

GOOD THINK TANK PRACTICES FOR ARCHIVING RESEARCH PROJECTS

by Raymond J. Struyk

*Struyk is an Offsite Senior Fellow at the Results for Development Institute in Washington, DC. In the 1990s he worked with local founders in Hungary and Russia to set up think tanks and since has helped two dozen others with a range of management issues. His latest book is *Improving Think Tank Management* (2015), available on Amazon.*

Over the past 25 years I have had substantial interaction with a couple dozen think tanks on various administrative questions, and the topic of archiving key products and background materials for research projects has never been raised. At the three think tanks at which I spent my working career, the only guidance provided to me, as the team leader or principal investigator, about organizing materials from completed projects for storage was implicit. I never had an actual conversation about it or saw a relevant policy statement. I am quite certain that my experience is typical.

WHY BOTHER?

Too often archiving does not get much attention until something goes badly wrong. A fairly common scenario is that the principal investigator (PI) on a project leaves a think tank sometime after a project is completed. She takes with her the digital files for projects she has directed; sometime she takes the analog (physical) files as well. The think tank may well not check to be certain that it has copies of the important files. This can be the source of a major crisis later. For example, the project's sponsor (shorthand here for the government agency, foundation, international aid agency or other type of funder) may a year or two later ask for the data set (if were not a deliverable) because it wants to exploit the data to address another policy question. The think tank may not have archived the data set. So it must turn to the PI. In reality, she may or may not have taken care in transferring various files. She also may not want to search for the data very quickly. In the worst case the think tank has to report that it cannot deliver the data set, casting its management in a very poor light.

One can easily imagine other instances in which a project data set could be needed after the project is closed. Given lengthy standard journal review times followed by the wait period for accepted articles to actually reach publication, two years or more may pass after the article is submitted for published quantitative analysis to be challenged. The think tank and author could be in a very awkward position if they were unsure which of the data sets they have should be employed to verify the published results.

Archives are more often exploited than one might imagine. One of the think tanks interviewed for this article reported the results of a survey of its research staff on their archiving practices. In response to a question on how often the respondent relies on information from previous projects to work on current projects, half said "sometimes" and 41 percent responded "very often." Obviously, well-ordered archives make exploiting prior projects more efficient.

Most think tanks have a separate overarching folder for proposals submitted so that they can be mined for future proposals. Similarly, there are cases where survey data or secondary data assembled with considerable effort can be exploited for a new project; so care is taken to archive them. One U.S. think tank told me that it had used data from a large 20-year-old household survey that addressed the specific topic of interest to a sponsor who wanted to learn how these respondents fared over time. The organization had preserved the questionnaires in storage. Sixty percent of respondents in the original sample were successfully contacted to create a nearly ideal data set for the study.

Archiving analog items, such as books, reports with custom printed covers, and banners and other materials for conferences, can play a role in future communications and in preparing accurate institutional histories.

One U.S. think tank told me that it had used data from a large 20-year-old household survey that addressed the specific topic of interest to a sponsor who wanted to learn how these respondents fared over time. The organization had preserved the questionnaires in storage. Sixty percent of respondents in the original sample were successfully contacted to create a nearly ideal data set for the study.

WHAT TO DO?

There seems to be little information published on actual think tank practices on archiving project-related materials to judge from the results of internet searches I recently conducted.¹ To gain a clear impression of current archiving practices and to identify strong practices actually being implemented, I had focused discussions with those with responsibility for archiving at four think tanks, three U.S. and one Russian: NORC at the University of Chicago, the Urban Institute, Results for Development Institute, and the Institute for Urban Economics. I also reviewed their relevant policy statements where available. All four think tanks are well-established; although one is only 10 years old, it already has a staff approaching 150; the other three have each operated for at least 23 years. All conduct quantitative analyses and carry out primary data collection, as well as executing more conceptual projects. There is a surprising range among these organizations in the attention they give to archiving, ranging from “benign neglect” to an impressive system being in place.

The following presents my distillation of a good model for project archiving—a model that captures the essential items and is not too resource intensive to implement. The discussion first covers when to actually archive a project and then moves to explicit consideration of archiving digital and then analog materials; it closes with thoughts on maintenance tasks.

A couple of clarifications before starting. First, I am assuming that all think tanks assign a unique number to each project and that all records are tagged with this number, including contracting documents, financial information, data sets, reports and other products, analytic files, and others. This does not mean that every digital file is labeled with the number but at least overarching project file folders (digital and analog) are so labeled. Second, when I use the term “digital,” I am including both “born digital” documents and digitized analog materials.

1. There is a literature on essential archiving for nonprofit organizations. This is heavily focused on meeting legal requirements for retention periods for various financial records and for key management documents, e.g., the charter and amendments to it, list of board members, minutes of board meetings that highlight decisions taken, and audit records among others. As an example, see the guidance provided for U.S. NGOs by the National Council of Nonprofits at: www.councilofnonprofits.org/tools-resources/document-retention-policies-nonprofits

WHEN TO ARCHIVE?

