Suggestions for Maintaining Records for Long-Term Field Studies
A supplement to papers by Berven et al. and Kenefic and Kern (this volume)

Laura Kenefic[1], Christel Kern[2], John Brissette[3], Matthew Russell[4], Lloyd Irland[5], Aaron Weiskittel[4] and Kate Berven[4]

Robert Lewis, former U.S. Forest Service Deputy Chief for Research and Development, used to say that "researchers leave tracks." By this he meant that part of the research process is recording all aspects of a study from initial plans to notes about installation and changes, as well as the raw data. This is especially important for long-term studies, which often span the careers of multiple generations of scientists, field assistants and technicians. This paper offers our perspective on the important tasks of recording long-term studies and preserving those records. This perspective is based on our experiences in re-opening (and attempting to re-open) discontinued long term studies.

We have been working collectively and independently to access data from a number of ongoing and closed studies. The studies of primary interest were initiated by the Forest Service between the 1920s and 1950s. Many of these studies were closed in the 1960s, and we have had variable success re-opening them. Although most of the files from the Partial Cutting Study (1926-1966) at the Dukes Experimental Forest (EF) in Michigan have been recovered (Kenefic and Kern, this volume), very few files exist from the Gale River EF (1927-1958) in New Hampshire (Berven et al., this volume). Those from the Finch-Pruyn EF (1934-1961) and Paul Smith EF (1948-1961) in New York were recently found but are held by a cooperator who allows only limited access (Berven et al., this volume). These long-term studies represent decades of investment, but the data are not readily available and require ongoing preservation efforts. These experiences, as well as years working with long-term data from experimental forests, including the Penobscot EF (1950-present) in Maine, have motivated us to make suggestions for stewardship of long-term studies and data, supported by real-world examples:

1. Scan or enter data from historic files to create electronic archives and improve accessibility. This is a tedious job that will likely take longer than you think is reasonable, but is necessary for data to be used in contemporary analysis.

   ➢ It took more than a year for student workers to scan 40 years of metadata files (memos and reports) from the Partial Cutting Study at the Dukes EF; it has taken three years (and counting) for a technician working part-time to enter the data.

2. Manage data as part of a relational database to ensure lasting availability. Relational databases

[1] U.S. Forest Service, Northern Research Station, Orono, ME

[2] U.S. Forest Service, Northern Research Station, Grand Rapids, MN

[3] U.S. Forest Service, Northern Research Station, Durham, NH

[4] University of Maine, School of Forest Resources, Orono, ME

[5] Yale School of Forestry and Environmental Studies, New Haven, CT

allow data to be organized in a hierarchical structure that reflects the relationships between similar but different types of data.

➢ Sixty years of data from a wide range of studies at the Penobscot EF in Maine have recently been incorporated into a relational database. Metadata (e.g. dates of harvests and geographic coordinates of plot locations) and measurement data (e.g. tree diameter and species) are archived in a form that is readily accessible for researchers to use now and in the future and new data are added as they are collected.

3. Back up all electronic data on a regular basis, and store them in a variety of formats (e.g. servers, external hard drives or online) and in more than one physical location.

➢ A technician working with decades of long-term EF data had the only electronic copies on his computer. Although he thought the system was automatically backed up, this was not the case and more than a year's work was lost when his hard drive failed.

4. Because electronic media become obsolete, it is important to maintain both electronic and paper copies of all files. It is better to have multiple copies than none. Consider using archival-quality media to ensure files will last. For paper records use high-quality paper and store out of direct sunlight.

➢ We have two-dozen magnetic tapes of Forest Service growth and yield data that we can't read because we don't have a tape reader, and there are no paper copies.

➢ George Furnival, the famous forest statistician, left behind a closetful of primary data on IBM punch cards, but there is no index of their content or instructions on how to access them.

5. Update data management programs as needed, or convert data to a new format. Electronic data cannot be accessed if they are of a file type unsupported by currently available computer programs.

➢ More than 7,000 publications in the Forest Service library at the Penobscot EF in Maine were stored in numbered boxes; the publication numbers are listed in an obsolete computer program and the installation disks are 3.5-inch floppies.

6. Study plans, as well as plan revisions and addenda, are critical for later interpretation of the work. Initial plans often change once field work begins; study plan documents and maps must be updated and attached (see #9).

