



Processing Workflow Analysis for Special Collections: The Center for the History of Medicine, Francis A. Countway Library of Medicine as Case Study

Citation

Novak-Gustainis, Emily R. 2012. Processing Workflow Analysis for Special Collections: The Center for the History of Medicine, Francis A. Countway Library of Medicine as Case Study. *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 13 (2): 113-128.

Published Version

<http://rbm.acrl.org/content/13/2/113.extract>

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:24830899>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Processing Workflow Analysis for Special Collections: The Center for the History of Medicine, Francis A. Countway Library of Medicine as Case Study

IN OUR EFFORTS to make the whole more than the sum of our parts, it is easy to forget that the better the parts, the better the whole. Our special collections and archives professions have placed a premium on the utility of our descriptive products and services to our end users, and rightly so.¹ Yet there remains, despite numerous calls in our professional literature,² a scarcity of data regarding the process inputs that lead up to the delivery of our product and services outputs, especially (and perhaps most notoriously) data pertaining to archival processing. This paper focuses on the Center for the History of Medicine's implementation of a time and labor tracking methodology to facilitate process analysis for speeding access to archives and manuscripts and serves to underscore the need for archival repositories to programmatically collect and surface operational data as a means of communicating the full cycle of activities that culminate in a research resource.³ Best practices related to acquiring, accessioning, processing, delivering, and interacting with records do not emerge in isolation, nor should performance improvement, benchmarking, and advocacy activities take place without evidence.⁴

1. The work of Elizabeth Yakel, Helen R. Tibbo, Wendy Duff, and their collaborators in developing the Archival Metrics Toolkits has produced invaluable evaluation tools to improve the user experience. See: www.archivalmetrics.org/node/10.

2. See Paul Ericksen and Robert Shuster, "Beneficial Shocks: The Place of Processing-Cost Analysis in Archival Administration," *The American Archivist* 58, no. 1 (1995): 32–52, and Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *The American Archivist* 68, no. 2 (2005): 208–63.

3. At the time of publication, Center for the History of Medicine holdings included 16,020 cubic feet of archives and manuscripts. Roughly 29 percent of its holdings are unprocessed.

4. It is the "holistic approach that archivists have been missing as we grapple with the phenomenon of large backlogs by focusing on only one component—archival processing—in the larger archival business process," according to J. Gordon Daines III, "Re-Engineering Archives Business Process Management (BPM) and the Quest for Archival Efficiency," *The American Archivist* 74, no.1 (2011): 128. In addition to time and labor tracking for processing, the Center is currently tracking acquisitions and accessioning work, remote reference services, select records management services, and grant-funded digitization initiatives.

*Please note: URLs for all websites referenced herein were valid at the time of article submission.

The Center for the History of Medicine in the Francis A. Countway Library of Medicine, Harvard Medical School, launched its archival processing metrics initiative in 2009 as part of its Foundations of Public Health Policy (FPHP) project, which was funded by a grant from the Council of Library and Information Resources (CLIR) Hidden Collections program. In 2010, the Center opened its work to the public through the establishment of a Processing Metrics Collaborative.⁵ Conditional to funding, the Center promised to capture metrics related to collections processed under the auspices of FPHP as a first step toward the systematic collection of hard data for the purposes of communicating the true cost of processing collections to both administrators and donors, forecasting the duration of processing projects, and establishing baselines for quantifying how processing collections and record series to varying levels of granularity could speed their opening to research access.⁶

To accomplish these goals, the Center recognized that it must answer certain evaluative questions. At what point can a collection be considered “processed” and suitable for delivery to users? What actions are required from the point of accession to the point of delivery? How much time do staff invest in the various activities that comprise that workflow? How do staff know if modifying a processing activity is effective? To respond to these questions and to consistently track and examine processing outputs, the Center determined that its processors needed:

- shared definitions of processing activities;
- clearly articulated definitions of processing levels;⁷
- guidance on communicating how “time-consuming” a collection appeared to be at the outset of processing planning (that is to say, “collection complexity”);
- a way to capture basic information about the collections being processed to contextualize the data, which could otherwise be misleading;
- a standardized approach to processing planning;
- alternate approaches to performing processing tasks, such as performing refoldering and folder listing in separate “passes” rather than performing multiple processes on each folder before moving to the next; and

5. Center for the History of Medicine, Francis A. Countway Library of Medicine, “Processing Metrics Collaborative: Database Development Initiative,” available online at <https://wiki.med.harvard.edu/Countway/ArchivalCollaboratives/ProcessingMetricsDatabase>.

6. The dearth of hard data to support such activities, and the frequently encountered reluctance to capturing it, is clearly illustrated in Greene and Meissner, “More Product, Less Process.”

