Uncovering California's Environmental Collections:
A Collaborative Approach (CLIR UCEC)

Processing Metrics Report
October 2011

Melanie Wisner
Intern, California Digital Library

Contents

Acknowledgements

Introduction

Discussion of rates of processing archival collections date back to early days of the modern archival era, an era defined by the establishment of the National Archives and Records Administration (1934) and the Society of American Archivists (1936). The collective recognition of the twentieth century "avalanche of paper" was evidenced in articles on the escalating dollar costs of archival administration-- discussions continuing with vigor into the electronic era as our focus moves to the collective backlog of unprocessed collections. Writings from the 1980s on (see bibliography) highlighted the importance of methodology, standardization, self- evaluation, cooperation, and meaningful processing levels in reducing backlogs. The "More Product, Less Processing" (MPLP) approach put forward by Mark Greene and Dennis Meissner in 2005 lit the fire anew for an overhaul of the archival mindset (Greene & Meissner, 2005). The professional literature and annual SAA conferences have seen papers and case studies of MPLP in practice since, as well as "rebuttals" (which in many cases serve to refine rather than throw out MPLP) (Cox, 2010; Meissner & Greene, 2010). In the past processing metrics were expressed as dollar cost per container or quantity processed per week. Currently, rates expressed in hours of processing per linear or cubic foot (regrettably, we don't all agree on the unit of measurement), are often cited or recommended: a competent archivist should be able to process large twentieth-century archival materials at an average rate of 4 hours per cubic foot, according to Greene and Meissner (Greene & Meissner, 2005, 253).

That said, five years later in reviewing the uptake of and response to MPLP ("More Application While Less Appreciation: The Adopters and Antagonists of MPLP," 2010), Meissner and Greene emphasize that "MPLP, fundamentally, is not about specific processing actions. It is about resource management..." and the goal, to "[e]stablish an acceptable minimum level of work, and make it the processing benchmark".

What is less often discussed is the means by which we might derive benchmarks and therefore accurately estimate resources required to process specific collections. Local choices are made within an MPLP framework aiming to reduce backlog while providing sufficient access; it is not a simple proposition to derive a formula from the multiplicity of these archival approaches. The best we can do is to share detailed description of our efforts to process more efficiently; a few significant projects have pointed the way forward in recent years.

Current related work

The Northwest Archives Processing Initiative (NWAPI) project (Northwest Archives Processing Initiative Phase II, 2007), bears close resemblance to the UCEC project. Eight repositories from Alaska, Oregon, and Washington shared MPLP techniques and a standardized finding aid tailored for a consortial finding aids database. With the help of Mark Greene as advisor, the consortium aimed for 7-8 hours' processing time per linear foot but achieved an average of 2.9 (excepting one repository that encountered a high proportion of photographic materials, an important exception mirrored below).

Work done in 2009 at the Center for the History of Medicine (CHoM) at Harvard Medical School as part of a Foundations of Public Health Policy CLIR grant (Gustainis, 2010) developed a valuable practical tool. CHoM's metrics database, MD, could provide a body of nationally comparable data were it to receive broader archival review and adoption; it is a relatively complex, granular, but adaptable tool. MD has been adopted by CLIR grant recipients North Carolina State University Libraries and The Free Library of Philadelphia, and willing participants (as judged by CHoM wiki members) number around 20.

The North Carolina State University Libraries' "Changing the Landscape: Exposing the Legacy of Modernist Architects and Landscape Architects" project (Walters, n.d.) headed by Emily Walters, another CLIR Hidden Collections effort (2009 cohort), has, as noted above, employed the MD tool from its beginning for cost analysis. Besides taking place at one institution, controls in this case include consistency of and development of expertise with record formats and a standard workflow employing Excel with the Archivists' Toolkit. The project includes a blog chiefly written by Emily Walters.

The Philadelphia Area Consortium of Special Collections Libraries (PACSCL) CLIR Hidden Collections project, 2008 cohort (Mengel, n.d.), managed by Holly Mengel, has employed a sharable methodology for consortial processing that includes capture of metrics in a scheme somewhat simpler than MD. While the metrics are less granular than in CHoM's database, the project includes significant controls in the form of shared training, tools and, most important, a roving processing team. A blog for this project is written by processors as well as by project staff.

It should be noted that detailed processing manuals maintained by individual repositories have been generously shared online and are used by other repositories looking for guidance in processing planning (Yale University's Beinecke Library's manual is one often cited (Beinecke Rare Book and Manuscript Library, 1997)).

