, describe, and increase access to a large, 100-year-old data set of analog fruit breeding data. This involved creating a spreadsheet to capture metadata about each data set, identifying data sets at risk for loss, and digitizing select items for deposit in our institutional repository.

Results/Discussion: We illustrate that large amounts of data exist within biological and agricultural sciences departments and labs, and how past practices of data collection, record keeping, storage, and management have hindered data reuse. We demonstrate that librarians have a role in collaborating with researchers and providing direction in how to preserve analog data and make it available for reuse. This work may provide guidance for other science librarians pursuing similar projects.

Conclusions: This case study demonstrates how science librarians can build or strengthen their role in managing and providing access to analog data by combining their data management skills with researchers' needs to recover and reuse data.

Introduction

While analog data can be found across universities, research centers, museums, and libraries, harvesting the data has been piecemeal. Past practices of data collection, record keeping, storage, and management have hindered data reuse. For example, individual investigators who are lacking historical data to calibrate models have resorted to extracting data points by estimating from graphs in older reports or journal articles or partially reconstructing data from incomplete data sets. In these cases, analog data together with modeling and preserved physical samples (e.g., ice cores) can give scientists a more complete picture. However, little attention has been spent on managing historical analog data and facilitating access to it. For the purpose of this paper, we define analog data as non-digital data, primarily in print, such as field books, lab notebooks, ledgers, data sheets, photographs, maps, drawings/sketches, slides, and so on. Over the last several years, mandates by government agencies and private funders that require grantees to make their data publicly available have driven discussions about research data management and reuse. In response to the new requirements, librarians have developed support and educational services for faculty and researchers to create data management plans, offer guidance around long-term data sharing and preservation, and make recommendations about data repositories (institutional or discipline-specific) and open access publishing.

The concepts of discoverability, preservation, reproducibility, and reuse that are being discussed around machine-readable data parallel similar issues with analog data. As space becomes a premium at institutions, and faculty and researchers retire, opportunities for librarians to step into the space to discuss the future of analog data are abundant. University archives are the place of record for many things related to the history of institutions, however, they may not have collection policies that address data. The proliferation of paper-based unprocessed archival collections have resulted in processing practices that restrict how raw data are acquired, described, and managed, with few exceptions in cases where the university is collecting the entire corpus of a person's record of work. When scientific data are found in archives they may lack initial context and description as well as archival descriptors to make them easily identifiable for potential researchers, as neither the description nor the finding aids were created with the idea that future researchers would be consulting this archival collection for use in longitudinal studies or other scientific research. In addition, many scientists may not be trained in how to use or search archival materials, as archives are outside their frequently consulted resources, such as databases, repositories, and association publications.

Since analog data are frequently hidden in labs, archives, and other locations, their use in current research studies remains relatively low. However, this may be more a lack of awareness and discoverability, than a lack of the data's importance or reuse potential. As seen with the data management plan mandates, scientists in the past may not have been prioritizing long-term access to their data as part of the research lifecycle. In informal conversations with the authors over the past several years, researchers have revealed uncertainty as to what will happen with their analog data long term, from possibly transferring materials to their successors, hoping the library or university archives may take it, or, worst-case, having it thrown away. Depending on the content and volume of raw analog data it may not be possible or make sense for libraries to house it all. However, librarians can be an active participant with researchers in conversations about how to manage this data, and raise awareness or sound the alarm before potentially significant and usable analog data ends up in the recycling

dumpster, or is unusable due to poor record keeping. Increasingly, librarians have the data management expertise and skill sets to engage with researchers and help them better describe, organize, and preserve their analog data for future scholars and scientists to access for longitudinal and historical studies. In this article, we will describe how librarians at the University of Minnesota worked with horticultural researchers to inventory and prepare their 100-year-old analog fruit breeding data to enhance discoverability and access.

