Cultivating the Scientific Data of the Morrow Plots

Visualization and Data Curation for a Long-term Agricultural Experiment

Bethany G. Anderson
University of Illinois
United States
bgandrsn@illinois.edu
0000-0001-6602-1312

Sandi L. Caldrone
University of Illinois
United States
caldron2@illinois.edu
0000-0001-6392-5279

Joshua Henry
University of Illinois
United States
jkhenry@illinois.edu
0000-0002-7826-5960

Heidi J. Imker
University of Illinois
United States
imker@illinois.edu
0000-0003-4748-7453

Hoa Luong
University of Illinois
United States
hluong2@illinois.edu
0000-0001-6758-5419

Kelli Trei
University of Illinois
United States
ktrei2@illinois.edu
0000-0002-6436-1011

Sarah C. Williams
University of Illinois
United States
scwillms@illinois.edu
0000-0001-7968-1870

Abstract – The Morrow Plots at the University of Illinois at Urbana-Champaign are the longest-running continuous experimental agricultural fields in the Americas. This paper discusses efforts to identify, curate, and preserve data from the Morrow Plots and visualization tools to enhance understanding of the historical and scientific context for the data. This ongoing effort to draw attention to the greater scientific value of the Morrow Plots and to test data curation and visualization methods underscores the importance of interdisciplinary collaborations to curate longitudinal scientific data sets.

Keywords – data, agriculture, archives, curation, visualization

Conference Topics – Community; Exchange

I. Introduction

The Morrow Plots at the University of Illinois at Urbana-Champaign (UIUC) are the longest-running continuous agricultural fields in the Americas. Established in 1876 by the College of Agriculture and professors Manly Miles and George E. Morrow, the plots were created to facilitate a long-term experiment with crop rotations and fertilization. In 1968, at UIUC's centennial, the plots were designated a National Historic Landmark. The duration and uniqueness of the experiment garnered the plots historical significance. At the plots' designation, Congressman William L. Springer noted that "...through scientifically proven practices, the productive capacity of an acre of land can be multiplied fourfold" [1]. Despite recognition of the plots' scientific value, it is the duration of the experiment that scholars typically cite [2]. Assessing and understanding the Morrow Plots' scientific value has been difficult due to the scattered nature of the plots' data across various archival sources, which have been published in a piecemeal manner over time. Additionally, sources offer different information about the maintenance of the plots, the factors that effected their yields, and the ways the experiment evolved over time. The distributed nature of the data—and information about the data and the plots more generally—thus poses challenges for understanding the greater impact of the Morrow Plots and for accessing the data of this significant longitudinal agricultural experiment.

18th International Conference on Digital Preservation
iPres 2022, Glasgow, Scotland.

To make the data set from the plots publicly available, and to celebrate the plots' sesquicentennial in 2026, the Morrow Plots Data Curation Working Group was established in 2018. Comprising agriculture and life sciences librarians, data management and curation specialists, an IT professional, and a science archivist, the working group seeks to identify archival records in digital and analog formats and aggregate, curate, and preserve the data in a usable and accessible format that can be broadly shared and preserved through the Illinois Data Bank [3].[1] Apart from curating the data set and identifying extant records to be transferred to the University of Illinois Archives, the working group seeks to create best practices and share lessons learned for curating and preserving a longitudinal agricultural data set. One of the challenges of curating the Morrow Plots data is displaying the data set in a format that can account for and illustrate variations and the ways the plots themselves evolved over time. The working group is also testing the ways that visualization can complement data aggregation and curation to provide a deeper understanding of the factors that influenced the plots and its scientific and historical context.

In this paper, we discuss the efforts of the Morrow Plots Data Curation Working Group in identifying and preserving relevant materials from the Plots, challenges in converting the data into reusable format to support open science, and the creation of a visualization that provides historical and scientific context for the data set to tell the story of the Morrow Plots. This case study of an ongoing effort to draw attention to the greater scientific significance of the Morrow Plots and test data curation and visualization methods and tools demonstrates the importance of interdisciplinary collaborations to curate longitudinal data sets. At the same time, we hope to demonstrate the value of visualization in complementing data curation efforts to facilitate historical understanding and scientific engagement.

