Amanda Goodman

Final Paper

November 28, 2010

Automation and Technology in the Archives

Technology is no longer just for geeks holed up in their garages or the rumored people who spend their entire lives living in their parents' basement. Instead, the latest technological gadget is highlighted on both the morning and nightly news with on the scene reporters asking customers why they are lining up days ahead of time for the newest product. In this technology saturated culture of the twenty-first century, the spread of fast paced and interconnectivity has changed the very fabric of society. However, the reach of technology into the archives via automation has been a long process.

Automation in archives refers to the process of using technology to hasten processing speed and data retrieval. Michael Cook wrote that "Much of the work which is done within an archives or records service can be described as the processing of data." He specifies that archival data is generated by the archivist to describe the document and "information derived from the archival documents themselves" (Cook 2010). Archival automation for this data is "fundamentally a mechanical or electronic extension of traditional logic -- or deductive and inductive reasoning -- to problem solving and the performance of work" (Fisher and Evans 1967). What Barbara Fisher and Frank B. Evans meant was that the logical organization of archival records and reliable search methods is nothing new to archivists. Every step of the process from accession to guiding researchers to new materials produces documentation that can help in the retrieval of documents. Therefore, making the change from a manual system to an automated one is only a change of medium, not a revolution of the profession itself. This paper

seeks to examine why it took so long for automation to take hold, reasons for and against automation and finally to provide some pointers to decide if automation is right for your archive. First though, a quick overview of what automation can accomplish is in order as to explain what tasks automation can ease the burden of.

Overview of the Benefits of Automation

A primary advantage of automation is to save time by avoiding the manual process of cross-indexing and as well as making other derivatives of online databases of archival records. The list below highlights some ways an automatic system can benefit an archive:

- Subject indexing
- Inventory lists
- Finding aids
- Statistics
- Shelf locators
- Text processing
- Establish intellectual control
- Access to born-digital records

History

The earliest articles that specifically use the term automation in an archival context are from the mid-1960s. However, automation had already appeared in offices as early as the 1950s in the United States. Where had the computer technologies come from that made automation possible?

During World War II, advances in communication as well as mathematics machines assisted in the warfare efforts around the world. Vannevar Bush, an American engineer and science administrator (Wardrip-Fruin and Montfort 2003), wrote an article, "As We May Think" in 1945 that discussed the "rapid development of electronics and electronic communication techniques and devices" (Fisher and Evans 1967) and how they may be applied in the future. Bush's influence encouraged others to continue the development of technology for applications

other than warfare. Shortly afterwards, "Murray G. Lawson foresaw the use of machines by historians and archivists to index collections" (Weiner 2010) in 1948. A bibliography of "Machine Techniques in Office Operations and Information Processing" resources appeared in the October 1954 edition of *The American Archivist* (Smith 1954). The spread of automated technology in the workforce was becoming common place as secretaries typed up reports. As the new typed documents began to be accessioned into the archives, archivists began to realize that they could look to the creators of documents for pointers on how to use automation in their work. However, the next major development in automation for archives would not happen for another ten years.

At the 1964 meeting of the International Congress of Archives (ICA), it was discovered that the only country that was experimenting with large scale automation was the United States National Archives and Records Services (NARS). Automation was used for "warehouse controls and could be similarly used for locating files" (Fishbein 1981). Following this, the Library of Congress stepped up to the plate and developed SPINDEX (as well as the following versions known as SPINDEX II and SPINDEX III), a computer program. SPINDEX II was created in 1967 to "expand the original Library of Congress card-based system into a tape-based system" (Weiner 2010).

These large scale efforts in the federal governments' archives were needed because of the volume of content that was being accessioned into the archive every day. As well, the United States had not begun its' archival program until late in the country's history, so there was a lot of backlog that needed to be processed. As well, the federal archives also had the monetary funding in order to experiment with automation that smaller archives were not yet privy to. So how many archives were using automation in the late sixties? The answer was not heartening.

