

The Information Ecology of Archives

The “information explosion,” about which so much has been said and written, is to a great extent an explosion of misinformation and badly organized information . . . The digital revolution has only made the problem more acute.

—Murray Gell-Mann

“Information versus knowledge and understanding,”
From *Information Ecology* by Thomas H. Davenport

Archives are often perceived as dim, dusty depositories of arcane knowledge and minutia haunted by historians fascinated by the past. Like all stereotypes, there’s a grain of truth in this belief. But, what’s often overlooked by many managers is that archivists (and their colleagues in the library) have significant skills in analyzing and synthesizing knowledge.

In many ways, information technology (IT) departments are reinventing archives as they seek means to preserve and provide access to electronic information. Unfortunately, the staff in IT departments often know programming and computers better than principles of information management.

I believe that in the evolving high-tech information ecosystem, a savvy manager will look at the strengths of these two disciplines and forge a new alliance between them. After all, nothing requires that electronic records be kept in a separate electronic archive.

If an organization is going to effectively manage its information resources, it’s going to have to look at the individuals who create, use, and preserve that information. Thomas Davenport argues in *Information Ecology*¹ that few organizations have a well-developed information management plan. Well, that’s putting it nicely. He says, “For the most part, . . . information environments are appalling. [Organizations] don’t know what they know or what they need to know” (p. 7).

Davenport argues that many of the individuals responsible for managing information resources tend to see information as data—as simple facts—not as knowledge or understanding that makes the data meaningful. Because the information managers come from a computer background, they tend to look to technology for solutions; but until machines possess judgment, they will only be able to respond with data, not knowledge.

I hope to give you an appreciation for what really goes on in an archive. Archivists don’t just put papers in boxes and consign them to oblivion. They ensure the quality of information through a process of appraisal. They arrange that information so that it remains useful. They house the records to preserve the information from deterioration. They describe the records, analyzing and abstracting the information so it’s easy to locate. And finally, they work with researchers who want to reference the records to ensure that they find all relevant information. The traditional principles of archives as embodied in these five steps can serve as a model for managing information resources and remain useful in the age of electronic records.

Appraisal and Acquisition

Given that we simply don’t have the resources to save all the data that comes into an organization, what shall be saved and what shall be lost? With electronic records, this question is even more compelling, as we must actively work against rapid deterioration and obsolescence.

When determining what records to preserve, archivists have traditionally looked for continuing administrative, fiscal, legal, or historical value. But what is administrative and fiscal value? When making a business decision, a manager wants to ensure that his or her information is accurate, timely, accessible, and relevant (see Davenport, p. 117-120). An archivist uses these same values when determining which information to save.

Accuracy. Archives have always placed a premium on acquiring authoritative information. Did these records come from an office charged with collecting or creating this information, or did they come from some other source? Is the source trustworthy? Are the records complete?

Good managers—like good historians and good journalists—look to more than one source of information to confirm facts. Are your managers consulting your organization’s past experience to verify information from other sources and to check the credibility of those sources?

Timeliness. A good archivist reads the company newsletters, talks to people throughout the organization, and knows what’s current. The archivist will tell people outside the archives about potentially useful information before it’s too late. Is your archivist in the information loop? Do you encourage your decision makers to consult with the archivist?

Accessibility. Archivists dislike acquiring collections with restrictions. They want to see that the information in their care is used. And, they take great pains to develop a variety of tools to make that information easy to find so that it can be used.

Do your managers have their own “archives” in their offices? In the best case, where these records are well managed, you might ask why they’re spending time doing the archivist’s work and hoarding information from others in the organization that might need it. In the worst case, the manager can’t find the information in their own file cabinets, nor is it accessible to anyone else. Consider implementing a strong records management program, rewarding individuals who deposit their corporate knowledge in the archives where it can be managed and made available to all people in the company who need it.

Relevancy. Archivists are more concerned about the present and the future than they are the history of your organization. They know it’s easier to try to collect historical information when it’s fresh and available. So they seek to acquire information relevant to your corporate mission and current activities, knowing it will become part of history.

Finally, archivists talk about the primary and secondary value of records (not to be confused with primary and secondary sources). These distinctions have little to do with the usefulness of the information; the secondary value of records may, in some instances, be more useful than their primary value. The primary value distinguishes the information the records creator intended to capture in the course of business; secondary value refers to other information captured as a by-product of the records creation process.

