SciDataCon 2025



Contribution ID: 68

Type: Presentation

## Reexamination of historical secondary data given federal funding cuts?

Tuesday 14 October 2025 11:52 (11 minutes)

Older data in paper or analog format (e.g., field/lab notebooks, photos, maps) held in labs, offices, and archives across research institutions are an often overlooked resource for potential reuse in new scientific studies. However, with the uncertainty around federal funding and research administration in the United States, there has been speculation that scientists across multiple disciplines will be utilizing more secondary data and that historical data will become more valuable. Reuse of historical data is particularly important in studies of biodiversity and climate change.

We have been examining the landscape of historical scientific data use and scientific researcher perspectives on their use of historical analog data. We have also been investigating scalable institutional workflows for organizing, describing and digitizing paper data at a large research University. When transitioning historical data that originates in paper to digital, we identified several unique challenges, including the presence of sticky notes, difficulty in interpreting handwritten notes, and unclear provenance and dates.

Our research tells us that scientific researchers care about this data, consider it valuable, and that these datasets exist in large amounts and are a potentially large, untapped resource. We explore how these datasets could be unearthed and shared to wide benefit. Currently, there are few mechanisms to help researchers find existing historical analog data in order to reuse it. This is a persistent problem, as evidenced by several large projects that have tried to address this over the last few decades. Large-scale solutions have not been identified to date and researchers are siloed by discipline. We are seeking solutions that can work at different scales, across disciplines, and for both researchers and data stewards.

We will discuss example projects involving fruit breeding data, and methodology including how data curators and researchers can work together to balance repository criteria (authenticity, data integrity, reliability and persistence of service) and expectations of the scientists, curators and the research community. The work to organize and convert these datasets from paper to digital formats and the time spent with data producers in improving the metadata and description increased the likelihood that this data would be more in alignment with the FAIR principles.

This presentation will be beneficial to data curators working with contemporary and historical data and scientists interested in exploring legacy data that could be applicable to current and future research.

Primary author: FARRELL, Shannon (University of Minnesota Libraries)

**Co-authors:** KELLY, Julia; MASTEL, Kristen (University of Minnesota Libraries); HENDRICKSON, Lois (University of Minnesota Libraries); GALLOWAY, Nikki (Old Dominion University); SPARROW, Stephanie (University of Minnesota Libraries)

Presenter: FARRELL, Shannon (University of Minnesota Libraries)

Session Classification: Presentations Session 4: Data Stewardship

Track Classification: SciDataCon Persistent Themes: Data Stewardship