In 1968, Frank G. Burke sent out a survey to two hundred archives in the United States and Canada to discover the extent that automation was being utilized in processing and search retrieval. The report showed that only fifteen institutions were using "manual, mechanical, or electronic devices" (Burke 1968) in their processing. Fourteen of these archives were associated with large institutions. Seventy other archives planned to use "some sort of automated or manual control device in their daily operations" (Burke 1968). Ninety-five archives had no plans to purchase any computer equipment, but their parent institutions were already involved in using automation technology. Therefore, these archives were still capable of potentially having access to computers. This was a foreshadowing of one of the five major ways archives would access computer for their automation in the 1980s. The answer to what percentage of archives was using automation in the late sixties was only 7.5%. However, with 35% planning to implement automatic processing systems in the near future, automation was finding a stronghold.

With the advent of the 1970s, automation was finally on the international radar. Fishbein wrote that the reason why automation had not pick up prior to then was because archivists "assumed that the recording media (chiefly tape) would merely perform calculating and transcribing functions." Computers were believed to be unable to meet the needs of archival work. However, Canadian archivists were able to show that "automation saved considerable money and staff time" (Fishbein 1981). Recognition of the usage of automation for archival work was officially recognized in 1972 when the ICA Committee on Automation was formed. This committee sought to create a bibliography, a journal, and a training curriculum on automation techniques as well as other material related to archival automation (Fishbein 1981). Two of the books that were written were *Guidelines for Administering Machine-readable Archives* (1980) by M. H. Fishbein and *An Introduction to Archival Automation* (1981) by A.

Arad and M.E. Olsen. The PARADIGM system that was developed by the University of Illinois Archives went from punched cards to networked retrieval in 1976. It had been developed "to provide administrative and intellectual or subject control of holdings" (Maher 1984). This demonstrated that a push for more streamline techniques meant automation was continuously being developed as a worthy invention in the archives. Then in 1978, the Society of American Archivists (SAA) offered the first workshops in machine-readable records (Walch 1993). MAchine-Readable Cataloging (MARC) assisted in the standardization of field creation, but the difficulties inherent in applying a new electronic technology to a manual process persisted. An official training program from the main American archivist organized meant automation had gained legitimacy within the United States.

The main hindrance for automatic systems was that each archive had to create their own in-house automation system. William J. Maher explained that these unique systems had developed "Because the content and format of archival descriptive systems vary from repository to repository, and because each repository has different priorities, independent development of automated systems has been a common response." He listed five reasons (Maher 1984) why such diverse systems existed:

- The archive was part of a national system or database.
- The archive used an off-the-shelf software package (e.g. WordPerfect).
- The archive owned their own computer and their programs were staff developed.
- The archive shared computer resources with their parent institution and were therefore not in control of the hardware or software and also had to time-share the equipment with other departments.
- A mix of the two of the previous systems was most common.

The benefits of an in-house develop system however had several benefits (Maher 1984). First, the custom-tailored system met the individual needs of the archives. There was no struggling through an abundance of features that were never utilized or too complicated for a layman data entry person to learn. Second, if the archive could use their parent institution's computer equipment, the archive would not have to foot the bill of purchasing all the technology. For example, the cost of an integrated data entry and processing program in 1984 was about $700 (Polsson 2010) in United States' dollars. To appreciate this cost in its contemporary equivalence, this one software program would have cost $1,426.80 in 2009 (Friedman 2010). Finally, by having an-house created system, an archive could keep the start-up costs low and expand as needed instead of having to purchase a fully operational system at the beginning which they may not fully utilize.

The literature about archival automation is scarce from the early 1990s onward. However, observations can be applied to note changes in computer technology and the World Wide Web over the last twenty years. These changes include memory devices getting larger and cheaper per megabyte as they evolved across several formats from the eight inch floppy disc to digital video disc (DVD). As well, the creation of HyperText Markup Language (HTML) made it possible to connect webpages together. The development of browsers and advances in programming languages made it easier for casual users to create content. This content creation led to the rise of social networking. Many archives now have Twitter feeds and blogs where they can announce their projects and processes on their own websites. The speed of internet connections has increased from 14.4k to 16.63 megabit per second (C. Smith 2010). These speeds allow for more primary source documents and scans to be put online for research purposes. Archival management software packages have also emerged which is targeted specifically at archives so

that archivists no longer have to either write their own programs or struggle with getting a large commercial product like WordPerfect to meet their specific needs.