II. Background

A. *History of the Morrow Plots*

In 1876, ten half-acre plots of land for corn, oats and clover hay were planted by Manly Miles, a professor of agriculture at the Illinois Industrial University. Initially known as "Experiment 23," the plots continued to be developed by the first dean of the College of Agriculture, George E. Morrow (1878-1894). It was Morrow who asked the university's Board of Trustees in 1880 for a "…. formal commencement of what is designed to be a long continued experiment to show the effect of rotation of crops, contrasted with continuous corn growing with and without manuring, and also the effect of clover and grass in a rotation" [4]. While Experiment 23 was one of several agricultural experiments at the university, it specifically focused on the study of crop rotation. In 1895, ten years after the Illinois Industrial University was renamed the University of Illinois, an astronomical observatory was built on plots 1 and 2. An expanding university further reduced the experimental fields in 1903 to three remaining plots which were subdivided (3, 4, and 5). Despite this reduction, faculty continued work with the plots, such as Professor Cyril G. Hopkins, head of the Department of Agronomy (1900-1919), who focused his research on soil fertility [5].

Data of yields from the plots were not recorded between 1876 and 1887 [6]. At that time, six plots contained corn, two of oats, and two of clover (the latter being introduced in 1881). The introduction of fertilizer enabled study of not only crop rotation, but also the ways that fertilizers could enhance yield. The UIUC's student newspaper in 1927 noted, "Soil receiving no treatment in the three year rotation averaged 50 bushels of corn, 45 bushels of oats, and two tons of clover per acre. The portion of the plot receiving treatment aver 67 bushels of corn, 63 bushels of oats and 3.6 tons of clover per acre during this period" [7]. Over time, commercial fertilizers began to be used (1955) and oats were eventually replaced with soybeans (1967). The latter coincided with the university's growing interest in soybean research, including the establishment of an international soybean program in 1966 [8]. Today, faculty, students, and staff continue to study crop rotation and factors that affect the Morrow Plots' yields.

[1] For more information about the preservation architecture of the Illinois Data Bank, see https://journal.code4lib.org/articles/15821.

18th International Conference on Digital Preservation
iPres 2022, Glasgow, Scotland.

B. Morrow Plots Working Group

The Morrow Plots Working Group was formed in 2018 by the College of Agriculture, Consumer, and Environmental Sciences (ACES) at UIUC. Given the significance of the plots, and their sesquicentennial in 2026, the working group's mission is to identify and make publicly available data from the plots to facilitate use of the data and engagement with the plots' scientific legacy. The working group comprises an interdisciplinary team from both ACES and the University of Illinois Library that includes agriculture and life sciences librarians, data management and curation specialists, an IT professional, and a science archivist. The working group has engaged in several activities, including oral history interviews with faculty and staff on the history of the plots; data curation efforts; identification of relevant archival records and creation of a topic guide; and digitization of materials for public access. These efforts aim to broadly promote the history and scientific value of the Morrow Plots, and ensure its data is preserved and made accessible.

III. Morrow Plots Data

Historical records tell us that we should expect to find crop rotation schedules for every year dating back to 1876, as well as yield and soil treatment data going back to 1888 [2]. One of the working group's aims is to clean and compile the data for all available years with the ultimate goal of creating learning objects for use in data science education. The data were originally recorded in ledgers but have been partially compiled by scholars who have previously published on the plots [2], [9], [10]. These were prepared for print and one of the chief challenges is creating a comprehensive machine-readable data set.

A. Legacy Data

Two farm/field managers from the Department of Crop Sciences at UIUC compiled existing data in two Excel files (in XLS format) that correspond with different phases of the experiment. The first file, which appears to be formatted for print, contains three parallel tables in the same sheet, one for each plot. All three tables are almost entirely complete and track year, plot, and soil treatment data from 1888 through 1954. The three-plot format with two layers of headings make it very easy for humans to interpret immediately, but the file needs to be completely reformatted, with new variables added, to make it machine readable.

The second file, which tracks planting and yield data from 1955 through 2021,[2] is formatted for analysis in Excel and takes advantage of some of that software's many special features, like embedded charts and color coding, which provide a richer context for the data, but create challenges for both machine readability and digital preservation. Like the first file, the second file tracks year, plot, and soil treatment data, and includes additional variables for hybrid/variety, planting date, removed stover amounts and population (plants/acre), some of which are rather sparse. The second file also includes some ambiguities common to data sets that have not yet been curated, such as duplicate copies of the table in additional sheets, columns containing more than one data format or unit of measurement, and color coding of both cells and text.