➢ The target residual basal area used in the widely influential "Overmature and Defective" (selection) treatments in the Partial Cutting Study on the Dukes EF (Kenefic and Kern, this volume) was lowered in the 1960s; this change was not recorded in the study plan addendum, but was fortunately discovered in a "Memo to the File" about a related study (see #8).

7. Records of treatments and associated costs have proven invaluable for reconstructing the financial aspects of treatments. Financial and other relevant data (e.g. person hours or equipment and fuel used) may be as worthy of retention as biological data.

➢ A graduate student recently used historic data from a time study of pruning to generate estimates of contemporary treatment costs for a range of scenarios.

8. Develop the habit of writing "Memos to the File" and keeping copies of study-related correspondence. Such documents can describe treatments, unforeseen events (e.g. a flood or windstorm that damaged the plots) or the rationale behind study decisions. They are invaluable (and a source of amusement) for later researchers.

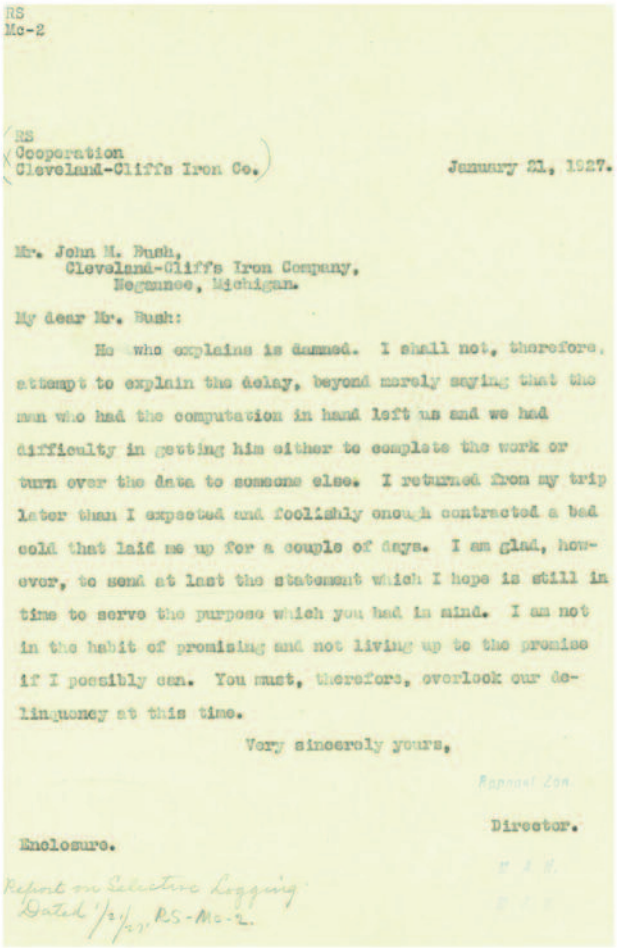
➢ The opening line of a 1927 memo from Raphael Zon (formerly Chief of Silvics for the Forest Service and the first Director of the Lakes States Forest Experiment Station) reads "He who explains is damned. I shall not, therefore, attempt to explain…" (figure 1).

9. Attach or link maps and inventory codes to the tally sheets and the study plan. Ensure that all metadata and supplemental information (study plans, maps, memos and reports) are maintained with the data. Such documents may (and likely

will) become separated from the referencing document over time. Data alone are of little use to later researchers.

- ➢ We have work plans, memos and some remeasurement data from Forest Service weeding experiments at Cherry Mountain and Waterville Valley in New Hampshire (1920s-1940s), but treatments are not described and plot locations are unknown.
- ➢ Species codes for 40 years of inventories in the Partial Cutting Study at the Dukes EF have been lost.
- ➢ Forest Service memos from the Gale River EF about studies in the 1930s-1950s by the father of spruce-fir silviculture, Marinus Westveld, say "see attached maps," but no maps are attached.

10. Adopt commonly used standards for variable names when archiving data.
    - ➢ Species codes in the USDA PLANTS Database (USDA NRCS 2010) and Forest Vegetation Simulator (Van Dyck and Smith-Mateja 2010) are widely used by researchers in multiple disciplines. Using your own system creates data that are meaningless to others, and can create confusion both now and in the future.
    - ➢ We have seen the same abbreviations used for different trees within a single inventory: "TA" for both Tilia americana (basswood) and tamarack, and "SM" for sugar maple, soft (red) maple and striped maple.