Reasons to Automate

The point about recent technology trends and its impact on archives is that storage space is cheap, broadband access makes interconnectivity between collections easier and with new archival specific software, it is easier than ever to process a collection and get it into researcher's hands. As well, "computers can speed up the processes of collection, handling and retrieval of information, and can also extend the range of information supply and use" (Cook, An introduction to archival automation 2010). Computer usage in an archive allows for "both the archivist's need to preserve original order and the researcher's desire for subject access" (Dürr 1984). Before computer technology was introduced to archives, it was impossible to file a collection under all the subjects which it should be filed under. Automation allowed an archivist to save time and money since the "rearrangement [of collections] is unnecessary and expensive" (Campbell 1967). If a collection could be indexed under several categories in a database, the researcher will be able to find their material and the archivist could skip most of the rearrangement process since it was not necessary to rearrange everything in order to make it accessible.

Campbell goes on to list a few tips on how automation can be implemented. First, she writes that the "cost of machine indexing can be kept within a reasonable range if no rearrangement of the materials is required" (Campbell 1967). Second, you must use a controlled vocabulary. A controlled vocabulary would make it possible to manage documents in a structured way that make them easier for retrieval since one only needs to know a few terms rather than having dozens or more terms to look under for every search inquiry. Third, in many

cases you can use high school graduates to index as long as you have a professional indexer to watch over them. She cautions against having more highly educated amateur indexers for fears that they may become interested in what they are indexing and will waste time reading the documents. Or worse, they may read into the documents something that is not there and connect them to other ideas. Her ideas can be argued with a classist attitude from the mid-1960s, but her points still stand that an archive can save money by using computers to index as well as employ paraprofessionals.

Reasons against Automation

Despite all the benefits of automation, historically archival professionals have been slow to implement automation techniques in their individual repositories. Archivists feared the "cost, obsolescence of software and hardware, and professional disagreements over priorities" (Maher 1984) which underline an ingrained misunderstanding of automation within the archival profession. Three years prior to Maher's paper, Fishbein had already published reports that automation was saving Canadian archivists money and staff time (M. H. Fishbein 1981). Therefore the fears of cost were unfounded overall. Yes, capital was required to invest in obtaining new systems, but the benefits eliminated this objection long-term.

In regards to software and hardware becoming obsolete very quickly, it is a necessary to remember that many document creators are using software and hardware tools in their content creation. If an archive cannot access, store, index, and retrieve this digitally created content, they are failing in their mission. These born-digital materials that never existed on paper needs to be accessible by archivists and researchers. However, there is hope because at places like Stanford University, located in California, has "opened a digital forensics laboratory" (Cohen 2010). FRED is short for "Forensic Recovery of Evidence Data" (Cohen 2010). FRED allows archivists

to access obsolete formats and recover it. While this solution is costly, it is one way to challenge the perception that computer technology is useless to an archive because it relies on hardware and software in order to access the data. The creation of digital data is not going to slow down, so archives must rise up to meet the challenge of handling it instead.

Professional attitudes about automation technology also need to be revamped. Rob Weiner says that "archivists have clung to a sort of pre-1950s mentality." While many archivists like to dream about handling ancient documents and preserving their content to be shared with researchers that come after them, the truth is that more documents are being created and put into archives than in the past. This increase in materials is because of better literacy so more people are capable of reading and writing, as well as more people being alive today, and the cost of document creation is lower. It costs very little to create a thousand documents on a computer as opposed to trying to purchase a thousand notebooks two hundred years ago. Archivists have also not been using computers for their full potential. The Washington University Special Collections Department used their computer only for correspondences while the "item-level finding aids, accession registers, and statistics were prepared, as they always had been, on a typewriter" (Prietto 1994). Other archives are still writing their finding aids by hand (Weiner 2010). In order to meet the needs of an information saturated and hungry society, archives will need to be able to grant flexible access to documents in a timely manner. This cannot be done using the former methods of pencil to paper.