7. At roughly the same time, Harvard University’s then Manuscripts and Archives Access Group (MAAG) was establishing processing levels to facilitate a 2009 Harvard-wide special collections survey. Three processing levels were established by the university, which the Center continued to use post-survey, splitting Level II into A and B categories. Information relevant to the processing levels is available at: <https://wiki.med.harvard.edu/Countway/ArchivalCollaboratives/CHoMMetricsDocumentation>.

- documentation relevant to Center processing practices, which resulted in a processing manual.

These internal needs mirrored and continue to reflect the ongoing needs of the archival profession for a common language to describe processing activities. While we may not need to formulate prescriptive methodologies, we at least should be able to talk about what we do in consistent and objective terms. In the Center's case, documentation was compiled, evaluated, and revised throughout the first year of the FPHP project, with major revisions and additions to the processing manual continuing through the second year.

Methodology

To facilitate data collection, the Center created a relational database (colloquially called MD for Metrics Database) using Microsoft Access for capturing timing data for use by staff throughout collections processing. In its initial iteration, this database was ready to use from the beginning of the FPHP initiative and continued to be modified throughout the project. Data collection was initially limited to FPHP staff, which consisted of one full-time project archivist and two half-time processing assistants. After six months of use by FPHP staff, the rest of the Center's processing staff, including interns, began using MD to track their processing activities.

Using MD, processors performed the following sets of tasks:

1. Recorded "top level" information about a collection that could quickly contextualize time and labor tracking data, including:
 - records content; pre- and postprocessing volume in cubic feet; a description of the staffing configuration for processing; type of collection;⁸ the processing level selected for the majority of collection; collection complexity ranking;⁹ primary collection format;¹⁰ and the date span¹¹ for the bulk of the records;

8. Itemized as: Personal papers; Professional papers; Personal and professional papers; Archival (institutional records); Corporate records; Government records; and Literary manuscripts.

9. Itemized as: 1: No rearrangement necessary. Processed as-is; 2: Some rearrangement required, but of the "minor housekeeping" variety. (Some folders need to be moved to different boxes, etc.); 3: Rearrangement necessary to restore (perceived) original order, but series are identifiable/recognizable; 4: Rearrangement necessary, but series are hard to determine "up front," requiring periodic re-evaluation of processing plan as records are handled; and 5. Full archivist-imposed organization necessary. Records are not in any discernible order or groupings, requiring record-by-record handling and grouping.

10. Itemized as: Computer disks; Drawings; Drawings, architectural; Gigabytes of data transferred (without a media or carrier); Microfilm/microfiche; Negatives and transparencies; Objects; Paper/textual records; Photographs; Printed matter; Reels of motion picture film; Sound recordings; and Video recordings.

11. Itemized as: 1800 or earlier; 1801–1900; 1801–1850; 1851–1900; 1901–present; 1901–1950; and 1951–present.

- a synopsis of critical information found in the processing plan relevant to pre- and postprocessing arrangement and condition;
 - commentary, such as problems encountered, major changes to processing plans, and specific preservation decisions; and
 - the date of completion for critical deliverables (processing plan, folder list, finding aid, MARC record, and deposit of curated scans to institutional repository).
2. Logged the date and the amount of time spent on discrete processing activities for specific collections to the quarter hour. Eleven tasks were initially identified and tracked; the list expanded after the first year of tracking to 19 processing-related activities (listed below in A–Z order):¹²
1. *barcoding and labeling boxes*;
 2. *collection review and processing planning* (for tracking the creation and revision of processing plans, as well as any presorting of records);
 3. *description* (for tracking the preparation of biographical notes, authoring scope and content notes and series descriptions, drafting bibliographic records, performing subject analysis and name authority research, and for anything else related to authoring and encoding finding aids not specific to folder listing, folder list export and tagging, and MARC cataloging in Aleph);
 4. *digitization* (for tracking curatorial digitization, in which selected items are scanned for online delivery, blog posts, and exhibition use, but are not scanned as part of large digitization initiatives);
 5. *electronic records review (collections processing)* (for tracking the opening, review, and transfer of electronic records as conducted as part of collections processing);
 6. *finding aid review (processing staff)* (for tracking staff review of their own finding aids or finding aids created by the interns they manage; work related to editing and publishing finding aids tracked separately by the Collections Services Archivist);
 7. *folder list export and tagging* (for tracking time performing Excel calculations on folder lists to generate encoded container lists);
 8. *folder listing* (for tracking the creation of folder lists in accordance with practices outlined in the Center’s processing manual);

12. The total number of tasks currently tracked in database is 49, which encompasses accession-level collection listing, intern management, and tasks specific to the work of the Collections Services Archivist, digitization initiatives, and special projects administration. As of July 27, 2012, the Center has logged more than 10,000 hours of staff time over 5,710 tracking records. The full list of activities is available at <https://wiki.med.harvard.edu/Countway/ArchivalCollaboratives/CHoMMetricsDocumentation>.

