Field Notebooks and Tally Sheets: Finding and Describing Reusable Analog Data on Campus

Shannon L. Farrell, Julia A. Kelly, and Kristen L. Mastel *

Introduction
As research is accelerating at an ever-increasing pace, so is the data that it creates. Recent mandates by funding agencies and governments across the globe to share data created from sponsored projects mean that much of the data being generated today has the potential to be available for future researchers to utilize. Libraries are key players in managing data from assisting in creating documentation so others can understand it, to helping preserve it, and maintaining access to it over time. The output of current research is often born digital, but the same issues also apply to data that exists in analog format. By identifying and preserving existing data in analog format on college and university campuses, we can reclaim it for current and future generations and new knowledge can be generated.

Even though funding agencies did not mandate the retention and preservation of data until quite recently, in the past some researchers took it upon themselves to keep their own data. In some cases they have even hung onto data from their predecessors. Although it is currently not typically part of the collection policies of college and university archives, analog data has also made its way into archives as well.

We have begun an investigation of the extent of analog data at the University of Minnesota. The University of Minnesota is a land-grant institution serving over 50,000 students and faculty. Our project has been multi-pronged, with a focus on agricultural and environmental analog data, and we are looking into holdings around campus and individual’s strategies for using, managing, and preserving this material.

Researchers on our campus who have analog data have expressed concerns about its security and have sought the help of library staff in addressing those issues. The researchers are often aware of the potential problems but are not familiar with the campus resources that could support them in solving their issues.

Without intervening and aiding researchers in preserving their analog data, it is quite likely that it will be hidden from future researchers or lost. In the past, scientists have had few constructive options for dealing with the analog data that they produced. Many people stored it in their labs and often left it there when they retired. Some kept it at home. Some discarded it. A few gave it to the archives at their institutions and it may or may not have been retained. A number of authors in the past have generously offered to share their data, noting in their journal articles that “data is available upon request”, but by the time the requests come in, the data may not be available, with the likelihood decreasing over time elapsed.¹ It may have been discarded, in a format that cannot be accessed, or be in a condition where it would be impossible for it to be used without a great deal of work on

* Shannon L. Farrell is Natural Resources Librarian at University of Minnesota, email: sfarrell@umn.edu; Julia A. Kelly is Science Librarian at University of Minnesota, email: jkelly@umn.edu; Kristen L. Mastel is Outreach and Instruction Librarian at University of Minnesota, email: meye0539@umn.edu
the part of both the original author and the requester due to issues such as poor description, lack of metadata, and technological obsolescence.

Finding ways to help researchers both safeguard the analog data that they have and also make its existence more widely known would be a service to science. Data that would otherwise be lost can be preserved. 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. It could spur new studies that compare the older data to newly-collected data. This could be especially useful to those studying climate change and other areas where longitudinal studies are of great value. Raising awareness among scientists about older analog data could spur more studies that utilize that data.

University libraries often have an archival unit and these units historically have not been focused on raw data. Many university archives, where space may be at a premium, do not currently include data in their collection policies and only take it in cases where the researcher is noteworthy and the entire corpus of the person’s work is included. This has not always been true, so some older analog data does exist in university archives, but it can be both difficult to locate and not always usable once it is found. Probably neither the person’s description nor the finding aids were created with future scientific data uses in mind. Scientists are not generally trained in how to use archival materials, and may not see archives as either a place to store their research output or as a potential source for data for their research.

The effort to identify, describe, and preserve older analog data supports the data management movement that librarians have undertaken in recent years, and the skills that data librarians have are easily transferable to the analog data effort. As librarians, we may be in a position to more easily take a system-wide view of the issues of analog data than individual scientists who are more focused on their particular area of study.

**Literature Review**

We suspect that analog data exists across college and university campuses, but there does not appear to be a much investigation around the topic of analog data by either data producers or data stewards. There are a few projects that address the preservation of older data, notably at the Smithsonian, Texas A&M, and the Biodiversity Heritage Library, and they are mostly scanning field notebooks. The University of Michigan Library has undertaken a citizen science project on Zooniverse to identify data in the papers written by students at the University’s biological field station. Similarly, a librarian at Stanford is working with the long-term historical analog data from the California Cooperative Oceanic Fisheries Investigations (CalCOFI) program to digitize, convert it to machine readable format and share it publicly. Additionally, the University Archivist and a librarian at Oregon State University have worked to digitize historical public health records in Oregon to improve access to the 19th and 20th century data (poster).