For the technology wary archivist, it must be acknowledged that technology cannot solve all of an archive's problems. A report from the Council of Nova Scotia Archives (CNSA) adds a disclaimed that "While it is recognized that increased use of technology could improve organizational efficiency and provide the public with better access to archival records,

constraints to adopting new technology must be acknowledged" (Stevens 2009). However, the cautious archivist should not use the CNSA's concession as an excuse to not study the use of automation in their archives, but instead look at it as acknowledgement that change is often difficult.

Is Your Archive Ready to be Automated?

Before beginning to purchase any hardware and software needed to create an automated system, you should first do a survey to make sure that automation is the right solution for your archive. These questions are adapted from William J. Maher's article, *Administrating Archival Automation: Development of In-House System*.

- Do a study of the conditions of the archives.

    ○ Is your archive poorly run? Does it have a lot of unprocessed back logs?

- How is staff time being used?

- What do you want the automated system to do?

- Why do you want to automate?

    ○ What is the common level of description across all collections?

- Staff resources

    ○ How long can you devote to planning, implementing, maintenance?

The answers to these questions should be answered honestly. Automation is not a cheap process and should not be taken lightly. While it has numerous benefits explained throughout this paper, it still requires staff resources which are defined above to include the planning, implementing, and maintenance. As Frank G. Burke explained, "While automation can be a tremendous aid in broad areas of archival work, it cannot remedy program deficiencies resulting from inadequate or weak management." If your archive has an unreasonable amount of back logs

or is short on staff time, it is best advised to wait till your archive has time to fully understand how automation can benefit you before introducing yet another system that will take time away from obligations which are already behind.

Conclusion

Automation seems to be a no-brainer to the modern archivist or student. The old ways of writing finding aids by hand were labor intensive and required someone with legible handwriting. Researchers also add need for technology in the archives as they demand access to archival records which can be searched by subject or keyword. As well is the undeniable existence of thick back logs and no end in sight to the amount of document creation that is going on every day. An archivist must be able to handle both the past, present, and future records in a highly efficient way and make them available to researchers who can put them to use.

Embracing technology with all its advantages as well as all its flaws is not an easy process. Theodore Dürr wrote, "The new world of archival automation in 1984 will require people with courage, vision, and, above all, tenacity." It has been nearly twenty-seven years since then and the importance of an automatic system has increased more than Dürr could have imagined. The modern archivist does have the advantage of archival management software which is archive specific which was written to address issues unique to archiving, but it is not a wonder drug that can solve all problems.

Therefore, no matter how processing has been handled in your archive in the past, technology can most certainly streamline some aspects of your workflow. Automation is "essential for future growth, effectiveness, and efficiency" (Dürr 1984).

Annotated Bibliography

Arad, A., and M.E. Olsen. *An Introduction to Archival Automation* . Paris, France: Internation Council on Archives, Committee on Automation, 1981.

Burke, Frank G. "Report on a Survey of Automation Activities in Archives and Manuscript Repositories in the." *The American Archivist*, April 1968: 208-210.

Campbell, Rita R. "Automation and Information Retrieval in Archives—the Broad Concepts." *The American Archivist*, April 1967: 279-286.

Cohen, Patricia. *Fending Off Digital Decay, Bit by Bit.* March 15, 2010. http://www.nytimes.com/2010/03/16/books/16archive.html?pagewanted=2&_r=1 (accessed November 30, 2010).

Cook, Michael. "An introduction to archival automation." *Selected guidelines for the management of records and archives: a RAMP reader.* 2010. http://www.unesco.org/webworld/ramp/html/r9006e/r9006e0m.htm#An introduction to archival automation (accessed November 28, 2010).