9. *integration* (for tracking activities specifically related to interfiling records, including photocopied cross references);
10. *item listing* (for tracking the production of item-level inventories/ descriptions as part of folder listing);
11. *MARC/bib record creation/revision* (for tracking the use of ALEPH to create and edit bibliographic records);
12. *MD tracking* (for tracking the amount of time staff spends capturing data in MD; usage records created daily or reported cumulatively for a week using one record per collection);
13. *preservation photocopying*;
14. *preservation tasks (misc.)* (for tracking the occasional removal of staples and paper clips, wrapping items in archival paper, and putting negatives, photographs, and slides in sleeves as necessary);
15. *promoting in-process and processed collections* (for tracking the authoring of blog and electronic mailing list posts as well as writing newsletter articles about recently opened collections);
16. *rearranging*;
17. *reboxing*;
18. *refoldering* (for tracking the refoldering of collection content and labeling folders); and
19. *review and separation of archival records from faculty (personal) papers.*

These activities, often performed for boxes in separate passes to facilitate tracking, had the effect of “atomizing” processing for the Center and served as a foundation for talking to other repositories about their workflows and what they were interested in tracking. It also suggested “buckets of activities” that the profession could use as a first step toward standardizing the way repositories talk about processing activities so that time and labor tracking could be conducted consistently across multiple repositories and aggregated for analysis,¹³ as well as used to illustrate the investment of time (and by extension, cost) in specific activities on the part of processors without prescribing a specific processing methodology. This emphasis on activities as opposed to methodology was a particularly important distinction as the Center began talking about and distributing copies of the database for use by other repositories.

FPHP processors also used additional database forms to track the time spent processing series by adding up the amount of time spent on performing specific ac-

13. An initial step toward aggregating data can be found in the activities mapping spreadsheet on the wiki: <https://wiki.med.harvard.edu/Countway/ArchivalCollaboratives/WorkflowIssues>.

tivities for each cubic foot of records processed,¹⁴ a practice ultimately deemed too cumbersome to sustain. No other repositories that have tested MD have pursued box-level or series-level tracking.

Findings

Processing Rates

Despite pervasive opinions that workflows at the repository level are too disparate to be juxtaposed, as archivists we nevertheless strive for a one-size-fits-all answer to outputs. What volume of records should we expect an experienced professional to process in a given year? What is a reasonable amount of records to put in my grant proposal? In part, these quandaries derive from our desire to apply processing methodologies uniformly rather than appropriately to the types of materials. It also stems from wanting to benchmark against a professional norm for processing and description rather than against the ways our researchers and our own public services staff experience and use the descriptive tools we produce. Between 1976 and 2011, roughly 52 processing rates have surfaced in professional literature and online resources, 37 of which were reported by nine University of California repositories to the California Digital Library as part of a grant initiative.¹⁵ The majority of those predating 2005 were referenced in Greene and Meissner's "More Product, Less Process" study (which posited that an average processing rate for large collections of twentieth-century materials should be four hours per cubic foot¹⁶), with the most recent analysis being the California Digital Library's 2011 *Uncovering California's Environmental Collections Processing Metrics Report*, in which the processing rate for 17 collections processed across multiple repositories averaged 13 hours per linear foot.¹⁷

For FPHP processed collections, the Center's time and labor tracking revealed that, in a team-based processing environment (a project archivist with one or two processing assistants, depending on the collection), the processing of predominantly paper-based, moderately complex, twentieth-century collections of personal and professional papers averaged 3.57 hours per cubic foot when measured by the initial volume

14. This required processors to maintain a separate box-level tracking form. The Center stopped box- and series-level tracking after the FPHP project concluded and after each staff processor had the experience of timing at least one collection to the box level. Timing per box yielded results similar to those obtained by dividing the total number of hours spent processing the collection by the number of boxes. For example, the average number of hours per cubic foot achieved by dividing cubic feet as a whole for the Rutstein papers is 2.92 hours per cubic foot on average using the start volume and 3.66 hours per cubic foot using the end volume; for Series 1: Correspondence and Subject Files (37 cubic feet), the rate was 2.09 hours per cubic foot for core activities (which exclude collection review and project planning, project tracking, digitization of selected items, collection-level description, and barcoding and labeling).