Literature Review

As of this writing there are very few large multi-institutional efforts to address the topic of older analog data in the sciences. This stands in contrast to the work being done around biological specimens, including the international iDigBio project which makes many millions of specimen records available (IDigBio 2019). Current analog data projects mainly focus on scanning handwritten field notebooks and making them freely available. Notable are the efforts by the Smithsonian and the Biodiversity Heritage Library (BHL), both of which focus on natural history topics (IDigBio 2019; Biodiversity Heritage Library 2019). The groups working together on the BHL project include museums, libraries, archives, and botanical gardens. They acknowledge that researchers move from institution to institution so in working together they can provide a more complete picture of both an individual's work over time or various aspects of multi-institutional projects.

Field notebooks are considered a rich source of information of various kinds, including both numeric and descriptive data as well as cultural and historical references. Jones notes their importance to many fields as well as the fact that they may be found in archives, museums, or libraries and they may be dealt with differently in various contexts (Jones 2017). In his opinion, they may need more description than other materials, given the breadth of information that they often contain. He also highlights the fact that they may be personal as well as professional, serving as a diary or travelogue in some cases.

Field notebooks are only one example of older analog materials that are of current interest to libraries and researchers. There are small-scale efforts to unearth, expose, and/or reuse analog data going on at individual libraries, archives, and museums, including one at Texas A&M which focuses on field notebooks, as well as specimen catalogs from one researcher (Davis 2019). Other work includes a project at Oregon State University where an archivist and librarian are digitizing historical public health records (Duckworth, Grayce, and Thornhill 2018). At Tulane, another archivist/librarian team have used the archives to locate weather and environmental data in a variety of formats, including crop ledgers, diaries, and personal papers in support of scientific research projects (Kearney and Mullins 2019).

Projects employing citizen science include a University of Michigan effort to locate data in hundreds of student papers associated with their biological field station; another project in Portugal to search old correspondence to find information about species and locations; and a third at the Chicago Botanic Garden's Lenhardt Library focused on the scientific notebooks of a German orchidologist and French lily specialist (Schell 2019; Dias da Silva et al. 2019; Lettner 2019).

Other projects have gone beyond digitizing. A librarian at Stanford is attempting to convert California Cooperative Oceanic Fisheries Investigations (CalCOFI) analog data to a

machine-readable format (Whitmire 2016). Field notebooks from the University of Colorado Museum of Natural History were digitized, transcribed, and annotated and the resulting files were cross-walked into Darwin Core-compliant record sets which revealed over 1,000 observations of species (Thomer et al. 2012).

Although government agency and private funder mandates have brought data management and data sharing topics to the forefront of researchers' minds, that does not mean that data reuse is a new phenomenon. There are cases of scientists incorporating existing data into their current studies and these range from simply referencing weather or geologic data to replicating entire studies. This is true for both machine-readable digital data and analog data. Examples of reuse of analog data include Abdesselam et al. who studied groundwater contamination in Algeria over the last 40 years, Munro and Horst who studied forest cover histories in Sierra Leone, and Buma et al. who used photographs to identify locations in order to document plant community succession in Alaska (Abdesselam et al. 2013; Munro and Horst 2016; Buma et al. 2017).

In addition to actually reusing data, scientists and librarians are highlighting historical data resources, creating new resources by utilizing older data, and writing about the importance of its reuse. Librarians Evans and Welch highlight collections of economic data that have been made freely available by libraries, universities, non-government organizations (NGOs), and governments but may be hard to locate (Evans and Welch 2014). Their focus is on pre-World War II data. Beltrano and colleagues draw attention to an archive and library for agricultural meteorology and phenology data that could be of interest to agricultural and climate scientists (Beltrano et al. 2012). Nicholson has created a database of precipitation data for Africa in the nineteenth century and writes about the resources that were tapped and the possible uses for the data (Nicholson 2001). Geological Survey of Canada researchers rescued historical bedrock field observations from 1968-1970 and they have already been reused in recent studies (Fallas, MacNaughton, and Sommers 2015). The writings of Thoreau and other historical records were consulted by researchers Primack and Miller-Rushing as they sought to compare earlier conditions to current ones in a particular locality (Primack and Miller-Rushing 2012). They note what they see as the underutilization of historical documents in current work.

