

Gregory P. Johnson. Quality or Quantity: Can Archivists Apply Minimal Processing to Electronic Records? A Master's Paper for the M.S. in L.S degree. April 2007. 36 pages. Advisor: Dr. Christopher Lee.

The study is an introduction into the debate of whether or not minimal processing, based upon the recent work by Mark Greene and Dennis Meissner, can be used in the processing of electronic records

With the recent publication of a "More Product, Less Process" approach by Mark Greene and Dennis Meissner, minimal processing of archival collections has received increased attention in the professional literature. Greene and Meissner posit that a change in archival philosophy toward processing collections at a less precise level will allow archivists to arrange analog collections at a faster pace and help to eliminate their processing backlog. Through an examination of the literature in the field and a comparison of the analog and electronic records media, I have determined that the processing of electronic records can be done according to the minimal processing protocols that are being examined for use in traditional archival collections.

Headings:

Archival description/Archives

Applications/Standardization

Electronic Records

QUALITY OR QUANTITY? CAN ARCHIVISTS APPLY MINIMAL PROCESSING
TO ELECTRONIC RECORDS

by
Gregory P. Johnson

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

April 9, 2007

Approved by

Dr. Christopher Lee

The recent publication of “More Product, Less Process” by Mark Greene and Dennis Meissner¹ posits that archivists need to start processing collections at a less precise level in order to address the overwhelming backlog of documents that continues to haunt modern archivists and their repositories. This use of minimal processing, either at a series or collection level, will also allow archivists the chance to create a catalog of what is housed in their repositories that has yet to be processed. According to Greene and Meissner, the former best practice of processing at the folder or, in rare cases, the item level has been one of the major causes of the backlog that archivists presently face. These backlogs are prohibiting the repositories from allowing public access to these collections, which is one of the major charges of an archive.

By processing at a less time consuming level, Greene and Meissner propose that archivists will be able to process at a faster pace, thus clearing out the unprocessed collections. According to the Association of Research Libraries Task Force on Special Collections, unprocessed collections made up 27%

¹ Greene, Mark, and Meissner, Dennis. "More Products, Less Process: Revamping Traditional Archival Processing", 240-241.

(manuscripts) to 46% (artifacts) of the collections housed in surveyed academic repositories in 2002.²

Currently archivists have an even larger problem: electronic records. Since electronic records are replacing paper-based documents at an alarming pace, a decision needs to be made quickly on how archivists are going to process them. Should archivists process these records on folder-by-folder basis, to be accessed through thorough and meticulous finding aids? Or, should archivists process them on a collection or series level, focusing their time on getting these collections out to the public?

The impact of the Greene-Meissner report on the ever-growing backlog of electronic documents and materials has yet to be studied. With the growing reliance on electronic records as a means of communication, what is being done to arrange and describe these records? From archivists and other information professionals, who need to discern methods to corral this ever-growing pool of documents, to citizens and researchers who would like to use the documents, this problem truly does have global implications.

The Greene-Meissner report studies how to alleviate the ever-growing backlog of analog collections, which are creating a “hidden archive” within archival repositories. Greene and Meissner posit that archivists need to follow the path set out by librarians, who also have had to wrestle with the problem of backlogs and have begun to solve this problem through the practice of cataloging

² Association of Research Libraries Task Force on Special Collection. “Hidden Collections, Scholarly Barriers: Creating Access to Unprocessed Special Collections Materials in North American Research Libraries”, 4.

library collections to the level of “good enough.”³ Greene and Meissner believe that by eliminating the non-essential tasks of processing, such as non-required re-foldering or excessive re-organization of files, archivists can create finding aids that are accurate enough to explain what is in the collection and will allow archivists to process at a much faster pace.

Greene and Meissner also posit that while the debate over processing our backlogs has been fierce over the past twenty years, it has been discussed in the wrong arena. While appraisal and acquisition have been the focus of the discussion, the authors believe that the debate should concentrate instead on processing. Once the archival community realizes that it needs to focus on processing in as much detail as appraisal, then it will be able to create new processing models and will hopefully be able to start to clean up the backlog of unprocessed collections.

Literature Review

This study is a preliminary study of electronic records arrangement. I will not focus on the topics of accessioning or description, as this has been touched on by other scholarly literature within the field.⁴

In this section, I will set a definition for traditional arrangement and describe how it has been traditionally done in analog collections.

What is Processing?

³ Greene, Mark, and Meissner, Dennis. "More Products, Less Process: Revamping Traditional Archival Processing", 255.

⁴ Adams, Margaret O'Neill. "Three Decades of Description and Reference Services for Electronic Records; Roth, James M. "Serving up EAD: An Exploratory Study on the Deployment and Utilization of Encoded Archival Description Finding Aids."

Currently, there is not a field-wide accepted definition of processing. The definition that I use for this paper comes from the Society of American Archivists (SAA) Glossary, which defines processing as:

The process of organizing materials with respect to their provenance and original order, to protect their context and to achieve physical or intellectual control over the materials⁵.

Arrangement of archival collections varies based upon the size, rarity, and usefulness of the collection. Collections deemed to be of less research importance may only require re-folding, re-boxing and basic preservation such as removal of fasteners, photocopying of frail paper, photocopy and removal of newsprint, etc. before they are ready to be used by patrons. Other collections, such as those that contain rare or fragile items, may require extensive preservation and the introduction of security measures to protect the collection. These works require a vastly greater amount of processing time before the collection is then able to be used by a repository's patrons.

Desnovers states that a repository should process its collections to serve three distinct groups: users, donors, and the archivists themselves.⁶ All three of these groups have different agendas, uses, and desires for collections. Users are looking for an easily searchable and quickly processed collection that will ease them in their research. Donors are looking to get maximum exposure for their collection and to have it processed quickly yet descriptively while respecting the tenets of the donor agreement. The archivist, on the other hand, has to try to

⁵ Society of American Archivists Glossary.
http://www.archivists.org/glossary/term_details.asp?DefinitionKey=294 (accessed April 8, 2007).