15. See: <https://wiki.med.harvard.edu/Countway/ArchivalCollaboratives/Resources>.

16. Greene and Meissner, "More Product, Less Process," 253.

17. Melanie Wisner, *Uncovering California's Environmental Collections: A Collaborative Approach (CLIR UCEC): Processing Metrics Report* (Washington, D.C.: Council on Library and Information Resources, October 2011), available online at www.clir.org/hiddencollections/resources/UCECfinalRev.pdf.

of records or 4.64 hours per cubic foot if measured from the volume of the processed collections.¹⁸ Collections processed by nongrant archivists over the same period of time averaged 4.43/5.66 hours by start/end volume.¹⁹ Programmatically, across the 39 collections processed between the implementation of MD in 2009 and May 2012, Center staff averaged 4.72 hours per cubic foot, logging about 4,992 hours of activities specifically flagged as “processing” activities. These averages include collections processed by interns in partnership with Center staff and the use of a processing assistant for non-FPHP collections when funding was available. They also include a number of collections that the Center considered statistical “outliers,” which must be contextualized to interpret overall rates.²⁰ Processing rates for FY2012, which did not include any such outliers, ran between 9 and 12 hours per cubic foot for “pure processing” and 6.88 hours start/7.94 hours end per cubic foot when including collections reprocessing and access improvement projects. Given that 20 collections were processed in FY2012, significantly more processing planning was required of staff in FY2012 than in previous years, raising the processing rate per cubic foot.²¹

Throughout, time and labor tracking was analyzed from the three following perspectives.

1. Perceived collection complexity (see figure 1):
 - Four collections assigned a complexity level 1 (least complex) averaged 12.25 hours per cubic foot (however, one of these collections was processed by a volunteer and one by and intern);
 - Fifteen collections assigned a complexity level 2 averaged 5.41/7.16 hours per cubic foot (start/end);
 - Thirteen collections assigned a complexity level 3 averaged 3.03/3.62 hours per cubic foot (start/end);
 - Six collections assigned a complexity level 4 averaged 4.48/6.3 hours per cubic foot (start/end);
 - One collection assigned a complexity level 5 (most complex) averaged 17.92/17.92 hours per cubic foot (start/end).

18. This rate was calculated from the 1,490.25 hours required to process an initial volume of 418 cubic feet of records. Calculating from the end volume of 321 cubic feet, the processing rate was 4.64 hours per cubic foot.

19. These rates were calculated from the 3,501.75 hours required to process an initial volume of 790 cubic feet with a resulting collections size of 618.6 cubic feet.

20. Most notably, the processing of the Saltpêtrière Hospital records, 1859–1942 (inclusive), for which physical rearrangement was deemed impractical and for which collection glass plates had already been rehoused by the Weissmann Preservation Center staff as part of a Harvard University-wide initiative.

21. Because processing planning is vital to the Center’s processing methodology, especially in team processing or processor/assistant configurations, the creation of processing work plans is considered by the Center to be a processing activity. Processing plans are revised as needed throughout processing in conjunction with MD data entry to ensure the processing and description of collections is appropriately contextualized.

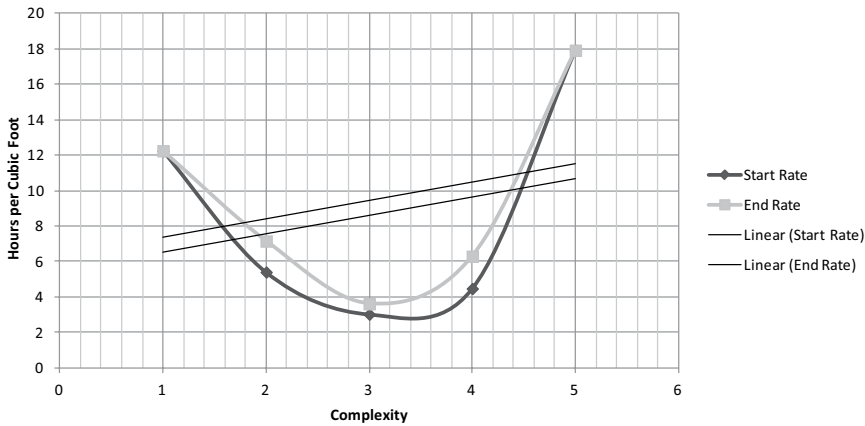


FIGURE 1: Outputs by Collection Complexity, 2009–2012

Contrary to what one might expect, collections of moderate complexity clearly took less time to process than collections of low complexity. This may be because lower complexity collections were typically assigned to interns and newer professionals who had relatively little prior processing experience, if any.

2. Processing level (see figure 2):

- Processing rates by level averaged 5.45/7.9 hours per cubic foot (start/end volume) for collections that were folder-listed in their entirety (Level IIB);
- Processing rates for collections employing folder lists for some series and box-level summaries for others averaged 2.91/3.33 hours per cubic foot (Levels IIA and IIB);
- The single Level III collection that was processed took 8 hours (a 1.0 cubic foot collection);

3. Staffing configuration:

- Rates for archivists processing collections in their entirety, from start to finish, without assistance, averaged 3.53/3.82 hours per cubic foot (519 start/480.85 end volume) over 20 collections (1,834.5 hours logged); the average of all processing rates for individually processed collections was 7.76 hours per cubic foot;
- Rates for archivists processing collections in a 1:1 processor-assistant/intern/volunteer team averaged 5.02/8.13 hours per cubic foot (393 start/242.75 end volume) over 14 collections (1,973.75 hours logged); the average across all processing rates for 1:1 processed collections was 9.84 hours per cubic foot;