- ○ Outlines five reasons why automation is now possible.
- ○ Guidelines on handling problems and the solutions in dealing with hardware, software, staff, costs, and systems planning.

Dürr, Theodore. "Some Thoughts and Designs about Archives and Automation, 1984." *The American Archivist*, Summer 1984: 271-289.

- ○ Includes automation wish list
- ○ Archival Automation Matrix figure which shows 1984 computer applications
- ○ Charts related to what databases and modules looked like
- ○ Discussion about archivists becoming information managers

Fishbein, M. H. *Automation in archives: a summary history.* Report in ADPA Journal, Liege, Belgium: Automatic Data Processing and Archives of the International Council on Archives, 1981.

Fishbein, Meyer H. *Guidelines for Administering Machine-readable Archives* . Washington, D.C.: International Council on Archives, Committee on Automation, 1980.

Fisher, Barbara, and Frank B. Evans. "Automation, Information, and the Administration of Archives and Manuscript Collections: Bibliographic Review." *The American Archivist*, 1967: 333-348.

- ○ Bibliography of the history of automation from 1945-1967

Friedman, S. Morgan. *The Inflation Calculator.* 2010. http://www.westegg.com/inflation/infl.cgi
(accessed December 1, 2010).

Library of Congress. *What is a MARC Record, and Why is it Important?* 2009.
http://www.loc.gov/marc/umb/um01to06.html (accessed November 28, 2010).

Maher, William J. "Administering Archival Automation: Development of In-House Systems."
The American Archivist, Fall 1984: 405-417.

- How automation is handled at different institutions
- How to create your own in-house system
  - Planning, development/implementation/maintenance, liabilities/limitations

Polsson, Ken. *Chronology of Personal Computer Software 1984.* September 09, 2010.
http://pctimeline.info/software/soft1984.htm (accessed December 1, 2010).

Prietto, Carole. "Automating the Archives: A Case." *The American Archivist*, Spring 1994: 364-373.

- From abstract: "This process consisted of a number of steps over a three-year period:
  evaluating the existing hardware and software, selecting a new database management
  package, installing and setting up the new software and training staff in its use,
  adding OCLC and NOTIS access to facilitate MARC-AMC cataloging, and, finally,
  adding our first MARC AMC records to a national bibliographic database."
- How to evaluate a database management system

Smith, Catharine. *Top 19 Countries With The Fastest Internet Connection Speeds Ranked By
Akamai.* November 18, 2010. http://www.huffingtonpost.com/2010/11/18/fastest-internet-connection-speed_n_783865.html#s184351 (accessed November 28, 2010).

Smith, Lester W. " Writings on Archives, Current Records, and Historical Manuscripts, July
1953-May 1954." *The American Archivist*, October 1954: 341-365.

Stevens, Amanda. "Archives Management Software Review Report." *Council of Nova Scotia
Archives.* March 9, 2009.
http://www.councilofnsarchives.ca/resources/SoftwareReview.htm (accessed November
28, 2010).

Walch, Victoria Irons. "Final Report: Automated Records and Techniques Curriculum
Development Project: Committee on Automated Records and Techniques." *The
American Archivist*, Summer 1993: 468-505.

- What to include in the curriculum to train archivists in automation

Wardrip-Fruin, Noah, and Nick Montfort, . *The New Media Reader.* Cambridge and London:
	The MIT Press, 2003.

Weiner, Rob. "Archives and Automation: Issues and Trends." *Texas Tech University Libraries.*
	March 25, 2010. http://thinktech.lib.ttu.edu/handle/2346/1527 (accessed November 28,
	2010).

- From abstract: "Topics include: resistance to technology and the need to educate
  about automation; the change in archival theory due to the information age; problems
  with technology use; the history of organizing archival records using automation; and
  the use of and problems with MARC (Machine Readable Cataloging)-AMC (Archival
  and Manuscripts Control) records for archives."
- Discusses archivists' reluctance to become computer literate