⁶ Desnovers, Megan Floyd. "When is a Collection Processed?", 309-310.

appease both groups while looking out for the best interest of the collection itself during processing.⁷

Desnoyers describes four “continua” of processing: the arrangement, preservation, description, and screening continua. I will focus on the arrangement continuum. According to Desnoyers arrangement is determined on a collection-by-collection basis by the archivist. This step of processing is completed when the archivist has determined that the collection is arranged and titled in a manner so that the researcher and reference archivist can understand its order.⁸

According to Greene and Meissner, archival processing is, at its core, “housekeeping,”¹² the act of re-folding, re-labeling, re-boxing, preserving (and conservation when necessary), and then describing what is in a collection. In their empirical study on the estimated costs of processing collections, Abraham, Baizarini, and Frantilla conclude that standard expense for the archival processing of a common collection is approximately \$200 to \$250 per cubic foot. For comparison sake, the authors state that the expense to a repository to

⁷ Ibid., 309-313.

⁸ Ibid., 313-315.

¹² Greene, Mark, and Meissner, Dennis. "More Products, Less Process: Revamping Traditional Archival Processing", 241

catalog and process a similar sized collection of library assets is estimated to range between \$4.58 and \$13.21 per cubic foot.¹³

Another major cause of the bottleneck has been description. Although Greene and Meissner believe that processing at a less precise level, usually at the series level or above, will help speed up the process, they also believe that this savings of time could lead archival professionals to allocate this extra time to extensive description and creating of finding aids.¹⁴ The description phase of the process for most repositories currently follows the near “one size fits all” descriptive models set forth by MARC (Machine Readable Catalog Record) and APPM (Archives, Personal Papers, and Manuscripts) formulas.¹⁵

Although archivists have longed for a form or template to ease and standardize the processing of paper based collections, Greene and Meissner believe that the archival world has been “inconsistent and even schizophrenic about defining the parameters of ‘processing’.”¹⁶ With all of this confusion within the archival world about the processing of analog collections, would it not be wise to find a process that could work for electronic records prior to the inevitable bottleneck? Greene and Meissner believe that “our professional fastidiousness, our reluctance to be perceived as sloppy or uncaring by users and others has encouraged a widespread fixation on tasks that do not need to be performed.”¹⁷

¹³ Abraham, T., Baizarini, Stephen E., and Frantilla, Ann. "What is Backlog is Prologue: A Measurement of Archival Processing", 32-33.

¹⁴ Greene, Mark, and Meissner, Dennis. "More Products, Less Process: Revamping Traditional Archival Processing", 215.

¹⁵ *Ibid.*, 217.

¹⁶ *Ibid.*, 227.

¹⁷ *Ibid.*, 241.

Another issue with the use of a template for processing collections is that all collections do not require the same level of attention to adequately process them. According to Desnovers, every collection has a different point at which it can be deemed to be “processed.” In attempting to process all collections at one level, the archival community creates “a small number of beautifully processed collections” and leads to “an extensive backlog of collections that are closed.”¹⁸ Desnovers believes that flexibility in processing is necessary to try to minimize the amount of collections archival repositories currently have unprocessed and unavailable for patron use.

Minimal Processing

With the maintenance of original order as one of the paramount goals for an archivist, a natural question that should be asked is why do archivists continue to dissect and dismember archival collections for the sake of making them more clear to researchers and in return elongate the processing stage?

In 2005, Greene and Meissner finalized and published the findings of a comprehensive study on the strategies and theories of archival professionals in relation to the processing and description of archival collections in the *American Archivist*. They sent a five-page survey instrument to all 1100 members of the Society of American Archivists Manuscript Repository and Description section to gauge how the archival community handles processing. They asked participants to describe the staff level and budget of their repository; the current number of unprocessed collections; how the repository preserves, appraises, and describes

¹⁸ Desnovers, Megan Floyd. “When is a Collection Processed?”, 315.

their collections; and the processing benchmarks of the repository. The results of the finding were based upon the 100 surveys that were returned.¹⁹

According to the data from the survey, most of the repositories that replied process their collections at a folder or item level. Most repositories re-folder their collections, weed out duplicated from within those folders, and arrange the items within the folders themselves. The Greene-Meissner report also showed that many repositories are performing preservation tasks below the series level: removal of staples and paper clips, segregating newspapers from the rest of the collection, photocopying newspaper and onionskin clippings, as well as other tasks of creating an acid- and contaminant-free environment for the collection.²⁰

Interestingly, the report revealed that 63% of the responders stated that 30-50% of their collections are unprocessed, 67% of responders stated that their backlog of documents has grown over the past decade, and 52% of repositories required more than 36 months processing a collection.²¹ A large portion of the reporting repositories also reported that they have upset or angered patrons, researchers, contributors and/or donors due to unprocessed collections and collections that are not processed in a timely manner.²²

The main conclusion of their work is that the archival community is squandering far too much time and resources processing collections at a level that is unnecessarily precise for the usefulness of the collection. The authors also contend that this misuse of time, energy, space, and money has been one of

¹⁹ Greene, Mark, and Meissner, Dennis. "More Products, Less Process: Revamping Traditional Archival Processing", 210.

²⁰ Ibid., 259.

²¹ Ibid., 258-263.

²² Ibid., 262-263.

the main culprits behind the growing backlog of collections yet to be processed. Greene and Meissner believe that the archival community needs to follow the lead set forth by the library community by rethinking processing specificity. The library community has trimmed its backlog of un-cataloged items, even while budgets have been slashed and staffing has been cut at most institutions, through the use of minimal processing. This lower level of precision in the cataloging of their books has sped up the process of cataloging while relieving catalogers of non-essential work. Greene and Meissner also advocate a repository-by-repository, user-centered finding aid policy, tailoring the finding aid to the prospective users, instead of a generic template.²³

Greene and Meissner have targeted as one of the main sources of the current bottleneck of unprocessed collections. In some repositories, newly accessioned collections can sit for decades prior to processing. According to Greene and Meissner, “processing is not keeping up with acquisitions, and has not been for decades, resulting in massive backlogs or inaccessible collections.”²⁴ The authors further their argument by revealing that 34% of the repositories surveyed claimed that more than half of their collections were unprocessed and unavailable to the public.²⁵ Greene and Meissner posit that what is necessary process more collections is a change in thinking within the archival community, toward a “minimal amount of steps necessary” to preserve and sufficiently describe the materials.²⁶

²³ Ibid., 216.

²⁴ Ibid., 208.