In their paper focusing on data in ecology and evolutionary biology, Poisot, Mounce, and Gravel make the point that "improving our data-sharing practices will improve both the quality of the science, and the reputation of the scientists" (Poisot, Mounce, and Gravel 2013). The scientific research value of the United States Geological Survey fossil collections and its associated records collection was the focus of a conference paper, with presenters maintaining that archive material should be inventoried and described in a standardized way to enhance its scientific research value (McClees-Funian et al. 2017). Researcher attitudes and actions concerning archaeological and zoological research data collection and management were investigated and common concerns emerged, including internal (e.g., disciplinary norms) and external (e.g., funding, legal) factors, as well as costs associated with data management. Analog data preservation was a higher priority to these study participants because of the greater likelihood of its survival and skepticism about the long-term viability of digital data (Frank, Yakel, and Faniel 2015). The authors of *Curating the Analog, Curating the Digital* (Archival Journal's June 2013 special issue) examine data curation across a range of domains, roles, perspectives, and practices of data use and reuse, calling for closer collaboration between archivists, data curators, and researchers to provide support for scientific research

through enabling long-term reuse of data (Hswe and O'Meara 2013). We argue that librarians and archivists can be significant partners in collaborating with projects, as described above, to facilitate the preservation and reuse of research data and we demonstrate the feasibility of this work in the following case study.

Case Study

We are collaborating on a project with researchers at the Horticultural Research Center (HRC) to help preserve and facilitate reuse of analog data. The HRC is part of the University of Minnesota but is located approximately 30 miles away from the main campus on 80 acres near Victoria, Minnesota. It was first established in 1908 by the University as the Fruit Breeding Farm. Although the HRC has conducted research on restoration ecology and the cold-hardiness of various plants, most of their work has focused on fruit breeding. The HRC has made nearly 100 fruit introductions over time, including varieties of apricots, blueberries, cherries, grapes, raspberries, and strawberries. However, the HRC is most widely known for their apple breeding program. Some of the most well-known apples that the HRC has released include: Haralson, Honeycrisp, SnowSweet®, Zestar!®, SweeTango®, Sweet Sixteen and First Kiss®/Rave®. Many of the apple varieties they have released have been commercially successful, with the Honeycrisp being the most dramatic example. The Honeycrisp had generated over six million dollars for the University by 2007 when it lost its domestic patent protection (Olson 2007).

Our team of science librarians, with experience in scientific research and expertise in data management and archival curation, began working with the HRC in 2016 when the director of the fruit breeding program mentioned having a lot of poorly organized analog data. The director of the center contacted the horticultural liaison librarians, who then enlisted the help of the co-chair of our library's research data services team. Together, we offered to take a look and provide some suggestions. The data in question fills a small, approximately 5x7 ft. closet-sized room with floor-to-ceiling shelves. It encompasses a variety of data, including breeding data, accession data, scion wood (grafting wood) data, and experimental data, along with historical maps from the farm. The vast majority of the data are original, and some of it, particularly that related to fruit introductions, is proprietary. Some of the data was organized (by topic or type), but it was not labeled. The researchers were not able to easily identify the content of each notebook or binder, for example. Over the course of three days, we worked in conjunction with the researchers to create a detailed inventory, including a data dictionary and controlled vocabulary. We described what date range the data covered, what genus/species were included, and whether the data were proprietary or not. The researchers identified terms/keywords that they would want to use to search the data, such as "field observations," "field maps," or "breeding material." We also included information about the data's physical location, format (what it looked like), and how many volumes existed with the same name in order to find it on the shelf (see Figure 1). In the end, records were created for 247 physical objects (notebooks, ledgers, folders) in a spreadsheet.