Archivists ask the same kinds of questions about the value of information in records, regardless of the records’ format. While people tend to associate archives with textual records, archival collections include photographs, maps, sound recordings, and three-dimensional objects. After all, “text” no longer means just paper records; word processing files on disk are textual records.

Determining the value of records is an inexact art. Some of the most interesting uses of archival materials comes from the innovative use of records. Information captured for one purpose re-used in a novel way can help us understand something in a new light or with renewed appreciation. As much as possible, archivists try to add value to their collections by suggesting ways in which their collections can be used.

Arrangement

Libraries generally keep their books in a single large collection, which is organized according

to a system established by the library—typically LC or Dewey call numbers. On the other hand, archives typically follow the principles of provenance and original order; collections from different sources are kept apart, and the records are kept in the order used by the creator. Practically, respect for provenance and original order saves the archivist time reorganizing the materials into some artificial order developed by the archive and allows the repository to exploit any existing access system developed by the records’ creator (which one assumes to be useful, especially for the purpose for which the records were created).

Provenance is the organization or individual responsible for creating the records as a whole. That doesn’t mean the provenance is responsible for authoring every letter, memo, or report in the collection. The collection is created through the aggregation of documents from a variety of individuals and organizations as the by-product of routine activities.

Respecting provenance—keeping the records of different sources separate in the archives—is a useful tool for accessing the information in those records as provenance provides a good clue to the type of materials likely to be in the collection. When looking for a certain type of information, one thinks of an agency that would have generated or tracked that information in the course of business, then consults the records of that agency.

Moreover, the nature of the office may suggest the character of information to be found in those records. For instance, if someone were researching AIDS and had access to the archives of the Center for Disease Control, Jerry Falwell, Dell Computers, and President Reagan, what’s the most likely source of information? The CDC jumps out as the “right” answer because it is a leader in studying infectious disease. But the papers of Ronald Reagan might be more useful in studying responses to the disease in federal policy, and the papers of Jerry Falwell might be more useful to gain a perspective (albeit it slanted) on socio-religious attitudes toward the disease. But, the likelihood of the corporate archives of Dell Computers containing much relevant information is pretty slim.

Original order is the organizational system used by the records’ creator, and generally reflects the routine activities in which the records were used. Original order is not necessarily the order of the records as received at the archive; some records may have been misfiled in the office, and order may have been disturbed during transfer. The archivist will determine how the records were organized in the office of creation, then arrange the records according to that order.

The relationship among the documents in the collection is itself a significant bearer of meaning. If the records are filed chronologically, an undated document may be dated by looking at adjacent records. Gaps in original order may be important clues to missing materials.

As noted above, maintaining original order is a time saving strategy that exploits any inherent retrieval system. As such, an archivist may describe the manner in which records are organized before describing the records themselves. The archivist will ensure that any indices, finding aids, or other access tools received with the collection are readily accessible.

Note that sometimes the records have no order, often because the records were not organized in the office of origin. And, respect for original order does not extend to respect for original chaos. In these situations, the archivist may impose a simple organizational scheme on the records.

Housing and Preservation

Because the records in archives are of enduring value, the archivist must preserve the physical carrier of the information. Acidic paper used for blueprints and field notes becomes brittle with age and must be reformatted. Photographs of important events fade. Electronic media holding accounting information, oral histories, and retirement records suffer bit loss and media deterioration. **Temperature, humidity, air borne pollutants, insects, and—most dangerous of all—humans, constantly put archival collections at risk.**

Archivists take several measures to minimize the damage. Documents are transferred from their original file folders and boxes that might damage the documents into archival quality containers (you don't want a collection of documents on the

floor when the box fails). In the process, staples and paper clips may be removed, papers unfolded, and fragile materials may be photocopied onto acid-neutral paper or encapsulated in Mylar. Materials that are very dirty may be cleaned. Electronic records may be copied onto new media to refresh the data and to circumvent obsolescent formats. Records are stored in environmentally controlled vaults away from vermin of all types.

However, the archivist is not a conservator restoring documents to like-new condition. Rather they are trying to stabilize the records against deterioration and protect them from damage or theft.

Description

Archival description creates a verbal abstraction of the collections, enabling researchers to browse the materials on paper rather than having to rummage through hundreds of boxes. Description summarizes a collection's organization and essential details so that researchers don't have to read every document to know what's in the records.