²⁵ Ibid., 210.

²⁶ Ibid., 212-213.

Some institutions are beginning to put the Greene-Meissner processing ideas into practice. Weideman, in surveying the collections at her home repository of Yale University, agrees with the ideas put forth by Greene and Meissner and has been practicing some form of minimal processing for five years. Weideman begins the processing of her repository's collections at the accessioning phase, creating catalog records with a minimal amount of description. These descriptions, which are usually done at the box level, allow patrons and reference archivists a glimpse into the collections that have yet to be processed, thusly eliminating a large portion of the repository's hidden collections.²⁷ In some instances, donors are asked to write a simple description of the collections they are donating, which will usually become the series description for the unprocessed collections, and donors are now being notified of the trend towards minimal processing at the time of the appraisal.²⁸

McCrea of the University of Montana agrees with Weideman's ideas and practices. According to McCrea, the University of Montana at Missoula has is experimenting with minimal processing, including the creation of collection level records for the repository's unprocessed collections.²⁹

Frank Boles, in an article originally written about preserving original order within paper collections, argues that the documents that are to be processed should follow the simplest form of processing that allows the most information to be given to the user. This theory, which he calls "simple usability," grants the

²⁷ Weideman, Christine. "Accessioning as Processing", 275.

²⁸ *Ibid.*, 277 and 281.

²⁹ McCrea, Donna E. "Getting More for Less: Testing a New Processing Model at the University of Montana", 290.

archivist time to create easily searchable metadata for the collection during time that previously had been used for traditional processing.³⁰

One drawback to the use of minimal processing in both analog and digital formats is that the emphasis for discovery is now placed on researchers and reference workers, not archivists, due to the limited inspection of the records by the processing archivist and creation of less detailed finding aids.³¹

For reference archivists dealing with collections that have been processed minimally, more boxes will now need to be retrieved for patrons in order to find documents that had been previously arranged together.³²

McCrea believes that these compromises of convenience will be repaid by access to previously unavailable collections. Donors will also be happy with the pace that their collections are processed, and the use of minimal processing may actually being useable as a selling point to potential donors due to its rapid results.³³

The unifying idea that comes out of the literature on minimal processing is that archivists who use this technique are not “boldly going where no one has gone before, but re-discovering a fundamental principle of archival processing.”³⁴ Every repository has its own unique collections and its own financial-, policy-, and staff-related realities, so a one-size-fits-all processing policy cannot be

³⁰ Boles, F. "Disrespecting Original Order", 30-31.

³¹ McCrea, Donna E. "Getting More for Less: Testing a New Processing Model at the University of Montana.", 289; Weideman, Christine. "Accessioning as Processing", 283; and Hyry, Tom. "More for Less in Archives: The Greene/Meissner Approach at Work at Yale", 8.

³² Weideman, Christine. "Accessioning as Processing", 282-283.

³³ McCrea, Donna E. "Getting More for Less: Testing a New Processing Model at the University of Montana", 289.

³⁴ *Ibid.*, 290.

universally applied to all collections at all repositories. According to Weideman, if repositories continue to experiment with alternative processing schemes and disseminate their results within the community, processing times can be improved for all repositories.³⁵

What are Electronic Records?

As with the idea of processing, there is not a community-wide accepted definition for what an electronic record is. I will begin the discussion with the definition put forth by the Society of American Archivists glossary:

Data or information that has been captured and fixed for storage and manipulation in an automated system and that requires the use of the system to render it intelligible by a person.

Electronic records can be made up of various formats or program types. Thibodeau, Moore, and Baru from the National Archives Records Administration (NARA) believe that electronic records belong in two “superclasses”: digital objects and records.³⁶ This is an important concept since both “superclasses” have their own characteristics and problems associated with them, such as the need to preserve digital objects quickly so that they won’t be lost but still trying to maintain their authenticity as a record. As digital objects, electronic records can be further broken down by software or operating system type. As records, they can be further broken down by traditional archival fonds and groupings.³⁷ Each superclass brings with it a special set of problems, with the archivist having to appraise, process, and accession these electronic records as both digital objects

³⁵ Weideman, Christine. "Accessioning as Processing", 283.

³⁶ Thibodeau, Kenneth., Reagan Moore., and Chaitanya Baru. "Persistent Object Preservation: Advanced Computing Infrastructure for Digital Preservation", 114.

³⁷ Ibid., 114.

and as records. When processing electronic records, it is important to recognize that any undocumented change to their format (i.e. change of software used or creating a paper copy) could damage their authenticity.

Processing Electronic Records

I will begin this discussion by looking at how two large governmental agencies, U.S. National Archives and Records Administration (NARA) and the Paradigm (Personal ARchives Accessible in DIGital Media) project in the United Kingdom (UK). Although their budgets, as well as the scope of their electronic records problems, are far larger than most repositories, these two are a good place to begin the discussion what is currently being done to try to solve the electronic records dilemma.

NARA

As one of the largest and oldest repositories for electronic records in the United States, NARA has worked for more than three decades accessioning, processing, and preserving electronic records. According to Ambacher, this has been done by following the regulations and guidelines that have been set forth by the various other media, such as film and paper, which are processed at NARA.³⁸ These guidelines call for a dedicated staff that works solely on the accessioning, preservation, description, and reference services of electronic records.³⁹ Ambacher states that, through this use of a dedicated staff, NARA has been able to keep a reasonable level of backlog, while still allowing other specialized staff time to study future procedures and standards for the processing of electronic

³⁸ Ambacher, Bruce. "The Evolution of Processing Procedures for Electronic Records", 59.

³⁹ Ibid., 43.

records and “ensures that an appropriate sense of a mission exists” among the staff.⁴⁰

Currently, NARA begins the processing of electronic records by making duplicates of electronic records that are injected and then placing them on media that has been certified by the government to safely house these documents. This policy is a byproduct of past records failures in which electronic records that were brought over were either lost or damaged during the inject process. These forms of media have changed over the past two decades, with the current certified media types including tape backups, CD-ROM's, digital linear tape, and open reel magnetic tape.⁴¹

The next major step in this process has been the standardizing of documents injected either into American Standard Code Information Interchange (ASCII) or Extended Binary Code Decimal Interchange Code (EBCDIC). These two file formats allow the files that have been injected to be easily accessible across various platforms and with multiple forms of software.⁴² Unfortunately, not all types of files can be changed into these formats. Files such as web based documents, email, GIS photographs, or other newer technologies that contain more complicated material will require newer programming language or the possible use of software emulators to access these records.⁴³

⁴⁰ Ibid., 59-60.