During the inventory process, we also asked the researchers to identify what items they would prioritize for digitization based on what was most at a risk for loss, what they considered most important, and what could be made available publicly. Following the completion of the inventory, based on researcher priorities, we applied for and received a small, internal grant from our library's Strategic Digitization program in 2017 that enabled us to digitize 28 of the

Title	Physical_location	Date_range	Volume_numbers	Format	Proprietary	Genus	Common_name	Multi_species	C_field_obs
Irradiated apple	1.1	1968	4	2-ring binder	0	Malus	Apple	0	1
Apple test 1	2.2	1961-1965	1	2-ring binder	0	Malus	Apple	0	1
Apple test 2	2.2	1961-1965	1	2-ring binder	0	Malus	Apple	0	1
2nd test apple	2.2	1946-1957	1	2-ring binder	0	Malus	Apple	0	1
Blueberry accession	2.3	1959-1964	1	Small folder, brass fastener	0	Vaccinium	Blueberry	0	1
European plums	2.3	1947-1952	1	2-ring binder	0	Prunus; P. persica	European Plum; Cherry Plum; American Plum; Nectarine	1	1
Grapes dead file	2.3	1942-1963	1	2-ring binder	0	Vitis	Grape	0	1
Peach dead file	2.3	1942-1955	1	Small folder, brass fastener	0	Prunus	Peach	0	1
Apple seedling notes	3.2	1929-1930	1	Unbound, brass fastener	0	Malus	Apple	0	1
Strawberry trials block 7	4.4	1958-1960	1	Small folder, brass fastener	0	Fragaria	Strawberry	0	1
Ornamental propagation	5.4	1975-1976	1	Unbound, binder clip	0	Many	Many	1	0
Storage apples	6.1	1967-1978	1	Bound ledger	0	Malus	Apple	0	0
Gooseberry	6.3	1930-1932	1	Bound ledger	0	Ribes	Gooseberry	0	1
Strawberry book	6.4	1928-1965	35	Bound ledger	0	Fragaria	Strawberry	0	1

Figure 1: Selected examples from fruit breeding inventory. Numbers ranging from 1.1 to 6.4 in the "Physical location" field refer to actual shelf numbers in the Horticultural Research Center. Numbers in the "Proprietary," "Multi_species," and "C_field_obs" fields refer to the presence or absence of the characteristic, with "0" meaning absent and "1" meaning present.

247 bound ledgers and binders. We limited it to this amount based on grant stipulations and saw this as a pilot to test efficacy. Our library's Digital Library Services used an overhead digital camera to scan the disbound and fragile materials, and the raw files were converted into pdfs. We then created a spreadsheet to capture metadata, using some of the data from the original inventory, and worked with the researchers to make sure that the data could be understood by others. The vast majority of the original data are handwritten and have not been made machine readable through its digitization. Although our institution has a data repository, it only accepts machine readable data. Our institutional repository, the University Digital Conservancy (UDC), which is committed to providing long-term preservation and access, does not function as a registry and requires full-text documents. Therefore, we created a collection in the UDC to store and preserve the data that we were able to digitize and thus enable its reuse (https://conservancy.umn.edu/handle/11299/201897). The HRC research collection has seven sub-collections organized by either type of research (e.g., pollination records, phenotype data) or documentation (e.g., maps, accession records). The pdf documents are paired with .txt readme files, non-proprietary open standards for long-term access. We could not find any models for creating metadata for historical analog horticultural data sets. Therefore, using a template (https://www.lib.umn.edu/datamanagement/dmp) created at the University of Minnesota for digital data sets, the metadata we created include title of the data set, author information, dates and geographic location of data collection, funding sources, creative commons attribution license, common and scientific names of species, and description of methods. We considered the pilot to be successful and hope to make the majority of the HRC's non-proprietary data publicly available in this manner as time and/or funding is available.

One of the main incentives for wanting to organize and make the data available is that researchers have been actively using the older data internally and have had outside requests

to use their data. Prior to our involvement, a Ph.D. student in Horticulture, Nicholas Howard, whose research focused on tracing the parentage of the Honeycrisp apple, attempted to find information in the historical data. The current apple breeders did not know the parentage because they could not find the breeding records from that time period. They presume that the data was lost at some point in the 1970s when researchers were attempting to digitize the data, because it should have been with all of the other analog records. Due to this loss of data, figuring out the parentage became a difficult and expensive process, involving digging through the old breeding data, figuring out where the HRC sent scions, calling nurseries who may have had old trees, and conducting genetic testing (Martin 2017; Howard et al. 2017; Howard, 2017). The project illustrated the importance of conducting proper data management with their old paper data.