B. Data Wrangling

We decided to employ the tidy data model for tabular data in which every column is a variable and every row an observation with one value per cell [11]. This format is supported by the tidyverse, a coordinated collection of R packages for data cleaning, visualization, and modeling. The tidy data format makes it easy to connect multiple data sets and pivot between different visualizations, which is useful for exploratory analysis. It can, however, require quite a bit of data wrangling up front. The tidy data format also addresses issues with data cleanliness and consistency that often do not arise until the publication and preservation stage. This is in a way a preemptive data curation strategy.

Although we plan to use R for the bulk of our work with the data, we started in Excel. Using Excel is risky because any transformations are made directly to the data. Excel also does not keep change logs or allow for easy backtracking. It is, however, expedient. We found it useful for quickly exploring the data and experimenting with different arrangements. One member of the working group, Heidi Imker, used Excel to transform a subset of the data into the tidy data format. To mitigate risk, she saved a series of versions at key points in the process. She was then able to import that data into RStudio, and link it to

[2] Data collection is ongoing, and we intend to expand our data set over time.

18th International Conference on Digital Preservation
iPres 2022, Glasgow, Scotland.

weather and crop price data, providing essential context for interpreting the data. Now that we have that end result as a guide, we can reproduce most of those steps in a safer tool like RStudio that allows for data manipulations in the software without changing the underlying data file.

This data cleaning is forcing us to grapple with difficult questions such as, how should we deal with notes made in cells that should contain numerical data? Text notes in the soil treatment variables are particularly challenging. Most treatments are measured in pounds with separate amounts recorded for the specific treatments applied. However, many cells contain complex text statements, which are specific enough to be valuable, but not structured enough to be machine readable. Some of these notes also record treatments measured in gallons instead of pounds, compounding the problem. To strike a balance between retaining the original information and cleaning the data, we shifted text comments to a Notes column, and flagged the plots as treated even though no specific amounts are recorded in the treatment columns. This at least preserves the original information in case we need it later. It also may be useful in keyword searches of the data set.

Overly complex notes can appear in any data set, but longitudinal data sets like this one face a particularly thorny challenge—how do we represent change to the experiment design over time? At first glance, the Morrow Plots themselves may look like they have not changed much while the campus expanded and grew around them, but the data tells a much more complicated story, especially when it comes to plot divisions.

Over time, the original plots were subdivided again and again as new variables were introduced. Each original plot eventually became eight subplots. One option would be to impose those subdivisions backwards in time and split old data into eighths. That would allow us to make clearer comparisons over time, but at a cost. It elongates the data for early years eight-fold. It also has a way of flattening time and presenting all the complexities of history at once. We could also widen the data set and create separate columns for plot and subplot. Whichever route we take, we will be sure to document our reasoning and include an explanation in the data documentation. Perhaps we will incorporate decisions like this one into the learning objects we eventually produce, giving students the opportunity to explore the pros and cons of various data wrangling strategies.

C. Communicating Data Context

Data cleaning decisions like these require a firm grasp on the experiment design and its history. We can look to publications about the plots for context, but it takes work to translate narrative text into a mental map, and even more to keep track of how that map changes at key points in time. Visualizations and visual aids are much better suited to communicating spatial concepts like plot divisions, and current design tools make it easy to layer in all kinds of symbols that communicate much more quickly and easily than words.

To aid our own understanding of the history of the experiment, we created a Morrow Plots infographic (Fig. 1) that visualizes the plot divisions, how and when they changed, the crops grown, the rotation schedule, and the key phases of the experiment. We also included a timeline of historic markers to provide additional context and emphasize the experiment's longevity. The visual medium allows us to communicate not just with words and labels but with icons and colors. We used Canva, an online visual design tool with libraries of drag and drop graphic elements. Although these additional layers make the infographic more complex than the charts and maps typically found in academic publications, they paradoxically make it much easier to understand.

The visuals make it more engaging as well. Although we have not yet published it, drafts have been shared with several stakeholders across campus because it is eye-catching and fun. When people see it, they want to learn more and share it with others. It is still in draft form, but once it is finalized, we expect to use it as part of the upcoming anniversary celebrations and as context for any learning objects we produce from the data. We also hope it will help us reach beyond the typical audience for academic publications and engage the broader community.