What is a readily identifiable “trigger event” for initiating the archiving process? A sensible point is when the final product has been formally accepted by the client. Usually the contracts officer will receive a formal message of acceptance. At this point project close out procedures are launched. A new item on the close out to-do list can be “Notify IT and the communications team to archive project materials.” (The communications team is involved for analog materials; see below.) In cases where there is no external sponsor, the contracts officer or chief accountant can ask the project leader or the designated manager if the project is complete, at the end of the performance period specified when the project number is set up.

What is a readily identifiable “trigger event” for initiating the archiving process? A sensible point is when the final product has been formally accepted by the client.

WHAT TO ARCHIVE?

It is advisable to have two research-related digital archives: (a) one for all project-related documentation and (b) one for data sets. There should be additional but separate archives for certain other items that are beyond this post: all proposals submitted, whether funded or not, so parts of them can be used in future proposals; all contracting documentation; and, financial records. The digital project folders maintained by the principal investigator and others will have some information on these “extra” topics that were of use to various staff during project execution, but full information is in the separate folders. The same project number is used to identify all the information stored for a project in all archives.

Digital materials other than data sets

Unless a think tank is quite small, all project information for active projects is kept on one of its servers during the life of project, organized by project number.

A couple of simple folder maintenance practices can make these materials when archived dramatically easier to use in the future. First, within each involved staff member’s overarching project folder the organization’s policy can be that there should be a series of folders with standard names for major topics (and perhaps even some subfolders) covering topics as: the contract, subcontracts, consultants, invoices, data, analyses, reports, events, presentations, and so on. (Defining this set of folders with substantial senior staff input is highly desirable to gain acceptance.) Each staff working on the project maintain the materials for the project in his master folder identified with the project number.² The full contents of all of a project’s folder are archived. This is quicker than attempting to review each staff member’s sub-folders for relevance.

The Principal Investigator will usually have the most complete set of folders. Other staff working on the project have folders containing sub-folders with identical labels but typically only a few have substantial content, typically having to do with data sets, analyses, literature reviews, and similar. At a project’s termination all files in all project folders are archived.

Data sets

These are kept separately because they may be needed for use in a future project (or proposal) or for results verification and it will be easier to properly label and locate this way. It follows that each “data set” needs to include proper documentation so that someone who has not worked with the data previously will be able to do so. Also, if multiple versions of the data exist, the version stored needs to be that set used in the analysis presented in the final report. Of course, multiple versions can be stored but differences among them need to be clearly documented.

It is highly desirable for both primary or secondary data sets be created for the archives following widely accepted standards for describing surveys, questionnaires, statistical data files, and social sciences study-level information. Such protocols include the World Bank’s Data Documentation Initiative (DDI)³ and the Nesstar⁴ protocol. These protocols are such that they permit an analyst new to the data to understand how the data were collected (e.g., sample attributes), examine the raw data, understand adjustments made to the raw data (e.g., imputation of missing values), and access the final data which were employed in the analysis reported in reports, journal, articles and books.

Because sponsors are ever more frequently asking for data sets as a deliverable under contracts, as noted earlier, think tanks should make it a general practice to prepare for this in project work plans and to include in budgets the costs to prepare data sets in good order for storage and future use as well as delivery to sponsors.

Analog materials

While the vast majority of materials to be saved are digital, there will be physical items that a think tank will want to save for future reference or display. Having a physical copy of books produced either by a project or based on a project's research is one example. Bound reports with project-specific covers may be another; conference documents such as handouts that cannot be digitized faithfully are other candidates. Where a project involves primary data collection with paper questionnaires, sponsors may require that the original forms be kept for several years even though the information has been digitized. Different organizations will have different preferences and policies on retention periods.

There is a strong case to assign primary responsibility for gathering these materials to the communications group because it will be involved in organizing events and in preparing publications. In contrast to the end-of-project action for research materials, these analog materials should be collected as they are developed to avoid missing some, which can easily happen if collection is postponed to the end of a multiyear project. Still an end-of-project review with the team leader of materials produced should be conducted. The review should include explicit questions about physical records that need to be retained.

It is a good idea to save at least two copies of everything (except completed survey forms and similar items) so that one can be accessible to staff on a routine basis in the library or similar space and another, the true archive, can be stored in a secure space.

It is advisable to have two research-related digital archives:

(a) one for all project-related documentation and

(b) one for data sets.

2. Ideally the full title of the overall project folder maintained by a specific staff member could have a title such as "6198-health-PI," where "6198" is the actual project number, "health" is the program in which the work is executed, and "PI", for principal investigator or other initials indicating the role of the person who maintained the folder. Other options could be SR for senior researcher, RA for research assistant, STAT for statistician and so on. This kind of labeling would make it much easier for those searching a project's archives to zero in on the most likely location of specific items more efficiently.