The old data has also been used by people from outside the HRC. Daniel Bussey consulted the data while writing his seven-volume set about the history of apples (Bussey 2016). During this process he also notified us about a rare handwritten and illustrated monograph from 1897 which he discovered in the University of Minnesota's library collection (Green 1897). The monograph consists of listings and descriptions of apple varieties accompanied by original drawings. It is the only copy in existence and represents data that cannot be found elsewhere. Along with the HRC data, we had the handwritten apple monograph digitized using the library's internal digitization grant. Several factors, including the book's provenance, scarcity, and importance to Minnesota's research heritage, informed the decision to put forward a proposal to the University of Minnesota Libraries Publishing Services to make it more visible and accessible. The proposal was accepted, put into the work queue, and several months later the final version was made available online (Green 2019).

Discussion

In this paper, we discuss a collaborative effort by HRC researchers and University of Minnesota librarians to preserve and facilitate reuse of analog data. Using the case study of fruit breeding records, we have illustrated the large amount of data that can exist within biological and agricultural sciences departments and labs, and how past practices of data collection, record keeping, storage, and management have hindered data reuse. We also demonstrate that librarians have a role in collaborating with researchers that could serve as a guide for science librarians to follow when working with researchers to preserve analog data and make it available for reuse. Our roles included working with researchers to find materials, creating inventories and metadata, and collaborating with library specialists responsible for digitization and preservation, publishing and repository management.

While the HRC researchers recognized the importance of preserving analog fruit breeding research data, there were numerous obstacles that prevented them from undertaking the task. These included the scope of materials gathered over the long life (100 years+) of the research, diverse taxonomies related to the data (e.g., lack of consistent or controlled vocabulary), varying degrees of data completeness (e.g., missing records from the 1970s), and the lack of data parameters required in a typical data set write-up (e.g., metadata or context that was poorly described or lacking). The HRC researchers were interested in recovering the 'parent' of valuable strains, locating data that laid out breeding history, and making potentially valuable data available for reuse by other scientists. They were motivated to collaborate on the project, and the timing was opportune, as a Ph.D. student was tracing the genetic strain of the HRC's

most profitable apple, Honeycrisp. Furthermore, they understood the advantage of working with science librarians who are experts in data management and data sharing education.

We found that preservation was an issue for most of the HRC analog records which were housed in a non-controlled environment. The vast majority of the data was original, some was organized but unlabeled, and some was still proprietary. In addition, discoverability and access were limited, and the data sets' existence was known only to a small number of people. Without richer descriptive metadata and platforms to make that metadata visible, the chance for reuse was low. To discern how to best preserve and increase accessibility, we had to examine the data's characteristics, including type of data, scope of data, copyright and intellectual property concerns, existing metadata or supporting documentation, and revision or addition to data (including ongoing research). We also considered various methods to preserve and/or increase access to the data: solely inventorying and organizing the analog data, digitization of materials in situ, applying optical character recognition (OCR) or other reformatting (e.g., into a spreadsheet) and transcription/translation/data input. Our goal was to find a feasible way to apply additional metadata and description to these historic analog data sets. Some of the challenges we encountered were various types of research being mixed together (e.g., multiple species in one ledger or multiple types of experiments in one ledger); lack of consistency or controlled vocabulary, particularly with naming conventions; missing information (e.g., who collected the data); and unexplained variables (e.g., unclear naming practices and acronyms, some columns/rows unlabeled). There were also issues with readability of handwritten data, and data that were crossed out or written over.

Improving the metadata may be the only way to make analog data accessible, usable, and discoverable. Creating data inventories and identifying a preservation plan is complicated and time-intensive. Many research data sets are orphaned, and the chance to describe materials to facilitate reuse in future research is lost when the creator dies or the project is abandoned. As this case study demonstrates, these pre- and post-custodial issues can be identified and managed, ideally by collaboration between researchers and librarians or archivists, in order to facilitate the organization of materials such that they can be easily understood by current researchers and for reuse.