Scientific researchers have probably been reusing older analog data for decades but it can be hard to identify this activity unless you carefully read the materials and methods sections of their papers. In many cases, authors have used data that was known to exist and was probably relatively easy to locate. Examples include weather data and other government data such as land use, agricultural production, and plant surveys. A few authors such as Buma utilized data kept in university archives to conduct longitudinal ecological research. Buma also drew attention to the value in using older data in a Science commentary.

**Known Issues with Analog Data**

Even though there has not been a great deal of work done to investigate the topic of discovering and reusing analog data, a number of issues are already known. Older analog data may be missing the metadata or other
documentation that would allow current researchers to understand how the data was collected and thus be able
to reuse it or repeat the study. The habits of researchers were very different in an era when funding agencies did
not value or require that data be saved with its future use in mind.

Some individual researchers as well as larger entities like centers and departments are holding on to analog
data because they feel that it might one day be useful, for either their work or the work of others, but they are
not sure exactly what to do with it or who to ask for assistance. They may be aware that their needs include or-
ganization and preservation but it may not be high on their priority lists and they may not even realize that their
institutions have experts in the library and other departments that can work with them.

Potential Roles and Considerations for Libraries
Data services have become one of the most visible services that college and university libraries provide and one
that is closely embedded in the research process. These services may involve tasks such as talking with individu-
als and collaborative groups about how to manage their own data, teaching sessions and workshops around
data management and data sharing topics, assisting in preparing and writing data management plans, helping
researchers find and use existing data, working with data producers to prepare data for sharing, curating data,
and hosting digital data repositories. For the vast majority, these services have focused on born-digital, machine-
readable data. However, the skills that we have developed in data management and data sharing are no less useful
in the analog environment. Libraries can adapt our services to help those on campus who hold historical analog
research data or are still collecting data in analog format.

Librarians should aim to work with researchers to avoid the loss of analog data and to help enhance access
to promote both reusability and replication. We envision this as a new opportunity to form meaningful relation-
ships with researchers, in an area where they need assistance and would expect the library to have a role.

There are some unique issues that will have to be taken into consideration when working with analog data,
such as making sure materials stay in order, that the only copy that exists is not lost, and that materials are safe
from physical degradation. Fortunately, university archives has been working with these kinds of materials for a
long time and has the expertise to help librarians and researchers navigate these issues. We further would hope
that this work would help raise the visibility of archives, by highlighting the services they can provide and the
resources they house, and by bringing more scientific researchers to utilize archives.

Our Progress to Date
Our interest in preserving analog data arose due to a request from the Horticulture department on what to do
with over 100 years of lab notebooks full of potato breeding data. The department was not sure they could safely
house the notebooks over the long term and wanted assistance from the library in deciding next steps. In this
case, we helped to provide temporary storage and are working to find long-term solutions for how to preserve
the data and how to increase both discovery and access so that others can use it.

Since this initial conversation, other researchers with similar problems have approached us (some via referral
and some on their own) and we selectively reached out to still others who we suspected were holding analog
data (see Table 1). Through these informal conversations and referrals we have reason to believe that a large
amount of analog data exists on our campus.

As we have held conversations with these researchers, we have discovered that there is no one-size-fits-all
approach for preservation and access to this data. Some situations that we have encountered:
• Researchers are still adding to the analog data collection and need to reference the historical data fre-
quently, so it needs to stay on site
Researchers have various types of analog data, including a combination of photographs, slides, videos, sound recordings, field notebooks, and drawings—all for the same project. Researchers are storing different parts of the analog dataset in two different places and only part of it has been digitized. Researchers have analog data that describes physical specimens that are still being held. Therefore, we realized that a myriad of solutions may exist for how to best preserve and provide increased access for these data. For example, when we learned that the Horticultural Research Center's (HRC) 100 years of fruit breeding data was still being actively used by researchers on a daily basis, we knew that having them transfer these datasets to our University Archives for preservation was not a workable solution. We also knew that digitizing the entire collection would not be possible in the short term, due to its immense size and inclusion of a small amount of proprietary data. However, we also learned via discussions that organizing and cataloging the lab notebooks would be a valuable exercise, as the HRC often got requests from outside researchers and did not know exactly where to find the appropriate notebooks. We worked with the researchers to create a data inventory, with a data dictionary and controlled vocabulary of terms/keywords (e.g., "field observations", "breeding material") to allow them to search the inventory quickly and efficiently. The researchers seem very happy with the present solution and we are currently working with them to selectively digitize high-priority materials for preservation. In another case, we worked with a former graduate student in the Ecology department to digitize the complete dataset from his doctoral degree, which contained numerous media formats. This was possible because it was limited in size and all formats could be packaged together to share via the University Digital Conservancy.\footnote{\textsuperscript{16}}