RS
Mc-2

RS
Cooperation
Cleveland-Cliffs Iron Co.

January 21, 1927.

Mr. John M. Bush,
Cleveland-Cliffs Iron Company,
Negaunee, Michigan.

My dear Mr. Bush:

He who explains is damned. I shall not, therefore, attempt to explain the delay, beyond merely saying that the man who had the computation in hand left us and we had difficulty in getting him either to complete the work or turn over the data to someone else. I returned from my trip later than I expected and foolishly enough contracted a bad cold that laid me up for a couple of days. I am glad, however, to send at last the statement which I hope is still in time to serve the purpose which you had in mind. I am not in the habit of promising and not living up to the promise if I possibly can. You must, therefore, overlook our delinquency at this time.

Very sincerely yours,

Raphael Zon

Director.

Enclosure.

Report on Selective Logging
Dated 1/21/27, RS-Mc-2.

Figure 1. U.S. Forest Service memo regarding the Dukes Experimental Forest, written by Raphael Zon in 1927.

11. Maintain consistency in codes and measurement protocol over time. Document revised codes and protocol in metadata and study plan addenda (see #6).

    ➢ The minimum tree size measured in the Partial Cutting Study at the Dukes EF varied from 0.5 to 10 inches in diameter at breast height over time, making it difficult to track long-term changes in whole-stand structure and basal area.

12. Maintain photographic images (see Fahey 2010a). Ensure that each photo is labeled with the place and date; otherwise future researchers won't know if the photo shows your work or your summer vacation. Make digital copies of old photos by scanning them at the highest resolution available. It is better to reduce the resolution of an image for a presentation than to not have a high-resolution copy for print media.

    ➢ Thousands of 1950s aerial photographs from Maine were recently discarded despite their potential relevance to studies of land use change because the photos had no geographic or other identifying information.

    ➢ Some photos may show both your work and your summer vacation, such as a circa 1950s photo from the Paul Smith EF that reads "Occasionally a brown trout like this one helps compensate for long winters, low wages and 'conferences' with our wives to explain why we didn't stay at home and cut the lawn..." (figure 2).

Figure 2. U.S. Forest Service photo from the Paul Smith Experimental Forest in New York (about 1950).

13. When possible, store records in fire-proof, climate-controlled facilities (see Fahey 2010b). If this is not possible, at least use metal file cabinets. Avoid storing records in paper boxes or on low sites prone to flooding. Do not use standard staples, paper clips or binder clips to hold paper records because humidity alone will cause them to rust over time. Never use "invisible" tape. Some old facilities have "mouse-proof" rooms; these are not dust-, mold- or insect-proof.

    ➢ A room lined with hardware cloth in the attic of an old Forest Service building yielded the few, intact (but unclean) files we have from the Gale River EF.

    ➢ Many files from long-term research on the Penobscot EF in Maine were recently destroyed when the basement flooded with three feet of water during a storm and power outage that disabled the sump pump; most remaining files are moldy.

    ➢ The original post-acquisition cruise and type map data for the Delta National Forest in Mississippi were lost in an office fire and no copies existed.

14. When handling old files from less-than-ideal storage facilities, wear long sleeves, gloves and a dust mask to minimize exposure to allergens.

    ➢ See #13.

15. Stay in touch with cooperators, the local professional community and colleagues in related fields. Present results at meetings of allied professional societies and invite people to field sessions at your sites. If others are informed about the study's value and potential impact, they can help protect the study's integrity.

    ➢ After the Forest Service research office near the Dukes EF closed in 1981, staff of the Hiawatha National Forest maintained some of the silvicultural treatments. This allowed two of us (Kenefic and Kern) to later reopen the study and generate new long-term results. National Forest staff also made helpful suggestions about decreasing vandalism and recreational damage in the study areas.

16. Take the opportunity to speak with retirees whenever possible; their first-hand knowledge of study specifics is unparalleled, and they may know where missing files are (or have them at home).

    ➢ We revisited a long-term study in Wisconsin using maps available in the study file, but were unable to relocate one of the experimental units. A retiree met us at the site and directed us to the missing unit; former staff knew that the map was wrong. We would never have known this (or have been able to update the map, see #6) without his help.

    ➢ Many of the data from more than 1,000 re-measured plots in a growth and yield study in Maine (1940s-1960s) were stored in our New Hampshire office and discarded during building renovations some years ago; copies taken home by a retiree allowed us to recover some of the data.