This study is not meant to be a one-size-fits-all approach to handling analog data sets. As we learned in undertaking this project, researchers may not want to archive their analog data, as they still may be actively using and adding to it; archives and repositories have limits and guidelines on what they will accept; and research data may be proprietary and not available for public consumption. In some cases, the data can be preserved in a digital format. However, the consensus among librarians, archivists, and researchers is that not everything can or should be digitized. And furthermore, digitization does not necessarily facilitate reuse or accessibility of the data. Although digitizing images of lab notebooks means that someone can, in theory, access it online, it does not mean that it will be machine readable and we need to carefully consider the utilization of resources to digitize data. The literature shows, however, an increased awareness about the potential value of analog research data, along with the need to analyze vast amounts of heterogeneous data from multiple sources for decision-making and longitudinal studies on key challenges such as the impacts of climate change, invasive species, and food security. Analog data are not being generated at the same rate as in the past. Most scientific researchers now collect data digitally, and the few who collect analog data promptly convert it to machine readable formats for analysis and preservation. So, what do we

do with older, non-digital data? How do we help our researchers and institutions preserve analog data that may be of value for future research?

This case study demonstrates that there is a desire to better describe and preserve historical, analog data, regardless of whether it is digitized or made machine-readable. It also provides lessons learned, and offers suggestions for how to improve data management and access to analog data. Helping researchers both safeguard the analog data that they have and also make its existence more widely known would be a service to science. We do not know how this older analog data might be utilized in the future, just as we do not know how the current data being captured in machine-readable format in data repositories will be used by others. As librarians, we may be in a position to more easily take a system-wide view of the issues with analog data than individual scientists who are more focused on their particular area of study. The case study of the HRC analog data shows that as we work with more researchers and data sets, it is likely that we will have a better understanding of the specific challenges researchers encounter, and opportunities for libraries to help.

Conclusions

This case study of fruit breeding records raises questions about analog data: where is it held, what format is it in, how can it be discovered, how is it currently described, and what can be done to either leverage or enhance its current description, discoverability, and accessibility. It points out that analog data exists disparately across and within institutions, and that science librarians are uniquely situated to develop frameworks for data reuse, so that the evidence held in these records of enduring value might be used to contribute to the creation of new knowledge and enhance understanding of present conditions.

Collaborative interdisciplinary research, particularly around critical issues like climate change, increasingly draws on historical data to gain a complete picture and to open up new areas of scientific inquiry. Librarians can work with diverse disciplines to identify and document the veracity and authenticity of historical research documents and analog data to help scientists discover, access, use, contextualize, understand, and scrape this data. This case study examines some of the practical issues involved in making data available for reuse—in reputable repositories, with adequate metadata, in usable formats, and the economic constraints. It demonstrates how science librarians can build or strengthen their role in providing access to this documentary evidence by combining their data management skills with researchers' needs to recover and reuse data.

There is much work to be done on this topic. Additional research is needed to understand the scope of the problem on university campuses and beyond. Further work is needed to understand how scientists are using and citing older, historical data, and what value they would assign to analog data. It is also unclear how well older scientific data are described and how much work would be needed to enhance its metadata for reuse. In addition, there is potential work to be done on understanding naming taxonomies in records held within special collections, as scientists' and librarians' vocabularies may overlap, but are frequently inconsistent. We are continuing this work at the University of Minnesota by surveying and interviewing researchers to learn how they are currently using, storing, and planning for the future of their analog data. We also will discuss concerns they have about analog data, potential services they would want, and their experiences and thoughts around analog data

reuse. If the scientific community sees the value in this analog data, they too need to be thinking about and preparing their data for long-term use. Librarians and archivists are positioned to help them in these endeavors by making use of our digital data management expertise.

Acknowledgements

Disclosures

The substance of this article is based upon a panel presentation on 'Data Services' at RDAP Summit 2019: "Value in those dusty file cabinets? Identifying analog data from our campus community" available at https://osf.io/3myvt.