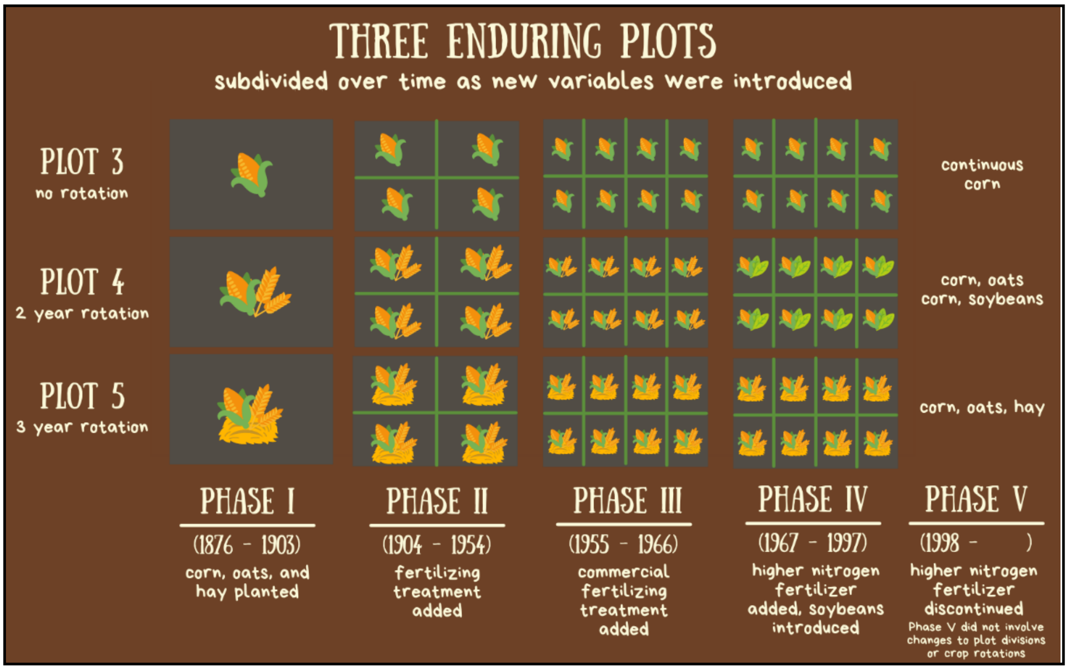


Figure 1 One section of draft infographic communicating plot divisions and crop rotations over time. See Appendix for full graphic.

IV. Conclusion

While the efforts of the Morrow Plots Working Group have made progress in discovering previously hidden data, curating the data, and enriching the connected context around it, there is still much work to do. Visualizations like the infographic included with this paper help to move the work forward by making both the richness and complexity of the data set more accessible and engaging for a broader audience. The tidy format produces preservation-friendly CSV files that distill the complexities of the data for broader accessibility. Efforts to fully curate the available yield data in the tidy data format for Plot 3 of the experiment are nearly complete, but this work still needs to be completed for the remaining plots. Additionally, work to identify and acquire additional archival sources continues, primarily through the work of the informational interviews with researchers and staff who through the years have worked with the Morrow Plots in some way. Finally, we hope for this work to culminate in 2026 with an interdisciplinary celebration and symposium of the historical and scientific contributions of the Morrow Plots.

REFERENCES

[1] "Board of Trustees Minutes –1970," *Board of Trustees Biennial Reports*, Record Series 1/1/802, p. 118, University of Illinois Archives.

[2] S. Aref and M. M. Wander, "Long-Term Trends of Corn Yield and Soil Organic Matter in Different Crop Sequences and Soil Fertility Treatments on the Morrow Plots," *Adv. Agron.*, 1997, doi: 10.1016/S0065-2113(08)60568-4.

[3] "Illinois Data Bank." https://databank.illinois.edu/ (accessed Mar. 07, 2022).

[4] "Board of Trustees Minutes –1878-1880," *Board of Trustees Biennial Reports*, Record Series 1/1/802, p. 232, University of Illinois Archives.

[5] *Cyril G. Hopkins Papers*, Record Series 8/6/21, University of Illinois Archives.

[6] *Morrow Plots Notebook*, Record Series 8/6/16, University of Illinois Archives.

[7] C. Stonberg, "Benefits of Soil Treatment, Crop Rotation Shown," *Daily Illini*, Jul. 08, 1927.

[8] "Board of Trustees Minutes –54th Report," September 21, 1966, p. 84., University of Illinois Archives.