17. Canvass your institution for archives and maps of historic value. State agencies, industrial landowners and large private owners may also have similar information (see Greene, this volume).

    ➢ There are many studies in which long-term ecological history has been reconstructed using type maps, cruise data and other information originally gathered for routine land management and commodity production purposes.

    ➢ The Great Northern Paper Company maintained re-measurement plots in Maine for many years; archived data recently served as the basis for a graduate student project on historical forest type and structure.

18. Permanently monument locations of studies, blocks and plots. Consider stakes that can be relocated with a metal detector if necessary. A combination of materials (e.g. metal stakes and plastic flags or ribbons) and bright colors increases the probability that markers will be found in the future; avoid using wooden markers or colors that occur naturally in the forest. GPS locations and references to other known locations will also prove helpful.

    ➢ Wooden stakes used to monument the Partial

Cutting Study at the Dukes EF (1920s-1960s) were either completely decayed or indistinguishable from downed woody material when the study was reopened in the early 2000s. In addition, the white paint used to number trees blended in with white lichen growing on the boles.

- ➢ Squirrels ate most of the aluminum tags nailed to trees in a university cooperator's long-term study in New York; tree numbers could thankfully be determined by recorded distances from plot center.

19. Maintain local copies of data; files sent to central locations are often difficult to get back. Be aware of records disposition schedules; do not send or leave files where they will be destroyed.

- ➢ A 1964 Forest Service memo states that the files from the Gale River EF were "Sent to the Federal Records Center for Permanent Retention." We do not have the record of file transfer needed for retrieval, and neither the Federal Records Center nor National Archives and Records Administration can find the files.
- ➢ For many years, files sent to the Federal Records Center by Forest Service Research and Development were improperly transferred and accepted; today there are hundreds of boxes of files in storage at a federal repository that have unknown contents.
- ➢ Files from the Finch-Pruyn and Paul Smith EFs were left in an office at Paul Smith's College in 1961; they were found in a disused dining hall in the basement of a dormitory in 2009.

20. Maintain files in a permanent and accessible location; do not use descriptors or locations that will be unknown to future researchers.

- ➢ A number of files from the Partial Cutting Study at the Dukes EF were sent to "St. Paul."
- ➢ Files from the Gale River EF were listed in a 1940 memo as in "West's desk."

21. Never let anyone borrow original copies of data or metadata; make copies on-site. Files may be lost in the event of an accident or death. You don't want to have to ask the survivors for them at the funeral.

- ➢ Files and photographs from research in the 1950s at the Sinkin EF in Missouri were recently found in an attic in Maine.
- ➢ Data from a long-term experiment at the Penobscot EF were loaned to a cooperator who died unexpectedly before returning the files.

22. Establish data management guidelines within your agency or institution. For helpful hints and advice, see Borer et al. (2009). Invest in staff training or an expert staff member to keep abreast of technological changes in software, media storage and database management. Assign duties as necessary to ensure that archiving data for future use is not overlooked. Require a senior administrator to sign off on all disposals of primary research information.

- ➢ A research team with long-term study responsibilities in another Forest Service Station maintains two technology experts. One designs and maintains relational databases for dozens of studies and the other keeps abreast of media storage, data collection equipment and programming. This type of investment both protects the data and allows scientists and field technicians to focus on conducting research instead of maintaining research records.

23. Post data online to allow greater access and increase the potential for the study to be continued through data analysis and publication. Consider drafting a data access policy (if one is not already in place) so that the data user and data provider have a clear understanding of the intended use of the data, and proper acknowledgement is given.

- ➢ The Penobscot EF data access policy was recently developed to guide collaboration; this is a good example of striking a balance between the needs of the data user and data provider (Appendix 1).

24. Publish the findings; publication is the best way to ensure that data are preserved (and that people will remember that the study exists). Appendix sections of papers can be used to pre-

serve metadata and even essential portions of primary data. Store electronic and paper copies of the publication in the study files.

- ➢ Publications from the 80-year-old Partial Cutting Study at the Dukes EF were not kept with the study files. Three years after we reopened the study, we are still discovering old and sometimes obscure publications that include important findings and interpretations.
- ➢ Some journals (e.g. the Canadian Journal of Forest Research and Forest Ecology and Management) offer online publication of supplementary data; this may prove helpful for sharing and preserving data that cannot be included in published articles.