Archival description does more than make it easier for the researcher. Because researchers look at just those portions of collections they need rather than entire collections, less staff time is spent pulling and retrieving boxes. And, because fewer materials are in the reading room, there's less chance that the records will suffer wear and tear or be stolen.

Archival description documents the archivists' experiential knowledge of the records. It summarizes all the facts the archivist has learned about the collection in a fashion that will be meaningful to those who may need that information. **Description, at its best, is the process of transforming the raw data of the records into knowledge.**

Archivists generally use three tools to describe their collections: a repository guide, summarizing all the holdings; finding aids for each collection, detailing their contents; and an index that complements provenance as an aid in identifying relevant collections. These tools work together progressing from a bird's-eye view of the forest, to maps of specific groves, to pointers to specific trees.²

A repository guide gives a researcher the big picture through a summary listing of all the collections in the archive. Because an archive organizes its collections by provenance, the repository guide is a list of the sources of collections with a brief description of the records from each source. The guide may include a short note about the provenance and a brief description of the materials.

A repository guide is a coarse sieve that helps researchers locate collections likely to contain records relevant to their research. In some

Archivist and records creator discuss transferring records from the Park Central File to the Historic Document Department archives. Maintaining the provenance and original order of the records will ensure their continuing usefulness as evidence of past actions for the administrative staff and for future users. Photo by Campbell/Danford. Courtesy San Francisco Maritime National Historical Park.



instances, the archive may have only a few records from a source, and the repository guide is the only place where those records are described.

Once researchers have identified promising collections, they ask to see the finding aid for that collection. The finding aid describes the collection's contents in greater detail than the repository guide. It facilitates access by bundling similar bits of information into manageable units and embodying the archivist's experiential in notes about the context in which the records were made, their creator, their significance, the reason for their creation, their contents, and their organization. These introductory notes—which may run several pages—give researchers a more complete sense of the collection, confirming the relevance of the materials or suggesting that time might be better spent on other collections.

Finally, a finding aid lists all the series and folder headings. Reading this list is equivalent to opening the drawer of a file cabinet. When researchers decide which folders they want to see, they request them from the archivist.

In some instances, collections that contain very valuable materials may list every item. An item-level inventory is useful evidence for security control, and is generally done only for items that would have a high market value. Ironically, highly detailed description is often a barrier, rather than a benefit, to increased access, as the researcher must read through much more text rather than a good summary. Although once the norm for archives, especially literary manuscript collections, item-level description is so time consuming, it is now relatively rare for entire collections to be described at this level, although description of a few selected items of great value is not uncommon.

Repository guides and finding aids are not perfect access tools. They emphasize a collection's primary value, but most collections contain interesting but tangential information. The hierarchical nature of repository guides and finding aids works well when the researcher is familiar with the names and organizations related to their subject. However, provenance is less useful when valuable information was captured by unexpected offices and, especially in the cases of personal papers, when even the best researchers are not going to be familiar with every source.

To complement traditional name-based top-down archival research methods, archivists index the repository guide and finding aids so that researchers can find relevant information hidden in unexpected places. The index provides more direct access to the contents of the collections than the hierarchical model of provenance and original order.

Description is enormously important for electronic records. While a researcher can call for documents and browse through them to find relevant materials, electronic records are not eye-readable; researchers cannot easily “browse” floppy disks and tapes. Because it's often hard to find the right software and hardware to read older electronic records, effective description is essential to help archivists and researchers know if the information contained in these electronic vaults is worth cracking.

To a large extent electronic records can be described using the same approach as their paper equivalents. Often electronic records contain raw, empirical data. In these instances, the archivist will try to describe the data fields; the software that originally created these records will probably not be serviceable in five to ten years, and the kind of information that the original software provided is likely not the same information needed for subsequent analysis. Because the data structure was documented, it's possible to write a new program to re-evaluate the data.

Reference

Once arranged, housed, and described, a collection is ready for researchers to use. How do those researchers find the collections that might contain useful information? How do researchers know which repository will hold the records?

The archivist is—in many ways—the most important means of locating materials. Although archivists try to translate their experiential knowledge into access tools, it's impossible to describe all possible uses of materials. The conversation between an archivist and researcher often has a collaborative nature, blending the researcher's novel view of a topic with the archivist's knowledge of the collections.