⁴¹ Ibid., 57.

⁴² Ibid., 58.

⁴³ Ibid., 58.

Due to the massive amount of documents created by the US government, which was estimated in 1993 to include 10 to 40 million email messages a year,⁴⁴ NARA is currently at the forefront of studying how archivists will be able to handle large amounts of electronic records. Thibodeau asserts that even with the lower estimate of 10 million messages, it would be impossible to perform even the most basic preservation on every document in the current NARA archival system.⁴⁵ Over the past decade, NARA has begun working with the various federal organizations whose documents NARA will be processing, such as the National Science Foundation (NSF), the Department of Defense (DOD), and the US Patent and Trademark Office, to create new records retention programs that will allow the government to process and use the many millions of records that are created yearly.⁴⁶ Due to the scale of the government bureaucracy, most of these projects aim to create automated systems that will be able to seamlessly process the massive amount of government work created with little or no preservation or processing. The mission statement for the Electronic Records Archives (ERA) program clearly summarizes the goals of NARA when it comes to electronic records. They are looking for a system that “will authentically preserve and provide access to any kind of electronic record, free from dependency on any specific hardware or software, enabling NARA to carry out its mission in the future.”⁴⁷

⁴⁴ Thibodeau, Kenneth. "Building the Future: The Electronic Records Archives Program", 92.

⁴⁵ Ibid., 92.

⁴⁶ Ibid., 93-97.

⁴⁷ Ibid., 102.

Paradigm

A study on the processing of electronic records collections is also currently underway in the UK. Martin and Thomas of Paradigm, a UK-based initiative investigating how archivists can process electronic collections, have begun using the personal papers of contemporary English politicians as a test bed for new technologies. One of the major issues that they are coming up against is the expense of the project, as well as issues of how to save the various forms of information and allow them to be accessed by patrons.⁴⁸

This project was created as a vehicle to begin the communal dialogue regarding how archivists will process electronic records. The Paradigm project is exploring the many issues that are inherent in the processing of electronic records, which include how to properly ingest them, if it is possible to maintain their authenticity and integrity when the files are ingested, how to create proper metadata for these records, and, finally, how these records will be used by the public after they have been processed. Ideas that have come from this project mirror those that are in place at NARA, which include designating a single form of electronic record, such as PDF or some form of an open source document, as the standard for all records. This would allow easier search-ability of electronic records due to the sameness of media, as well as offer easy ingest of records. The major problem that has been noted by this project is the expense of electronic records projects, such as Lockheed Martin's work with NARA, the

⁴⁸ Martin, J. and T., Susan. "Using the Papers of Contemporary British Politicians as a Testbed for the Preservation of Digital Personal Archives", 47.

Espida project in Glasgow, Scotland, or the creation of the National Academy Archives in London.⁴⁹

Grimard agrees with the idea that the only way to maintain the usefulness of electronic records is to change the medium of the record, creating a standard format for electronic records and converting all present electronic records into this new format. Only through these forms of preservation and processing would the record be able to maintain its usefulness in society.⁵⁰

Bantin posits that our traditional forms of archival processing and description, which have their roots in the archival principles of the first French Republic, are not adequate for electronic records. Due to the complexity of the records themselves, as well as the fluidity in which they can be changed or even destroyed, a major challenge to the archival community is defining and categorizing the provenance of electronic records.⁵¹ Bantin, as well as Cunningham, believe that this process needs to be “pre-custodial” in nature, i.e. that the archivist needs to be involved as soon as records are created, or even earlier, at the point of system design.⁵² This is necessary to insure the integrity of the linkages between documents as well as maintain format types.

Stollar and Keihne describe the ability to allow users the chance to arrange the items in a manner that they deem fit through the use of metadata and any number of applications, in this instance, DSpace. The authors describe

⁴⁹ Ibid., 47-49.

⁵⁰ Grimard, Jacques. “Maintaining the Long Term Preservation of Electronic Archives, or Preserving the Medium and the Message.” 156.

⁵¹ Bantin, Phillip. “Strategies For Managing Electronic Records: A New Archival Paradigm?”, 25.

⁵² Cunningham, Adrian. “Waiting For the Ghost Train: Strategies for Managing Electronic Personal Records Before It Is Too Late.” 57.

how the collection has been processed using traditional archival hierarchies, such as being divided into communities (archival “fonds”), sub-communities (series and subseries) and so on.⁵³ Even though this work has been processed in this manner, the data is stored and searchable through DSpace, which allows users to manipulate the data from this collection without damaging the collection or the individual items.⁵⁴

Hyry and Onuf concur on the usefulness of metadata. They state that an advantage of some forms of electronic records is that they already have their own searchable metadata when they are accessioned into the collection, thusly allowing easier processing and description. Email messages and some word processing documents, for example, already have metadata at the time of ingest. This can support searching and organization by that metadata that was created at the beginning of the life of the record, such as sender/recipient name or message title in email documents or the metadata sections of most word processing document.⁵⁵ Metadata within documents unfortunately also has a drawback: forms of technology that were used in the creation of these records, either software or hardware, are often necessary to access the original document, which could be cost prohibitive for a repository.⁵⁶

Chapman also believes that with the massive influx of electronic documents, processing at a higher level of aggregation, usually at the series or

⁵³ Stollar, Catherine, and Thomas Kiehne. "Guarding the Guards: Archiving the Electronic Records of Hypertext Author Michael Joyce", *Ibid.*, 4.

⁵⁴ *Ibid.*, 4-5.

⁵⁵ Hyry, Tom., and Rachel Onuf. "The Personality of Electronic Records: The Impact of New Information Technology on Personal Papers", 38-39.