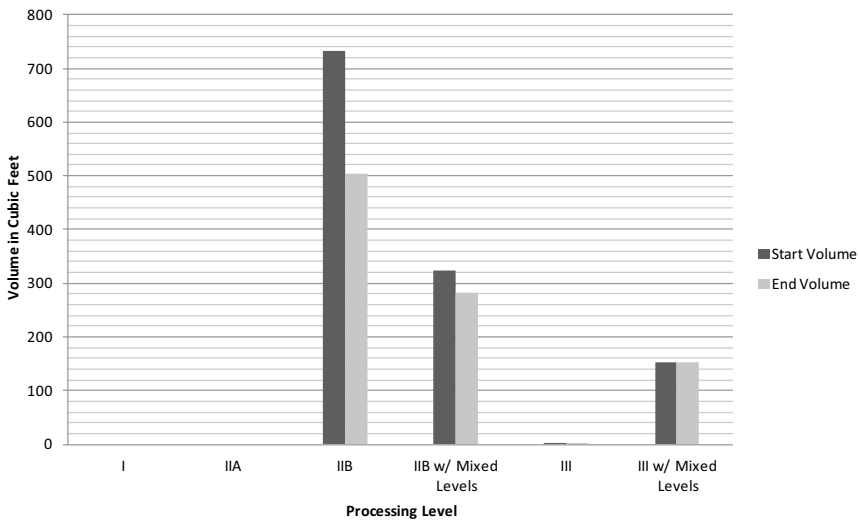


FIGURE 2: Outputs by Processing Level, 2009–2012

- Rates for archivists processing collections in a 1:2 processor-assistant team averaged 4/5.48 hours per cubic foot (296 start/216 end volume) over 5 collections (1,183.75 hours); the average across all processing rates for 1:2 processed collections was 6.36 hours per cubic foot.

While the data indicate that having one processor with multiple assistants provided the most efficient processing model, anecdotal evidence suggests that the most effective staffing configuration for larger collections was one processor working with one processing assistant, as staff commented that it simplified project management.

Almost all collections processed by the Center between FY2009 and FY2012 were twentieth-century paper/textual records; two were hybrid collections (Hay and Simon); and one collection was composed almost entirely of glass plate negatives. Data supporting the analysis of processing rates for collections dating prior to 1900 or containing a significant number of nontextual records (such as datasets and architectural drawings) will need to be collected as processing opportunities emerge. Such a knowledge base can, however, be built by the special collections community as a whole through data sharing, with the objective of illustrating that different types of collections, of varying complexities, processed to different levels result in different outputs, as well as emphasizing that the sought-after “predictability” of processing rates depends upon the particular intersections of collection characteristics, divisions of labor, and experience levels of the processors. For example, two Processing Metrics Collaborative participants, North Carolina State University

and the Free Library of Philadelphia, have generated extensive timing information related to processing collections of modern architectural drawings and the papers of literary authors, respectively.²²

Processing Activities

Timing for specific processing-related activities yielded few surprising results (see table 1 below). It is obvious that the manual labor inherent in processing is the most time consuming, even when not everything is refoldered and almost no clips or fasteners are removed, as in the Center's case. It is not only apparent that we can better leverage the skills of professional archivists by providing support staffing options for the refolding and listing of collections, but that we *must* do so if we are to combat backlogs and advance technical processing methodologies for digitized collections to offset the need for physically arranging collections. Not insignificant is the amount of time that processors spend formulating work plans, documenting how a collection was processed, and recording the rationale for processing decisions.

TABLE 1: Time Required to Perform Specific Processing Tasks²³

Activity ²³	Hours	Percentage of Total Hours (4991.75 hours)
Refolding	1,050.50	21.05
Folder Listing	1,005	20.13
Description	700.25	14.03
Rearranging	520.25	10.42
Collection Review and Processing Planning	471.5	9.45
Folder List Export and Tagging	208.25	4.17
Preservation Tasks (Misc.)	143.75	2.88
Preservation Photocopying	140.5	2.81
Active Records Survey (Preprocessing Planning)	129.5	2.6
Editing and Publishing Finding Aids (CSA)	91.75	1.84
MD Tracking	87.5	1.75
Intern Management	68.75	1.38
MARC/Bib Record Creation/Revision	68	1.36
Processing Research	62	1.24

22. See North Carolina State University's *Changing the Landscape* blog, available online at <http://news.lib.ncsu.edu/changinglandscape/>; and The Free Library of Philadelphia's activities timing reports, available online at <https://wiki.med.harvard.edu/Countway/ArchivalCollaboratives/FreeLibrary>.