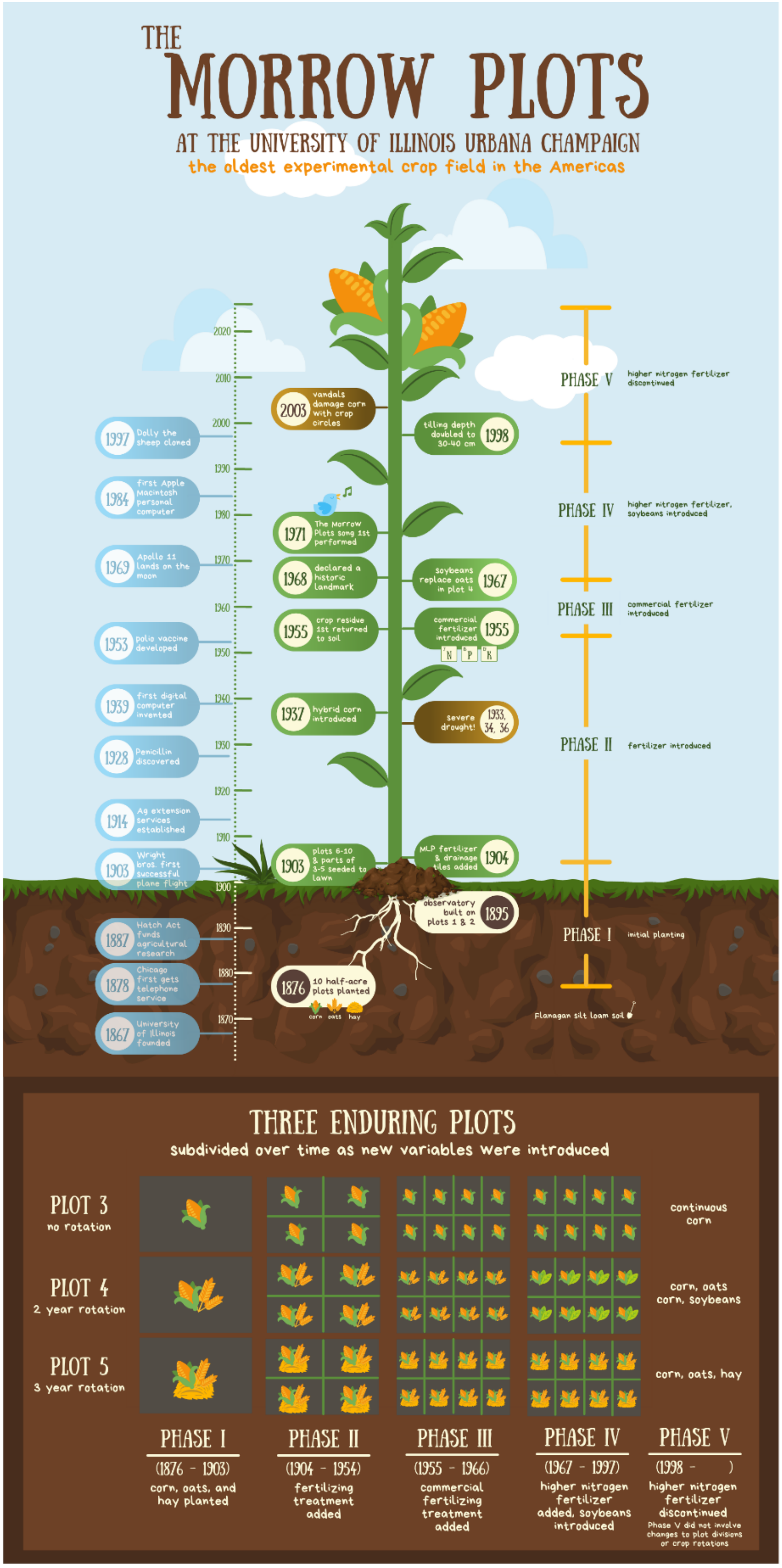
[9] R. T. Odell, S. W. Melsted, and W. M. Walker, "Changes in organic carbon and nitrogen of morrow plot soils under different treatments, 1904–1973," *Soil Sci.*, 1984, doi: 10.1097/00010694-198403000-00005.

[10] E. D. Nafziger and R. E. Dunker, "Soil Organic Carbon Trends Over 100 Years in the Morrow Plots," *Agron. J.*, vol. 103, no. 1, pp. 261–267, 2011, doi: 10.2134/agronj2010.0213s.

[11] H. Wickham, "Tidy Data," *J. Stat. Softw.*, vol. 59, pp. 1–23, Sep. 2014, doi: 10.18637/jss.v059.i10.

APPENDIX

Morrow Plots Infographic (draft)



18th International Conference on Digital Preservation
iPres 2022, Glasgow, Scotland.