⁵⁶ *Ibid.*, 42.

even collection level and using descriptive metadata, is necessary to allow smooth navigation and use. This metadata would have to include enough information to allow the user the ability to discern whether or not the information they desire can be found within the record or collection. Chapman believes that this level of metadata description, which would be akin to road signs on a highway, will allow the user to easily navigate through a collection at this higher level through well-defined metadata, rather than using traditional archival finding aids. This information would also allow the metadata to be placed into an easily accessible open database, which would allow remote users the ability to accurately search a larger set of collections at a single time.⁵⁷

According to McInnes, although the records would still be physically housed at the archive in whatever medium that archive would choose to use, the metadata created during the initial steps of the records processing would be used for access to the record. Now, “metadata elements transfer part of the responsibility for intellectual control to the system itself,”⁵⁸ with the rest of the responsibility falling to the processing archivist. Another possibility is that the records may not even be transferred to a server or storage medium at the archive; they could stay with the records creator.⁵⁹

A major issue that is present within the discussion of processing of electronic records is the need to maintain the authenticity of the original record during the various parts of the processing of the record, even if there needs to be a change in file format or operating system in order to access the record. Duranti

⁵⁷ Chapman, Ann. "Collection-level Description: Joining Up the Domains", 151.

⁵⁸ McInnes, Sally. "Electronic Records: The New Archival Frontier?", 216.

⁵⁹ Ibid., 216.

and McNeil put forward the idea that the physical form of the record, which could be made up of font types and sizes, symbols, document seals, etc., is present in electronic documents just as it is in analog. During processing, any change in these characteristics, even if they are caused in the migration process to a new system or file type, could be damaging to the original form and content of the record, and thusly the authenticity as well.⁶⁰

The “How Much Information” study estimated that email alone generates about 400,000 terabytes of new information each year worldwide. Instant messaging adds 274 terabytes each year along with the more than 170 terabytes added by the World Wide Web. This new information is greater than what is housed in the Library of Congress and three times the amount of new information from 2000.⁶¹ These figures do not include peer-to-peer file sharing or any other new form of information transfer or documents that have been created on personal or work computers. With this massive influx of information, archivists and records managers will need to find a way to get a “snapshot” of the information housed in large files and folders. It would be a nearly insurmountable task for archival professionals to re-folder every single folder or grouping of information that is collected, personally read every single email message and attachment, or weed through every single message to find all of the authors.

With electronic records, archivists must once again look at the utilitarian aspects of minimal processing over the classical theories of full processing. With the massive amount of information that is being sent in various electronic formats

⁶⁰ Duranti, Luciana. and Heather McNeill. "The Protection of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project", 50.

⁶¹ Berkeley School of Information, "How Much Information: 2003", 2.

in a plethora of operating systems, some on proprietary software, it is not possible to separate all email attachments from their host messages or accurately save a copy of all hyperlinked media from within word processing documents as would have been done with their paper counterparts, stapled or paper clipped documents. With software being updated often, in some cases monthly, any record older than a year could possibly be lost due to updated software. Thus, a considerable backlog of electronic records would most likely mean a loss of that information. Documents that were written in the early versions of Microsoft Word or Corel Word Perfect are not always compatible with the newer versions and older PDF files cannot always be read on newer versions of Adobe Acrobat, to cite two examples. These files are possibly lost, unless archivists can find and install older versions of these programs.

Comparison of Paper and Electronic Records Processing

Menkus states that “people will create those records that they want to create and will retain those records that they want to retain and will destroy almost any records that they feel like destroying.”⁶² This is just one of the many ways that paper-based records are similar to electronic records. Although Pyatt believes that the traditional forms of archival processing are not applicable to the processing of electronic records,⁶³ the potential transfer of techniques requires further investigation.

⁶² Menkus, Belden. “Defining Electronic Records Management”, 42.

⁶³ Pyatt, Timothy. “Acquisitions: Assessment, Scheduling, and Transfer of Public Affairs Records”, 1.

Electronic Record Archival Collections⁶⁴	Analog Archival Collections
1. Receive and identify physical media	1. Appraisal and accessioning of records ⁶⁶
2. Catalog the physical media	2. Create and initial catalog record ⁶⁷
3. Copy files to newer media	3. Perform basic preservation ⁶⁸
4. Perform initial file processing	4. Survey the collection to develop fonds and groupings for what is present ⁶⁹
5. Create an item-level index of all recovered files	5. Process the collection and create finding aid ⁷⁰
6. Create/process working copies of all files, retaining the original bitstream copies	

Figure 1. Comparison of processing procedures between electronic and analog records.

As seen in figure 1, many of the traditional archival principles can be translated into the arrangement of electronic records. This begins with the initial process of receiving and identifying the collection and creating an initial catalog record. As seen in the articles by McCrea and Weideman, archivists are beginning processing collections as early as the appraisal process.⁷¹ Stollar and Kiehne believe that this is happening with electronic records as well.⁷²

Glick and Wilczek describe the process of ingest in terms normally used for traditional analog processing. They believe that the process of ingesting documents closely mirrors the tasks required in processing an analog collection,

⁶⁴ This entire section of the figure is taken from Stollar, Catherine, and Thomas Kiehne. "Guarding the Guards: Archiving the Electronic Records of Hypertext Author Michael Joyce", 2-3. These ideas are also covered in Glick, Kevin L., and Wilczek, Eliot. "Changes in Acquisition: A Guide to the Ingest of Electronic Records," 3-6.

⁶⁶ Roe, Kathleen. *Arranging and Describing Archives and Manuscripts*, 8.

⁶⁷ Weideman, Christine. "Accessioning as Processing." 276-277.

⁶⁸ *Ibid.*, 278.

⁶⁹ Desnovers, Megan Floyd. "When is a Collection Processed?" 312-313.

⁷⁰ Desnovers, Megan Floyd. "When is a Collection Processed?" 312 and 322.

⁷¹ McCrea, Donna E. "Getting More for Less: Testing a New Processing Model at the University of Montana", 290; and Weideman, Christine. "Accessioning as Processing." 277.

⁷² Stollar, Catherine, and Thomas Kiehne. "Guarding the Guards: Archiving the Electronic Records of Hypertext Author Michael Joyce", 2-3.

with the archive following many of the same guidelines and steps.⁷³ In this ingest process, the archivist appraises the usefulness of the records, decides if his repository is able to handle the records, and works with the donor on a donor agreement and how the donor would like the records to be used. All of these steps have their origins in traditional processing techniques.