23. The table includes four activities not initially tracked as "core" processing activities: Processing Research, Preprocessing Box and Folder Listing, Selection and Deduplication (Special Projects), and Intern Management.

Digitization	60.5	1.21
Promoting In-Process and Processed Collections	51	1.02
Barcoding and Labeling	31	0.62
Review and Separation of Archival Records from Faculty Papers	23.75	0.48
Electronic Records Review (Collections Processing)	21.5	0.43
Reboxing	21	0.42
Finding Aid Review (Processing Staff)	18.5	0.37
Preprocessing (Accession Level) Box and Folder Listing	16.25	0.33
Selection and Deduplication (Special Projects)	0.75	0.01

Projecting Outputs and Cost

The question of how we can contextualize processing data to predict outputs requires us to reconsider what constitutes processing activities. Despite the realities of work environments, we nevertheless expect processors (particularly project archivists) to devote all of their working hours to processing. Yet professional archivists often have public services obligations, participate in staff initiatives, manage interns, and take on special projects as needed. Tracking of the Center's staff Processing Archivist over a 24-month period (June 2009–June 2011) revealed that he invested roughly 20 percent of his time on activities defined as processing, 10 percent staffing the public services desk, and 70 percent on other activities (depository account management, event preparation, meetings, supply orders, continuing education, special assignments, sick/paid days off, and holidays). During this period, he opened a total of 350 cubic feet of records with the aid of a processing assistant for two of the six collections he processed.

Expectations of project archivists are much narrower. While the FPHP project resulted in the processing of 414 cubic feet start/318 end (including 128 cubic feet more than proposed), processing activities conducted by the Project Archivist over a 24-month period (April 2009 to February 2011) with two part-time processing assistants comprised only 13 percent (441 hours) of his time, with 87 percent delegated to outreach and publicity, engagement studies, exhibition work, meetings, sick/paid days off, and holidays. By comparison, a part-time processor who started off as a processing assistant, and who contributed over 22 months of activity (October 2009–July 2011) with no “added value” assignments, spent 51 percent (approximately 757 hours of 1,496.25) of her time processing and 49 percent on outreach, meetings, and the preparation of supply orders. She processed 182 cubic feet (start/170 end), the entirety of H MS c334 (the Lauriston Sale Taylor papers, 1904–1999 inclusive), by processing series in the collection to either the box (IIA) or folder level (IIB). This enabled the collection to be processed faster while ensuring

folder-level description was performed for the records Public Services targeted as being of most interest to researchers.²⁴

Irrespective of position type, over the past three years, the Center has dramatically increased its processing outputs through the implementation of a uniform processing program.²⁵ Short of dramatically changing the quality of our processing products, we have little efficiency left to glean with current staffing levels; only by hiring additional processors or by freeing up more staff time to devote to processing can we hope to further increase outputs. While the Center does not feel administrative pressure to open the same volume of records (if not more) with each passing fiscal year, given grant/special projects staffing fluctuations and the recent high volume of acquisitions,²⁶ it is important for the Center to be able to contextualize its outputs, illustrating that, while the volume of collections processed may vary from year to year, Center staff are still processing as efficiently while participating in the greater work of the Center. Isolating processors from other dimensions of work can detract from staff unity and affect morale; in some cases, however, we have been able to redirect work to different or more appropriate staff members to free up some additional processing time. Additionally, our profession's dependency on grants and project funding has bred contradictory expectations: to obtain funding to process collections, we find ourselves in the position of having to commit ourselves to delivering so many "extras" to sell our projects (conferences or project events, constituency engagement and outcomes analyses, and developing project websites, blogs, and Web 2.0 presences, among others) that processing takes the backseat, leading to what some have colloquially referred to as the creation of a "postgrant" environment in which the benefits of smaller, more controlled experimentation at home are preferable to the constraints of major funding.²⁷

As a consequence of understanding better the limited amount of time we have available to process collections, the Center implemented a tiered approach to estimating the cost of collections processing. Rather than leveraging a single flat rate (which was \$800.00 per cubic foot until 2010), estimates presented to donors and potential funders for processing a collection are now predicated on the intersection of access objectives

24. While one FPHP collection employed different processing levels, the Taylor collection was the first to incorporate the use of multiple processing levels in processing planning. Subsequently, box-level processing and description has been applied to collections containing reprints and publications meriting long-term retention, as well as to patient records that cannot be opened for decades to researchers without Internal Review Board approval.

25. In FY2009, the Center produced 612 cubic feet of fully processed, reprocessed, or box- and folder-listed collections. In FY2010, that total was 778.5 cubic feet; and, in FY2011, 1,169.35 cubic feet.

26. Volume of acquisitions was 487 cubic feet of manuscripts and archival records for FY2008; 519 cubic feet for FY2009; 219 cubic feet for FY2010; 1,292.25 cubic feet for FY2011; and a projected 1,207 cubic feet for FY2012.