3. www.en.wikipedia.org/wiki/Data_Documentation_Initiative

4. www.nesstar.com/help/4.0/webview/getting-started/getting-to-know-nesstar-webview.html

STORAGE AND ACCESS

Digital archives

A fundamental goal of archiving documents in any format is to preserve their integrity. This can especially be challenging for digital archives where files can be easily deleted or overwritten—think of a data set being updated through further work on imputing missing values. To maintain integrity access to folders has to be restricted and even when access is granted limitations on types of file manipulation must be imposed. For example, standard rules can be that no new files, including replacements for existing files, can be saved to a project’s archive folder without case-by-case permission and action by the IT group.

Archives should be kept on a separate, limited access drive on the think tank’s server or in a similar space in the organization’s cloud account. Access to the archives for viewing files or downloading them can be restricted, with the IT group allowing access to those cleared by senior management. All files can be made “read only.” The case for adding files or modifying a file would have to be cleared by a senior manager and the project’s principal investigator to justify the changes proposed.

For small think tanks or start-ups these procedures may be too onerous. In such cases the archives can be placed on a designated computer or equivalent in cloud storage with access limited to those who have received permission from the designated senior manager to access a specific project’s archives.

All the archives should be backed-up regularly to make sure new additions are secure, presumably on the same schedule as other files.

Analog archives

These items must be kept in a secure, i.e., locked, location. There is an infinite number of experiences of someone taking an archived book, report or other material off a shelf or from a box and failing to return it, usually from simple neglect or absent mindedness. Archived physical items could be placed in standard storage boxes in a (lockable) standard closet fitted with shelves. A moderate-sized closet will likely accommodate documents from a surprising number of projects. Ultimately, off-site storage may be needed, but it must be insect-free and temperature controlled.

If the organization has a librarian, giving her the responsibility for maintaining these files seems sensible, after they are handed over by the communications team. Again, archival items should not be intermingled with other items in the library’s collections. When an item is needed, the librarian finds it, has the borrower sign out for it for a defined short period of time, and then follows-up with the borrower if it is not returned promptly. If there is no librarian, then a well-organized person on the administrative staff is a good candidate.

A fundamental goal of archiving documents in any format is to preserve their integrity. This can especially be challenging for digital archives where files can be easily deleted or overwritten—think of a data set being updated through further work on imputing missing values.

ARCHIVE MAINTENANCE

The principal tasks here are the removal of certain items and the proper storage of those retained. Starting with data sets, some contracts require that data sets cannot be retained for more than a specified period, often three years. Other contracts require that they be maintained for longer periods of time. At the time a data set is archived the required removal date should be recorded and maintained in a digital log book by the IT team and a “removal date” put on the calendar. Other data sets can be saved indefinitely: even large digitized data sets do not cost much to store. The volume of analog items on the other hand can build up.

Since proper storage of analog documents requires climate control and safety from destruction by bugs (so, no, boxes of documents cannot be put in the CEO’s garage), cost can mount up over time as the volume of documents expands. Hence, rules should be established for the time intervals between reviews of an organization’s holdings of different types of records. It may be that some, such as hard copies of professionally printed annual reports and books published based on the organization’s research should not be subject to review. It would be very helpful if materials subject to periodic review would be flagged in the document inventory.

Rules should be established for the time intervals between reviews of an organization’s holdings of different types of records. It may be that some, such as hard copies of professionally printed annual reports and books published based on the organization’s research should not be subject to review.

POLICY AND PROCEDURE STATEMENTS

The foregoing indicates that basic P&P statements be developed that cover the topics just discussed. Those responsible for the various actions will not be able to execute them well without clear guidance. It is very important that these be developed with full input from lead researchers, the IT team, the communications team, and the librarian who will be responsible for organizing the materials to be archived and managing the actual archives. Getting their buy-in is essential to the system actually working. One of the four think tanks interviewed is now following exactly this process in upgrading its procedures. Thought could be given to including explicitly an item on the annual performance evaluation on how well those with responsibility for various archival actions have executed them during the course of the year.

One task essential to codify is the identification of the person responsible for checking that the system is actually working. Whoever has the oversight task should be notified by the contracts office when a project is to be archived so that he or she can check a few days later on whether the archiving is complete or at least underway. This “quality control monitor” could be consulted by rating supervisors at the time annual performance reviews are done.

Whoever has the oversight task should be notified by the contracts office when a project is to be archived so that he or she can check a few days later on whether the archiving is complete or at least underway.

VALUE FOR MONEY?

At this point the reader is probably saying that this sounds like a lot of work. It is. Of course, if the tasks involved are done on a regular and timely basis by the various people assigned responsibility, the burden is not so great for anyone. It is when archiving has been neglected that a catch-up operation is really burdensome. Here it makes sense to implement a new system for projects as they close out and delay applying the new procedures to past projects until resources are available.

My guess is that many readers of this post have had awful experiences with missing or very poorly prepared archives, examples of which were given at the start of this post. This should motivate them to consider developing a well-functioning archival system if one does not now exist.

It is when archiving has been neglected that a catch-up operation is really burdensome. Here is makes sense to implement a new system for projects as they close out and delay applying the new procedures to past projects until resources are available.