One area where ingest does differ great from traditional processing regards the state of the records at ingest. Traditional archival collections are accessioned into the archive in various states of order, from organized to completely un-organized. With electronic records, Glick and Wilczek state the organization of the records is now the responsibility of the producer, not the archivist, due to the fragility of the electronic records.⁷⁴ This includes ensuring that the records are in compliance with a repository's ingest guidelines for format. Although this aspect of electronic records ingest puts the onus on the producer to create a logical order for the records, it also allows the archive to maintain a level of authenticity within the collection. According to MacNeil, if the producer is creating and maintaining their records in compliance with a repository's ingest guidelines, the producer will be able to ensure the authenticity of their records because the archive will not have to make any format changes to the records.⁷⁵

The next steps in traditional processing, which include surveying the contents of the collection and performing basic preservation, also mirrors the process set forth by Stollar and Kiehne of copying files to a standard medium and

⁷³ Glick, Kevin, and Eliot Wilczek. "Ingest Guide for University Electronic Records," 7.

⁷⁴ *Ibid.*, 5.

⁷⁵ MacNeil, Heather. "Providing Grounds for Trust II: The Findings of the Authenticity Task Force of InterPARES," 47.

beginning basic processing. With both models, the next step would be to process the collection and create some form of finding aid. The final step of the process for electronic records is the creation of working copies of files.

Electronic Records Archival Collections	Analog Archival Collections
Determine provenance ⁷⁸	Same ⁷⁹
Deleting duplicate files ⁸⁰	Weeding duplicates ⁸¹
Scanning newly ingested documents for viruses or other attachments that can damage your repositories infrastructure ⁸²	Removing and copying, when possible, all damaged or infested sections of a collection ⁸³
Watermarking valuable documents/transfer damaged documents to new formats ⁸⁴	Replacing valuable or damaged documents with photocopies ⁸⁵

Figure 2. Comparison of processing task between electronic and analog records.

When looking at the tasks involved in processing more closely, there are more similarities between the two formats. The similarities between the two types of media begin in the accession (or ingest) stage of processing. As stated by Glick and Wilczek, traditional archival skills are required to work in conjunction with new technical archival skills in the ingest stage of processing.⁸⁶ In addition to technical knowledge, such as format types, server capabilities, and knowledge

⁷⁸ Stollar, Catherine, and Thomas Kiehne. "Guarding the Guards: Archiving the Electronic Records of Hypertext Author Michael Joyce", 2-3

⁷⁹ Desnovers, Megan Floyd. "When is a Collection Processed?" 316-317.

⁸⁰ Stollar, Catherine, and Thomas Kiehne. "Guarding the Guards: Archiving the Electronic Records of Hypertext Author Michael Joyce", 2-3

⁸¹ Desnovers, Megan Floyd. "When is a Collection Processed?" 316-317.

⁸² "Portico Technical Overview: A Format-Registry-Based Automated Workflow for the Ingest and Preservation of Electronic Journals," Slides 7-9;

⁸³ Desnovers, Megan Floyd. "When is a Collection Processed?" 316-317.

⁸⁴ Stollar, Catherine, and Thomas Kiehne. "Guarding the Guards: Archiving the Electronic Records of Hypertext Author Michael Joyce", 2-3

⁸⁵ Desnovers, Megan Floyd. "When is a Collection Processed?" 316-317.

⁸⁶ Stollar, Catherine, and Thomas Kiehne. "Guarding the Guards: Archiving the Electronic Records of Hypertext Author Michael Joyce", 5-6.

of various programming languages, archivists are still going to be required to understand how copyright law affects their choices for ingest, how to properly use metadata for new collections, and how to accurately process and describe the collections that are being ingested.

A common task in the appraisal/ingest of all records is the determination of origins. For analog records, the provenance and original order of the records is essential to be able to properly process a collection of records. With electronic records, this process is equally important. Hedstrom states that provenance is vital to understanding the “nature, timing, and authenticity” of electronic records.⁸⁷

Other steps in the two processes that are alike include the need to scan collections for viruses (electronic) or pests/mold (analog), weeding of duplicate files (both), and watermarking (electronic) or making photocopies (analog) of valuable documents. Although the actions may be different between the two types of collections, the processing theory is the same.

As noted earlier, most of the tasks required in processing electronic records have their roots in tasks that are already required during analog processing. It is the massive scope of electronic records, however, that causes the divide between the two. It is not possible to handle terabytes of information with the same level of precision that is to be expected of an archivist working on smaller analog collections. I do not believe that there is a single model that will work for everyone.

⁸⁷ Hedstrom, Margaret. "Descriptive Practices for Electronic Records: Deciding What Is Essential and Imagining What Is Possible", 56.

Summary and Conclusions

Electronic records and how to process, describe, and preserve them, will be one of the foremost conundrums faced by archivists in the 21st century. Since the records themselves, as stated previously by Thibodeau, Moore, and Baru, belong to two classes of classification, the question of how archivists should process these collections has to be addressed. Should archivists process these records in line with analog processing procedures? Or should they process them as digital objects and create new procedures such as NARA and Paradigm?

Cook, through the lens of post-modernism, states that the field is on the cusp of a major paradigm shift in archival thought, switching from worrying about the process to worrying about the product. Cook borrows the words of Thomas Kuhn from *The Structure of Scientific Revolutions* when putting forth the idea that this paradigm shift is caused by archivists having to completely rethink how they process collections in order to adapt to processing electronic records.⁸⁸ The processing of electronic records requires archivists to break away from century-old tenets of archival thought, with the archivist now most likely being required to change the physical integrity of the record in order to preserve it. In order to properly house, describe, and preserve these new records, an archivist, according to Cook, must “recopy ...its structure and functionality ... into new software every few years.”⁸⁹

Wallace also believes that our current practices of processing, which are based on the traditional principles of processing paper-based collections, will not

⁸⁸ Cook, Terry. “Archival Science and Postmodernism: New Formulations For Old Concepts.” 4-5.

⁸⁹ Ibid., 20.

be adequate for processing electronic records. Wallace states that “electronic records cannot be left to sit on shelves for years before being processed. New archival organizational structures must be created to ensure that records can be maintained in a useful form.”⁹⁰ This new organizational structure could be minimal processing.