27. This was a point of discussion at the Second Radcliffe Workshop on Technology and Archival Processing, Radcliffe Institute for Advanced Study, Harvard University, Cambridge, Mass., June 8, 2012.

and perceived collection complexity and are based on either a survey of sample records or the appraisal notes of the Acquisitions Archivist. Proposed rates now range between \$400.00 and \$900.00 per cubic foot.²⁸ More complex collections take staff longer to process in a given fiscal year, meaning that the possible number of collections that can be processed is greatly reduced and should cost more. Tackling our backlog, which is rife with complex collections, therefore means that more professional staff and processing assistants will be needed if the Center wants to avoid adding to its backlog.

TABLE 2: Costs per Cubic Foot for Varying Degrees of Processing Complexity

Complexity	Criteria	Cost per Cubic Foot
I	No rearrangement necessary (processed as-is).	\$400.00
II	Some rearrangement required, but of the “minor housekeeping” variety (some folders need to be moved to different boxes, for instance) and there is a folder list.	\$400.00
	Some rearrangement required, but of the “minor housekeeping” variety (some folders need to be moved to different boxes, for instance) but there is no folder list.	\$500.00
III	Rearrangement necessary to restore (perceived) original order, but series are identifiable / recognizable and accompanied by an electronic folder list.	\$600.00
	Rearrangement necessary to restore (perceived) original order, but series are identifiable / recognizable. Not accompanied by an electronic folder list.	\$700.00
IV	Rearrangement necessary, but series are hard to determine “up front,” requiring periodic re-evaluation of processing plan as records are handled. Accompanied by electronic folder list or other documentation.	\$700.00
	Rearrangement necessary, but series are hard to determine “up front,” requiring periodic re-evaluation of processing plan as records are handled. Not accompanied by electronic folder list or other documentation.	\$800.00
V	Full archivist-imposed organization necessary. Records are not in any discernible order or groupings, requiring record-by-record handling and grouping.	\$900.00

28. Costs per cubic foot assume processing to the folder (IIB) level; the actual cost of processing may be less (or more) if series (or subseries) are processed to different levels. The base cost of \$400.00 is approximately the cost per cubic foot for a professional archivist working at the Center, as predicated on the ratio of processing time to other assigned tasks (per the processing timing analysis made possible through the use of MD). The Center added \$100.00 to the base cost to account for each level of increased collection complexity, which would require more time, more project management / supervisory time, might require hiring processing support staff, and so forth. Costs per foot do not include processing supplies or fees for retrieving collections from off-site storage.

Finally, data capture enabled the Center to judge the efficacy of new processing practices. For example, in 2006, the Center estimated the processing of the Lauriston Sale Taylor papers, 1904–1999 (inclusive) to take a professional, full-time archivist three years at a cost of \$145,600.00 employing the flat processing rate of \$800.00 per cubic foot. In 2010, the Center created a new processing plan for the still-unprocessed collection that included a total preprocessing volume of 182 cubic feet. The plan called for applying box-level processing for 80 feet of expert witness testimonies and serials, reprints, and gray literature, and folder-level processing for 90 feet of committee and organizational records and Taylor’s research, subject files, and unpublished manuscripts. Levels of description for different series in the collection were selected based on which series were more likely to be used by researchers. With this new methodology, the collection was processed in one year by an archivist working 17 hours per week at a fraction of the 2006 estimate, which was fair and appropriate for its time.²⁹

Conclusion

Recently, as Center staff were working out how acquisitions and records management activities could be grouped for the purposes of workflow analysis, one member commented that it was difficult to think of herself in hours, to which the Deputy Director artfully replied, “You were always a statistic. Now, at least you can be an activity, too!” Perhaps we could all benefit from thinking about the processes that constitute our day-to-day work. In early 2010, the Center conducted a poll regarding how repositories measured outputs. There were 65 respondents, the majority from college or university archives and special collections in an academic or library environment. The results were surprising: 24.6 percent said they did not keep statistics on collections processed in a calendar or fiscal year, 54.5 percent did not keep statistics on creating and encoding finding aids, 87.3 percent did not measure the amount of time they spend on creating processing plans, and about half did not maintain any statistics on digitization, among other processing and descriptive activities.

Since the survey, the Center has worked with a number of repositories on using MD with some success to examine these gaps in data collection, notably two institutions that received CLIR project funding in 2009: North Carolina State University Libraries (*Changing the Landscape: Exposing the Legacy of Modernist Architects and Landscape Architects*; Emily Walters, Project Archivist) and the Free Library of Philadelphia (*Milestones in 20th-Century American Children’s Literature at the Free Library of Philadelphia*; Adrienne Pruitt, Project Archivist). At Harvard, the

29. In 2006, processing was estimated at \$145,600.00. In 2010, processing costs were estimated at \$800 per cubic foot (Level IV complexity rate), for a total estimated cost of \$72,800.00 if the entire collection was processed to the folder level (IIB). Actual cost in 2011, with 90 cubic feet processed to the folder level (IIB) and 80 cubic feet processed at the box level (IIA), was approximately \$34,000.00.