References

Abdesselam, Salah, Amor Halitim, Anthony Jan, Fabienne Trolard, and Guilhem Bourrie. 2013. "Anthropogenic Contamination of Groundwater with Nitrate in Arid Region: Case Study of Southern Hodna (Algeria)." *Environmental Earth Sciences* 70(5): 2129-2141. https://doi.org/10.1007/s12665-012-1834-5

Beltrano, Maria Carmen, Giovanni Dal Monte, Stanislao Esposito, and Luigi Iafrate. 2012. "The CRA-CMA Archive and Library for Agricultural Meteorology and Phenology: A Heritage to Know, Preserve and Share." *Italian Journal of Agrometeorology-Rivista Italiana Di Agrometeorologia* 17(3): 15-24. http://agrometeorologia.it/documenti/Rivista2012_3/AIAM_2012_03_015BELTRANO_DALMONTE_ESPOSITO_IAFRATE.pdf

Biodiversity Heritage Library. 2019. "BHL Field Notes Project." Accessed July 5, 2019. https://www.biodiversitylibrary.org/collection/FieldNotesProject

Buma, Brian, Sarah Bisbing, John Krapek, and Glenn Wright. 2017. "A Foundation of Ecology Rediscovered: 100 Years of Succession on the William S. Cooper Plots in Glacier Bay, Alaska." *Ecology* 98(6): 1513-1523. https://doi.org/10.1002/ecy.1848

Bussey, Daniel J. 2016. *The Illustrated History of Apples in the United States and Canada.* Mount Horeb, WI: JAK KAW Press, LLC.

Davis, William B. 2019. "William B. Davis Collection." OAKTrust. Accessed July 5, 2019. http://oaktrust.library.tamu.edu/handle/1969.1/129120

Dias da Silva, Ana Margarida, M. Teresa Girão da Cruz, Joana Cabral-Oliveira, Helena Freitas, and Antonio C. Gouveia. 2019. "Plant Letters: A Citizen Science Project Uncovering Historical Biodiversity Data." *Biodiversity Information Science and Standards* 3: e36153. https://doi.org/10.3897/biss.3.36153

Duckworth, Steve, Mack Grayce, and Kate Thornhill. 2018. "The Trouble with Legacy Public Health Data." Presented at the *Research Data Access & Preservation Summit 2017*, Chicago, IL. https://osf.io/ujhn2

Evans, Kelly J., and Jeanie M. Welch. 2014. "Measuring the Past: Free Digitized Sources of Historical International Economic Information." *Journal of Business & Finance Librarianship* 19(4): 287-305. https://doi.org/10.1080/08963568.2014.946363

Fallas, Karen M., Robert B. MacNaughton, and Matthew J. Sommers. 2015. "Maximizing the Value of Historical Bedrock Field Observations: An Example from Northwest Canada." *GeoResJ* 6(June): 30-43. https://doi.org/10.1016/j.grj.2015.01.004

Farrell, Shannon L., Kristen L. Mastel, and Julia A. Kelly. 2019. "Value in Those Dusty File Cabinets? Identifying Analog Data from Our Campus Community." *Open Science Framework*. https://osf.io/3myvt

Frank, Rebecca D., Elizabeth Yakel, and Ixchel M. Faniel. 2015. "Destruction/Reconstruction: Preservation of Archaeological and Zoological Research Data." *Archival Science* 15(2): 141-167. https://doi.org/10.1007/s10502-014-9238-9

Green, Samuel B. 1897. *Apples*. Place of publication not identified: publisher not identified. Retrieved from the University of Minnesota Digital Conservancy. http://hdl.handle.net/11299/208485

Green, Samuel B. 2019. *Apples*. Minneapolis: University of Minnesota Libraries Publishing. https://doi.org/10.24926/8666.0201

Howard, Nicholas P., Eric van de Weg, David S. Bedford, Cameron P. Peace, Stijn Vanderzande, Matthew D. Clark, Soon Li Teh, Lichun Cai, and James J. Luby. 2017. "Elucidation of the 'Honeycrisp' Pedigree through Haplotype Analysis with a Multi-Family Integrated SNP Linkage Map and a Large Apple (Malus×domestica) Pedigree-Connected SNP Data Set." *Horticulture Research* 4(17003). https://doi.org/10.1038/hortres.2017.3