In the literature on electronic records, one major idea that resonated was the need for archivists to become savvier in the use of new information and communication technologies. The authors called for archivists to develop “a thorough grounding in the various operating systems”⁹¹ and the need to gain better “knowledge of file transfer protocols.”⁹²

While I completely agree that archivists are not going to overcome the inherent issues with the processing of electronic records without further mastering the technical aspects of these records, I believe that archivists cannot forget the foundations of the field which are based in analog record processing. Although the formats of the collections that are being processed are changing daily from paper to electronic records, the core principles of how to process them are not. Without a firm grasp of basic archival processing principles, archivists will not be able to properly process electronic records. As Cox surmises:

We have made this era more confusing and more challenging by filling it with the promises of technological solutions rather than realities. It is a time in which we have become deluged with what Langdon Winner has called 'mythinformation,' the

⁹⁰ Wallace, David A. "Managing the Present: Metadata as Archival Description", 12.

⁹¹ Stollar, Catherine, and Thomas Kiehne. "Guarding the Guards: Archiving the Electronic Records of Hypertext Author Michael Joyce", 4.

⁹² Pyatt, Timothy. "Acquisitions: Assessment, Scheduling, and Transfer of Public Affairs Records." 3.

'almost religious conviction that a widespread adoption of computers and communications systems, along with easy access to electronic information, will automatically produce a better world for human living."⁹³

Cox posits that one of the causes for our issues is the community's continual desire to create more elaborate descriptive techniques, or as he states, "for the past 20 years, North American archivists have been up to their noses in their navels in developing either elaborate bibliographic (the USA) or archival control (the Canadian approach)."⁹⁴ Cox further posits that by the time that archivists have figured out how to process and describe electronic records and the World Wide Web, another format will most likely be upon us and all of those previous records will most likely be lost.

The literature on processing paper based archival collections is quite vast, with a smaller, but yet still impressive and growing subset of that literature focusing on the idea of minimal processing. With the recent work by Greene and Meissner, this idea has been brought to the forefront of archival processing discussions. Currently, there is no literature focusing on how archivists can use these protocols with electronic records.

One idea that seems to hold true in almost every article that I have found on minimal processing is this: there is not, and cannot feasibly be, a "one size fits all" guide for processing every type of archival collections at all repositories. It is likely that a large portion of local and community archives may not take steps to deal with the problem of how to process electronic records until well after

⁹³ Cox, Richard J. "Access in the Digital Information Age and the Archival Mission: The United States", 28.

⁹⁴ Ibid., 33.

some of the larger institutions have “solved” this problem. Others may continue to save electronic records onto portable devices, such as CD’s, DVD’s, flash disks, or other portable storage options.

The ideas of processing collections at a less precise level and limiting description of archival collections to only what is necessary can be transferred from analog collections directly to electronic records. Electronic records are far more fragile than their paper based counterparts, and leaving them un-processed while an archivist creates a long and eloquent description endangers the record.

For the past two decades, the discussion of electronic records has influenced the writing, thinking, and discussion of the archival community. This debate impinges upon all fields across the full span of academia, with the implications of our decisions on how to preserve electronic records directly affecting research in nearly all of the liberal arts and hard science programs. Outside of the academic world, electronic records from government agencies such as National Aeronautics and Space Administration (NASA), the United States Geological Society (USGS), or even the White House are being lost daily due to our lack of ability to properly process and preserve them. This lack of processing creates a dangerous lack of transparency within the government, a loss of valuable cultural treasures, and a complete erasing of some pieces of our recent history.

To conclude, Cox, citing the Australian Society of Archivists, puts forth the following idea to archivists:

Archivists ensure that records which have value as authentic evidence of administrative, corporate, cultural and intellectual

activity are made, kept and used. The work of archivists is vital for ensuring organizational efficiency and accountability and for supporting understandings of Australian life through the management and retention of its personal, corporate and social memory. Our task is to figure out how to do this in an age when information and evidence, and access to information and evidence, are changing.⁹⁵

I believe that the best way to keep up with evolving technologies and records types is to continue to evolve with them. One way is through the adoption of minimal processing.

⁹⁵ Ibid., 34.

References

Abraham, Terry., Stephen E. Baizarini,., and Ann Frantilla, "What is Backlog is Prologue: A Measurement of Archival Processing." *American Archivist* 48, Winter (1985): 31-44.

Ambacher, Bruce. "The Evolution of Processing Procedures for Electronic Records." In *Thirty Years of Electronic Records*, edited by Bruce I. Ambacher, 43-62. Lanham, MD: Scarecrow Press, 2003.

Association of Research Libraries Task Force on Special Collection. "Hidden Collections, Scholarly Barriers: Creating Access to Unprocessed Special Collections Materials in North American Research Libraries". June 6, 2003. <http://www.arl.org/bm~doc/hiddencollwhitepaperjun6.pdf> (accessed on February 10, 2007).

Bantin, Phillip. C.. "Strategies For Managing Electronic Records: A New Archival Paradigm?" *Archival Issues* 23, 1: (1998): 17-34.

Berkeley School of Information, "How Much Information: 2003." http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/printable_report.pdf (accessed February 10, 2007).

Boles, Frank. "Disrespecting Original Order." *American Archivist* 45, Winter (1982): 26-32.

Catanzaro, Christine and Hackbart-Dean, Pamela. "The Strongest Link: The Management and Processing of Archival Collections." *Archival Issues* 27, no. 2(2002):125-36.

Chapman, Ann.. "Collection-level Description: Joining Up the Domains." *Journal of the Society of American Archivists* 25, 2(2004): 149-155.

Cook, Terry. "Archival Science and Postmodernism: New Formulations For Old Concepts." *Archival Science* 1, (2001): pg 3-24.

Cox, Richard J. "Access In The Digital Information Age And The Archival Mission: The United States." *Journal of the Society of Archivists* April, Vol. 19, 1 (1998): 25-40.

Cruikshank, Kate, Caroline Daniels, Dennis Meissner, Naomi L. Nelson, and Mark Shelstad. "How Do We Show You What We've Got? Access to Archival Collections in the Digital Age." *Journal of the Association for History and Computing* 3, no. 2 (2005).
<http://mcel.pacificu.edu/jahc/JAHCVIII2/articles/cruikshank.htm> (accessed March 6, 2007).