Arnold Arboretum Archives, Houghton Library, the Historical & Special Collections Division of Harvard Law School Library, the Harvard University Archives, and Baker Library Historical Collections have experimented with MD, with Houghton currently using MD to track accessioning work and the Law School Library using MD to track collections processing. Repositories that tested, but did not adapt MD, reported that the process was too cumbersome or time consuming, that Microsoft Access was difficult to use and made data entry too confusing, or that they had difficulties enforcing data collection.³⁰ Similarly, while presenting MD to potential users/repositories, the Center was frequently told that, although the resulting data was valuable, their staffs were too small to commit to this type of work or that they simply did not process collections in a way that could be tracked.

The Center has shared its MD documentation with the California Digital Library (2010), Harvard's Productivity Tools Strategic Conversation (2011), and the University of California's Next-Generation Technical Services (NGTS) initiative (2012), as well as at professional conferences. The Center also held Processing Metrics Collaborative gatherings at the 2010 and 2011 Society of American Archivists annual meetings. All along the way, we have received excellent feedback and an appreciation for the work conducted, but we have had to be alert to opportunities for such comments. Because MD can be downloaded directly from the Center's wiki, the total number of institutions that are currently using it, have used it, or have modified it to suit local needs is unknown, and any data collected have yet to be shared with the greater archives community.

To succeed in our endeavor to understand better how we, as archivists, work, we must:

1. become more comfortable as a professional community with operational transparency;
2. develop a shared processing vocabulary so that benchmarking is possible, both within our local institutions and beyond; and
3. create a shared space for disseminating processing-related data³¹ or encourage users of shared collection management/archival description software to use system fields in the same way to facilitate reporting.³²

30. Repositories are listed on the Processing Metrics Collaborative wiki, available online at <https://wiki.med.harvard.edu/Countway/ArchivalCollaboratives/CollaboratorDocumentation>.

31. See the Center's unfunded 2010 Library Lab application, available online at <https://wiki.med.harvard.edu/pub/Countway/ArchivalCollaboratives/CHoMMetricsDocumentation/CountwayMetricsProject20101122.doc>.

32. See Harvard University's comments regarding ArchivesSpace and MD in its Collection Management specification feedback: Rachel Wise, "Harvard University: Collection Management Records Specification Comments," November 15, 2011, GoogleGroups, Archives Space, available online at <https://groups.google.com/forum/?fromgroups#!topic/archivesspace/98kjGAAYU8>.

The shift from outputs-based to outcomes-evident assessment is vital to the profession. Just as process evaluation is void of meaning without public-facing assessment, understanding how we produce a product is essential to responding to constituent feedback and facilitating change. In 1995, Paul Eriksen and Robert Shuster proposed that archives conducting cost analyses of their processing work should be encouraged to report their findings, suggesting that the Society of American Archivists might serve as a hub for disseminating findings.³³ One repository cannot do it alone, nor would it be desirable.³⁴ A national processing metrics initiative continues to be of interest to numerous people in the profession,³⁵ but questions of responsibility for data collection and project sustainability loom large, not to mention fear of transparency and the misuse of data. Data collection is not always perfect. People are human and make mistakes; certainly, examinations of populated versions of MD supplied on the wiki bear out this truism. Yet the more we make data collection part of our routine operations, the more responsive we can and must be in an increasingly data-driven environment.

33. Paul Eriksen and Robert Shuster, "Beneficial Shocks: The Place of Processing-Cost Analysis in Archival Administration," *The American Archivist* 58, no. 1 (1995): 32–52.

34. See: Wendy M. Duff et al., "The Development, Testing, and Evaluation of the Archival Metrics Toolkits," *The American Archivist* 73, no. 2 (2010): 569–99.

35. The Center's Processing Metrics Collaborative wiki had 674 page views between February 1, 2012, and July 26, 2012, the period during which this article was written.

CALADEX

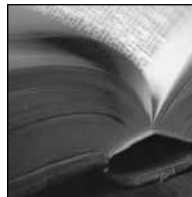
LOGISTICS

PERSONALISED TRANSPORTATION MANAGEMENT

Specializing in fine book packing and shipping

On-site packing by our qualified team

Door to door transportation from any USA point



For information on Caladex LLC and the services we offer, please contact us at:

USA & CANADA

Tel: 800 643 4315
Fax: 718 233 4159

UK

Tel: 020 7993 8511
Fax: 020 7681 3435

Email: info@caladex.com • www.caladex.com