Along with investigating the amount and kind of analog data found within departments and labs, we looked for data within our University Archives. Although it has not been the focus for the University Archives to collect raw life sciences data, we suspected that there was some data in older collections and initial searches of the finding aids supported this hypothesis. We started by searching Archives’ finding aids for the term “data” and limited our search to collections within the agricultural and environmental sciences. Since there is no controlled vocabulary to describe datasets in the Archives’ finding aids, we recorded words and phrases that could be used to locate potential datasets, such as “field notebook”, “tally sheet”, “observations”, and “survey”. We examined 15

<table>
<thead>
<tr>
<th>Campus group or Department</th>
<th>Topic(s) of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bell Museum of Natural History</td>
<td>Plants, Mammals &amp; Birds specimen labels and collecting records</td>
</tr>
<tr>
<td>Cloquet Forestry Center</td>
<td>Forestry studies</td>
</tr>
<tr>
<td>Department of Horticultural Science</td>
<td>Potato breeding, Roses, Chrysanthemums, Turf Grass, Golf course surveys</td>
</tr>
<tr>
<td>Horticultural Research Center at the Arboretum</td>
<td>Fruit breeding</td>
</tr>
<tr>
<td>Insect Collection</td>
<td>Insect collecting records</td>
</tr>
<tr>
<td>Department of Applied Economics</td>
<td>Economics of agricultural research</td>
</tr>
<tr>
<td>Itasca Biological Station</td>
<td>Plant, animal, and aquatic studies</td>
</tr>
<tr>
<td>Hubachek Wilderness Research Center</td>
<td>Plant, animal, and aquatic studies</td>
</tr>
<tr>
<td>Bee Lab</td>
<td>Honey bee breeding</td>
</tr>
<tr>
<td>Department of Ecology, Evolution &amp; Behavior</td>
<td>Mountain plover observations</td>
</tr>
</tbody>
</table>

\[TABLE 1\]

List of groups and departments that we have talked to about their analog data
collections, including both individual researcher’s collections and those from research centers or departments. In order to determine how well the data was described for potential reuse, we created an inventory recording data such as: date range and location where data was collected, if column and row headings were explained, if there was a description of the methods provided anywhere, and if the methods were detailed enough to be able to repeat or replicate the work. This initial review has proved that there is raw data on various scientific topics in the University Archives that could be reused.

**Next Steps**

As the project moves forward, there are several issues that will need to be explored. In order to determine the scope of the services that libraries could offer around analog data, we will need to ascertain the extent of data held on campus. From sampling collections in our University Archives, we have an idea of how much data exists there; however, it is unclear how much data researchers have held onto in analog format in their labs, departments, and personal collections off-campus. Surveying and talking to faculty will be necessary to get an accurate representation of the extent of the issue. A survey of agricultural and biological sciences faculty is currently underway and follow-up interviews will be conducted in the coming months.

We will also need to figure out how we can best assist scientists with their analog data—especially if they do not give their data to archives. At this stage, we do not know what other library-based services we could offer for analog data. However, as we work with more researchers and datasets, it is likely that we will have a better understanding of the specific challenges researchers encounter and opportunities for libraries to help.

Additionally, if one of our goals is to increase access to analog data and facilitate reuse, we need to examine metadata schema (including those specifically for archival materials) to figure out a workable application for analog datasets. We will need to determine if additional access points would compel more scientists to indeed find and use these datasets. This would also mean that we need to be working with the broader data community to discuss how to make analog data more discoverable, whether its via an existing searchable registry or index or some new mechanism. Appropriate solutions and steps forward may be revealed as we work with more researchers in labs and departments, discuss the kinds of analog data they retain, and aid them in cataloging and describing their data.

Based on our experiences, we know that our institution is not alone in having analog data stored in various places all over campus. It would be highly beneficial at this stage to work with other academic institutions to see if their analog data landscape is similar and to work together to figure out large-scale solutions to better preserve this data, and increase its discoverability and reuse.

**Notes**