Cunningham, Adrian. "Waiting For the Ghost Train: Strategies for Managing Electronic Personal Records Before It Is Too Late." *Archival Issues* 24, 1 (1999): 55-64.

Desnoyers, Megan Floyd. "When is a Collection Processed?." In *A Modern Archive Reader: Basic Readings on Archival Theory and Practice*, edited by Maygene F. Daniels and Timothy Walch, 309-325. Washington D.C: National Archives and Records Service, 1984.

Duranti, Luciana. and McNeill, Heather. "The Protection of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project." *Archivaria* 42, Spring (1997): 46-67.

Eastwood, Terry "Review of McKemmish and Piggott." *Archives and Manuscripts* 23, 1 (1995): 109.

Glick, Kevin L., and Wilczek, Eliot. "Changes in Acquisition: A Guide to the Ingest of Electronic Records." Paper presented at New Skills for a Digital Era, Washington, DC, May 31 - June 2 2006.
http://rpm.lib.az.us/NewSkills/CaseStudies/1_Glick_Wilczek.pdf (accessed March 6, 2007).

Glick, Kevin, and Eliot Wilczek. "Ingest Guide for University Electronic Records." Tufts University and Yale University, 2006.
<http://dca.tufts.edu/features/nhprc/reports/ingest/index.html> or
<http://repository01.lib.tufts.edu:8080/fedora/get/tufts:UA069.004.001.00006/bdef:TuftsPDF/getPDF> (accessed March 6, 2007).

Greene, Mark, and Meissner, Dennis. "More Products, Less Process: Revamping Traditional Archival Processing." *The American Archivist* 68, Fall/Winter (2005): 208-263.

Grimard, Jacques. "Maintaining the Long Term Preservation of Electronic Archives, or Preserving the Medium and the Message." *Archivaria* 59, Spring (2005): .

Hedstrom, Margaret. "Descriptive Practices for Electronic Records: Deciding What Is Essential and Imagining What Is Possible." *Archivaria* 36, (1993): 53-63.

Hyry, Tom. "More for Less in Archives: The Greene/Meissner Approach at Work at Yale." Presented at RLG Member Forum: More, Better, Faster, Cheaper: The Economics of Descriptive Practice, August 8, 2006. 1-10.

Hyry, Tom., and Onuf, Rachel. "The Personality of Electronic Records: The Impact of New Information Technology on Personal Papers." *Archival Issues* 22, 1 (1997): 37-44.

Jones, William, Ammy Jiranida Phuwanartnurak, Rajdeep Gill, and Harry Bruce. "Don't Take My Folders Away! Organizing Personal Information to Get Things Done." Paper presented at the Conference on Human Factors in Computing Systems, Portland, OR, April 2-7, 2005.

Ketelaar, Eric "Archives of the People, By the People, For the People," *South Africa Archives Journal* 34, (1992): 5-16.

McCrea, Donna E. "Getting More for Less: Testing a New Processing Model at the University of Montana." *American Archivist* 69, Fall/Winter (2006): 284-290.

MacNeil, Heather. "Providing Grounds for Trust II: The Findings of the Authenticity Task Force of InterPARES." *Archivaria* 54. (2003): 24-58.

Martin, Janette. and Thomas, Susan. "Using the Papers of Contemporary British Politicians as a Testbed for the Preservation of Digital Personal Archives." *Journal of the Society of American Archivists* 27, April (2006): 29-56.

Menkus, Belden. "Defining Electronic Records Management," *Records Management Quarterly* 30, January (1996): 38-42.

McInnes, Sally. "Electronic records: the new archival frontier?" *Journal of the Society of Archivists* 19, October (1998): 211.

McKemmish, Sue, Glenda Acland, Nigel Ward, and Barbara Reed. "Describing Records in Context in the Continuum: The Australian Recordkeeping Metadata Schema." *Archivaria* 48, Fall (1999): 3-37.

Owens, Evan, Suku Sukumar, John Meyer, David Copperman, and Roland Mesde. "Portico Technical Overview: A Format-Registry-Based Automated Workflow for the Ingest and Preservation of Electronic Journals." Paper presented at the NDIIPP Partners Meeting, Berkeley, California, November 2005. http://www.digitalpreservation.gov/partners/portico_tech.pdf (accessed March 6,

2007).

Producer Archive Workflow Network in Support of Digital Preservation (PAWN)
<http://umiacs.umd.edu/research/adapt/papers/UMIACS-TR-2004-49.pdf>
(accessed March 6, 2007).

Pyatt, Timothy. "Acquisitions: Assessment, Scheduling, and Transfer of Public Affairs Records." Paper presented at New Skills for a Digital Era, Washington, DC, May 31 - June 2, 2006.
http://rpm.lib.az.us/NewSkills/CaseStudies/3_Pyatt.pdf (accessed March 6, 2007).

Roe, Kathleen. *Arranging and Describing Archives and Manuscripts*. Chicago, IL: Society of American Archivists. 2005.

Rosenzweig, Richard. "Scarcity or Abundance? Preserving the Past in a Digital Era." *American Historical Review* 108, June (2003): 735-762.

Stollar, Catherine, and Thomas Kiehne. "Guarding the Guards: Archiving the Electronic Records of Hypertext Author Michael Joyce." Paper presented at New Skills for a Digital Era, Washington, DC, May 31 - June 2, 2006.
http://rpm.lib.az.us/NewSkills/CaseStudies/4_Stollar_Kiehne.pdf (accessed March 6, 2007).

Thibodeau, Kenneth. "Building the Future: The Electronic Records Archives Program." In *Thirty Years of Electronic Records*, edited by Bruce I. Ambacher, 91-104. Lanham, MD: Scarecrow Press, 2003.

Thibodeau, Kenneth., Moore, Reagan., and Baru, Chaitanya. "Persistent Object Preservation: Advanced Computing Infrastructure for Digital Preservation." Proceedings from the DLM-Forum on Electronic Records, Panel Session #1/2.. Standards and Specifications. Brussels, BE. (December 1996): 113-118.

Wallace, David A. "Managing the Present: Metadata as Archival Description." *Archivaria* 39, (1995): 22-32.

Weideman, Christine. "Accessioning as Processing." *American Archivist* 69, Fall/Winter (2006): 274-283.