25. Recognize and reward the work involved in the preservation of long-term field sites and associated data.

- ➢ Forest Service research work units are guided by documents ("research work unit descriptions") that describe the problems, or topics, to which the unit will devote its resources. The Forest Service authors of this paper belong to a unit that has four problems, one of which is to "maintain long-term manipulative research and monitoring." This explicitly includes managing (collecting, entering, processing and archiving) the datasets associated with the unit's long-term studies.

Summary

Reprising Robert Lewis's thought, we want researchers to not only leave tracks, but to leave durable ones, so that later scientists can carry forward – perhaps with entirely different research questions – work that has been left behind. We expect that financial pressures and changing research priorities will continue to result in the closure of long-term studies; our experiences suggest that many of those studies will prove valuable to future researchers if properly archived. The steps we advocate here will not add much cost at the present, but will surely contribute to making our tracks durable for the future.

Acknowledgements

We thank Judy Terrell of the U.S. Forest Service, Northern Research Station for her helpful feedback on an earlier version of this manuscript.

References

Borer, E.T., E.W. Seabloom, E.B. Jones and M. Schildhauer. 2009. Some simple guidelines for effective data management. Bulletin of the Ecological Society of America 90(2): 205-214.

Fahey, M. 2010a. The care and preservation of photographic prints. http://www.thehenryford.org/research/caring/prints.aspx. Accessed 30 June 2010.

Fahey, M. 2010b. The care and preservation of documents and works of art on paper. http://www.thehenryford.org/research/caring/paper.aspx. Accessed 30 June 2010.

United States Department of Agriculture, Natural Resources Conservation Service. 2010. PLANTS Database. http://plants.usda.gov/index.html. Accessed 30 June 2010.

Van Dyck, M.G. and E.E. Smith-Mateja. 2010. Keyword reference guide for the Forest Vegetation Simulator (revised). U.S. Department of Agriculture, Forest Service. Forest Management Service Center. http://www.fs.fed.us/fmsc/ftp/fvs/docs/gtr/keyword.pdf. Accessed 30 June 2010.

Appendix 1. Penobscot Experimental Forest Data Access Policy (2009-2012).

DATA ACCESS POLICY

U.S. Forest Service, Northern Research Station

Penobscot Experimental Forest

Long-Term Silvicultural Research

A research unit of the U.S. Forest Service has been conducting a long-term silvicultural and ecological study in northern conifers at the Penobscot Experimental Forest in Maine since the 1950s. A series of treatments and periodic re-measurements of numerous response variables on permanent sample plots has produced a dataset that is one of the most comprehensive records of changes to forest structure and composition in North America.

Data and information derived from this publicly funded research are made available with as few restrictions as possible. Making this dataset available has the potential to greatly increase communication, collaboration, and synthesis within and among disciplines. Use of the dataset is supported and encouraged and permission to use these data is granted subject to the following terms:

1. Use the data for academic, research, educational, government, recreational, or other not-for-profit professional purposes.
2. The dataset is provided for use by the Data User and is not to be redistributed.
3. The Data User will acknowledge the support of the U.S. Forest Service in any publication or presentation using these data and documentation. For example:

   "Data for this study were provided by a unit of the Northern Research Station, U.S. Forest Service, located at the Penobscot Experimental Forest in Maine. Significant funding for collection of these data was provided by the U.S. Forest Service."

4. These data have been provided in the spirit of open scientific collaboration. Data users are thus strongly encouraged to consider consultation, collaboration, and/or co-authorship with the U.S. Forest Service investigators conducting this long-term research.

Disclaimer

While substantial efforts are made to ensure the accuracy of data and documentation contained in the dataset, complete accuracy of data and metadata cannot be guaranteed. All data and metadata are made available "as is."

Long-term Silvicultural & Ecological Studies

Results for Science and Management: Volume 2

Ann E. Camp
Yale School of Forestry and Environmental Studies

Lloyd C. Irland
Yale School of Forestry and Environmental Studies

Charles J.W. Carroll
Yale School of Forestry and Environmental StudiesJanuary

January 2013
GISF Research Paper 013

Yale University
School of Forestry and Environmental Studies
Global Institute of Sustainable Forestry
360 Prospect Street, New Haven, Connecticut 06511 USA
www.yale.edu/gisf