Despite these efforts, and despite the profession-wide conversation surrounding minimal processing and its effects, there is no widespread agreement on what to measure, how to measure, or how to analyze what is measured. There are many points to finesse in this discussion, particularly as grant-funded processing initiatives encourage more consortial work. Differences between the work of trained archivists and paraprofessionals or students is one. Making sure that the gathering of metrics does not verge on individual performance evaluation is another. It will be helpful to see what may be learned from processing initiatives in progress, including what is difficult about capturing metrics and how it may be improved. Despite the relative simplicity of method and small sample size for the UCEC project, we hope the following are useful observations based on field data.

UCEC metrics background and method

The goal for the processing metrics aspect of the CLIR UCEC grant was straightforward: seize the opportunity to gather and compare processing data from nine institutions sharing consortial processing goals and certain processing conditions. The particular metric chosen was processing rate, measured in hours per linear foot. One driver for this effort is found in the University of California Next Generation Technical Services (NGTS) initiative (Regents of the University of California, 2011), a systemwide examination of tools, workflows, and structures begun in 2009 that includes important vectors for special collections and archives. The December 2010 system priorities from the NGTS report include two bullet points found under the rubric of "New Modes for Organizing and Providing Access to Special Collections, Archives, and Digital Formats": "implement efficient 'More Product, Less Process' (MPLP) tactics for processing archival and manuscript collections"; and "support streamlined processing workflows and reuse descriptive data with systemwide use of the Archivists' Toolkit" (p. 1). More broadly, these potential implementations resonate with the NGTS concept that "[c]ost savings and cost avoidance are strategic" (http://libraries.universityofcalifornia.edu/sopag/ngtsframework.html, 2011). The CLIR UCEC metrics effort is well-timed as a testbed for these action items and for the overall strategic approach.

Shared conditions in the UCEC project included consortially-provided training in the Archivists' Toolkit (AT) and MPLP processing, though prior experience and implementation levels with both varied among the participating campuses; and the

use of and support from the Online Archive of California (OAC) as discovery system for the resulting finding aids.

The focus on metrics came later in the planning process than other elements of the UCEC project. The method of data gathering, therefore, needed to be lightweight for quick and easy implementation; it was conceived as an integrated component of the data capture planned for monitoring grant progress. The CHoM metrics database, MD, was cited as a resource for CLIR UCEC participants and interested campuses directed to contact its creators; however, there was no time to arrange training in MD or to encourage buy-in from grant participants, both key elements for success with such a system. No institutions opted to use it.

The metrics effort relied instead on self-reported data entered into a shared Google Docs Excel spreadsheet on the UCEC wiki. The spreadsheet offered a separate line for each collection to be processed, with 20 fields to be completed:

1. Repository
2. Collection
3. Scope/Content
4. Lin. ft. prior to processing
5. Condition rating prior to processing (PACSCL ratings, see below)
6. Condition notes prior to processing
7. Processing plan available?
8. Processing plan notes
9. Number of processing staff
10. Archival management/description application used? (E.g., Archivists' Toolkit, etc.)
11. Processing start date
12. Processing end date (arranged, described, and MARC/EAD records generated)
13. Processing level completed (Princeton ratings, see below)
14. Processing notes (pertaining to completion of processing)
15. MARC record publication date
16. EAD finding aid publication date (OAC)
17. EAD URL (in OAC)
18. Lin. ft. processed
19. Hours required to process 1 lin. ft.
20. What activities were included, within your calculation of lin. ft. processed, per hour?

Two numerical ranking systems were "embedded" in the spreadsheet (numbers 5 and 13 above): first, pre-processing condition ratings, borrowed from PACSCL (see appendix A); and second, completed processing levels for collections, borrowed from Princeton University Archives (see appendix B).

During semi-quarterly check-in calls with the CLIR UCEC Project Manager, participants were encouraged to enter data, but it was voluntary (as opposed to mandatory narrative and budget reports). Eight of the nine repositories have contributed data to the spreadsheet.

In addition, conference calls were conducted with all campuses (a campus visit in one case instead) by the project coordinator and CDL intern. Besides scanning the data fields listed in the spreadsheet, we asked a series of supplementary questions to paint the larger picture of opportunities and challenges encountered in processing.