As noted earlier in sections on arrangement and description, provenance is the principal mechanism for locating relevant materials. Even in repositories with well-indexed collections that allow researchers to look up specific subjects directly, provenance remains useful as an access tool. Looking for a subject by identifying the names of those individuals or organizations potentially responsible for collecting information on that subject forces researchers to think about their topic more broadly. When researchers look up a subject in an index and find no entries, they logically assume that the archive has no records on that subject. In fact the archive may hold records with relevant information, but that information is was not immediately recognizable during description. **Researchers are often so motivated by a time deadline that they are more focused on getting an answer than asking the right questions;** approaching the subject indirectly through prove-

nance helps discipline researchers to formulate their queries with care.

Of course, researchers have to know which archive to visit. Provenance remains the best first step in knowing which repository will hold relevant records. Probate records for someone who died in Austin will be at the Travis County Recorder's office. Records of the Phoenix Indian School are probably at the Bureau of Indian Affairs or, for older records, at the National Archives branch in Laguna Nigel. But, while the individual may have died in Austin, his estate may have gone through probate elsewhere, and an important body of Indian School records may have been disposed in accordance with the records schedule but wound up at a museum.

Finally, it's anyone's guess where an individual's personal papers might wind up. Aunt Hattie's photographs may be at the city library, the county historical society, the state archives, or a university special collection library.

Archivists have developed a number of tools to help researchers identify which repositories hold relevant collections. The National Union Catalog of Manuscript Collections (NUCMC) is similar to a repository guide, but includes collections from many different repositories. The guide has now been migrated to the World Wide Web

<<http://lcweb.loc.gov/coll.nucmc/nucmc.html>>.

A little more than 10 years ago, archivists began using online databases to share information about their collections. OCLC and RLIN, two bibliographic utilities with union catalogs of books and other materials, began assembling online archival union catalogs. Unfortunately, access to those utilities was problematic; the interface was difficult to use and access was limited to specialized terminals usually found in libraries. Possibly a more significant barrier was the fact that searching these databases was not free; even now that the utilities have better Web interfaces and are widely available on the Internet, access fees continue to reduce their use. Finally, the collection-level descriptions of these records were often so general that searches for information in the series and folders was not included, limiting the retrieval value of these databases.

In the last five years, archivists have begun to use the Internet to publicize their collections. **The Web is radically changing how archives provide access to their holdings.** Repositories that were not members of OCLC or RLIN could make their finding aids widely available at little or no expense. Because the entire finding aid was available, information in series and folder headings was accessible. And, the Web has made it easy to provide access to the documents themselves through

digital images. With the exception of the Heard Museum's homepage, the Webpages for archival collections are downloaded more than any page on the entire site. The museum has started receiving many more inquiries for use of the archival collections. Fortunately these queries have not significantly added to the reference workload; because the entire finding aids is available the queries generally refer to specific materials, so that the reference archivist doesn't have to consult the finding aid for the researcher.

Ultimately, many researchers find out about archival collections through word of mouth. Often it's through conversations with colleagues or through footnotes in articles. However, the archivist is often an excellent source of information about collections in other repositories.

The Archives of Tomorrow

Possibly the aphorism of the 21st century will be "Death, taxes, and technological development." The last 15 years have seen enormous changes in archives, largely driven by ready access to automation. When PCs made computing inexpensive, archivists immediately adopted them to produce finding aids and indices. Now that desktop machines have the power of 1970s mainframes, we'll be seeing more and more innovation in access as archivists develop more sophisticated mechanisms to manipulate data.

As archivists acquire more and more word processing documents and databases, it will be more common to provide access to the entire contents of the archival collection. As search engines become more sophisticated, locating relevant documents will be easier. And archivists that are specialists in retrieving data will be able to "push" information about relevant collections to researchers who have registered their interests with the archive.

But in the long run the core functions of the archivist, if not unchanged, will remain essential: acquiring and appraising information, housing and preserving it, describing it, and helping researchers reference it. **Throughout this process, transforming data into knowledge will also remain the principal strength of archives.**

Notes

1. New York and Oxford: Oxford University Press, 1997.
2. Readers may want to look at complete samples of the access tools developed by the author on the Heard Museum's Web site <www.Heard.org/library/reguides/>.

Richard Pearce-Moses is the Archivist for the Heard Museum in Phoenix, Arizona. This article is adapted from his speech at the Information Ecosystem Conference.