Howard, Nicholas. 2017. *Elucidation of the extended pedigree of 'Honeycrisp' apple and genetic architecture of its susceptibility to soft scald and soggy breakdown postharvest fruit disorders and zonal leaf chlorosis disorder*. Retrieved from the University of Minnesota Digital Conservancy http://hdl.handle.net/11299/200171

Hswe, Patricia and Erin O'Meara, eds. 2013. "Curating the Analog, Curating the Digital." *Archive Journal*. http://dev.archivejournal.net/?p=4691

IDigBio. 2019. "IDigBio Home." Accessed July 5, 2019. https://www.idigbio.org/home

Jones, Michael. 2017. "From Personal to Public: Field Books, Museums, and the Opening of the Archives." *Archives and Records* 38(2): 212-227. https://doi.org/10.1080/23257962.2016.1269645

Kearney, Courtney and Andrew Mullins III. 2019. "How Can Historical Archives Support STEM Research?" https://osf.io/2u6kv

Lettner, Amanda. 2019. "Citizen Science Literature Project: Transcribing Scientific Notebooks." Chicago Botanical Garden. Accessed July 5, 2019. http://cbhl.libguides.com/scientificnotebooks/transcribe

Martin, Echo. 2017. "Nicholas Howard Uncovers the Honeycrisp Family Tree." Department of Horticultural Science. Published May 24. https://horticulture.umn.edu/about/horticulture-news/nicholas-howard-honeycrisp-family-tree

McClees-Funian, Ricardo, Casey McKinnery, Natalie Latysh, Kathy Hollis, Holly Little, and Rebecca Schmidt. 2017. "Getting the Records Straight: Digitizing and Making Connections among USGS Fossil Collections and Associated Records for Broader Access and Discovery." Presented at *The Society For The Preservation of Natural History Collections Program 2017*, Denver, Colorado. https://prd-wret.s3-us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/atoms/files/Digitizing_paleontology_collections.pdf

Munro, Paul George, and Greg van der Horst. 2016. "Contesting African Landscapes: A Critical Reappraisal of Sierra Leone's Competing Forest Cover Histories." *Environment and Planning D: Society and Space* 34(4): 706-724. https://doi.org/10.1177/0263775815622210

Nicholson, S. E. 2001. "A Semi-Quantitative, Regional Precipitation Data Set for Studying African Climates of the Nineteenth Century, Part I. Overview of the Data Set." *Climatic Change* 50(3): 317-353. https://doi.org/10.1023/A:1010674724320

Olson, Dan. 2007. "Honeycrisp Apple Losing Its Patent Protection, but Not Its Appeal." MPRNews. Published October 21. https://www.mprnews.org/story/2007/10/11/honeycrisp

Poisot, Timothée E., Ross Mounce, and Dominique Gravel. 2013. "Moving toward a Sustainable Ecological Science: Don't Let Data Go to Waste!" *Ideas in Ecology and Evolution* 6(2): 11-19. http://doi.org/10.4033/iee.2013.6b.14.f

Primack, Richard B., and Abraham J. Miller-Rushing. 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. https://doi.org/10.1525/bio.2012.62.2.10

Schell, Justin. 2019. "Unearthing Michigan Ecological Data." Accessed June 24, 2019. https://www.zooniverse.org/projects/jmschell/unearthing-michigan-ecological-data/about/research

Thomer, Andrea, Gaurav Vaidya, Robert Guralnick, David Bloom, and Laura Russell. 2012. "From Documents to Datasets: A MediaWiki-Based Method of Annotating and Extracting Species Observations in Century-Old Field Notebooks." *ZooKeys* 209: 235-253. https://doi.org/10.3897/zookeys.209.3247

Whitmire, Amanda. 2016. "Creating a 23-Year Oceanographic Time-Series: Or How I Learned to Stop Worrying & Love Data." Poster Presented at *The Rescue of Data At Risk Workshop 2016*, Boulder, CO. https://stacks.stanford.edu/file/druid:hp802gp8515/Whitmire_CalCOFI-Poster_pg1.pdf