Findings and interpretations

UCEC project overview

- 9 California institutions participating
- 42 discrete collections to be processed
- all collections concerning California's environment and environmental history
- 1102 total linear ft. before processing (a minimum, does not include oversized material and maps)
- 890 total linear ft. processed to date (a minimum, does not include oversized material and maps)
- average size of collection, pre-processing: 26 linear ft.
- 7 of 9 institutions have posted UCEC finding aids to the Online Archive of California (OAC) to date
- 25 finding aids total posted to OAC to date
- Date ranges of material represented in 25 posted finding aids, showing the bulk from the 1920s-1970s:

Number of collections with material from specific decades
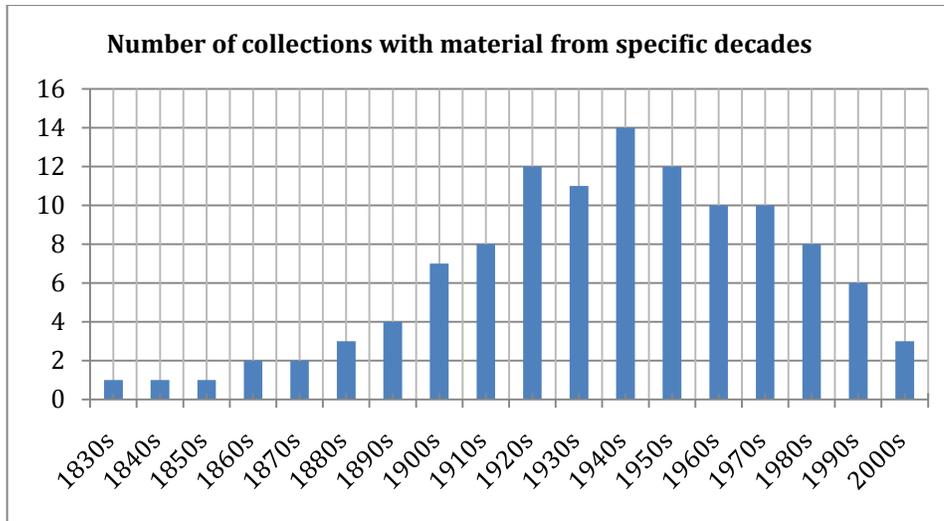
Figure 1: Date ranges for material represented in 25 completed collections

Analyzing a subset

In order to take a closer look with some of the most complete and unambiguous data, we considered a subset of seven repositories with a total of 17 completed collections:

- CSU Fresno:
  - National Land for People collection
- UC Berkeley
  - Edgar Wayburn papers
  - George Marshall papers
- UC Davis
  - Nikola P. Prokopovich papers
- UC Irvine
  - F. Sherwood Rowland papers
- UC Riverside
  - Citrus and horticulture collection
  - Archibald D. Shamel papers
  - Frederick Ferdinand Halma papers
  - Herbert J. Webber papers
  - William T. Horne papers
  - William L. Paul papers
- UCLA
  - Pinal Dome Oil Company records
  - Richard Gordon Lillard Papers
  - Unocal Corporation records
- USC
  - Citizens Committee to Save Elysian Park collection

- o   Mineral King Development collection
- o   Union Pacific Railroad Tidelands records

The metrics and narrative comments and conclusions that follow pertain to these 17 collections.

1. Workflow and staffing

Participants took either one or two steps, as shown below, to produce a valid EAD finding aid ready for the Online Archive of California (OAC):

- AT > OAC
- MS Access > OAC
- local EAD templates > AT > OAC
- EADExpress > WebGenDB > OAC

Though four different workflows are represented, they all use templating/data entry and conversion routines. Differences in workflows probably did not impact processing rates significantly as routines are well-established.

The average reported staffing level was .8 FTE per collection. In practice this meant that each institution had on average one professional archivist working with one student assistant, though one or both were not full-time. The students generally performed tasks such as labeling, interfiling, and transcribing into templates. The lead archivist was generally responsible for overall arrangement and high-level description including biography/history and scope and content notes. Creation of MARC records was generally not included in reporting but was reported by a number of repositories to take negligible amounts of time--these were often created by catalogers or other technical services staff and often involved cutting and pasting from description provided by processors.

2. Processing activities

Participants were asked to report their own processing rates, deriving figures for time spent by whatever tracking means they chose. We did, however, ask which of eight specific archival activities these repositories included in calculating a processing rate for each collection, and the results were as follows:

Surveying	71%
Accessioning	59%
Description	100%
Arrangement (rearranging, reboxing, refoldering, barcoding and/or folder labeling)	100%

Preservation (photocopying, sleeving, removing paperclips, etc.)	65%
Separating and/or de-accessioning materials	71%
Generating EAD	35%
Generating MARC	24%

While arrangement and description were universal, other activities were not. Producing EAD and MARC records were least included but, as noted above and in interviews, accounted for relatively little time. The remaining four activities were included by the majority of participants but require sharper definition; surveying and accessioning activities, for instance, may overlap depending on definition.
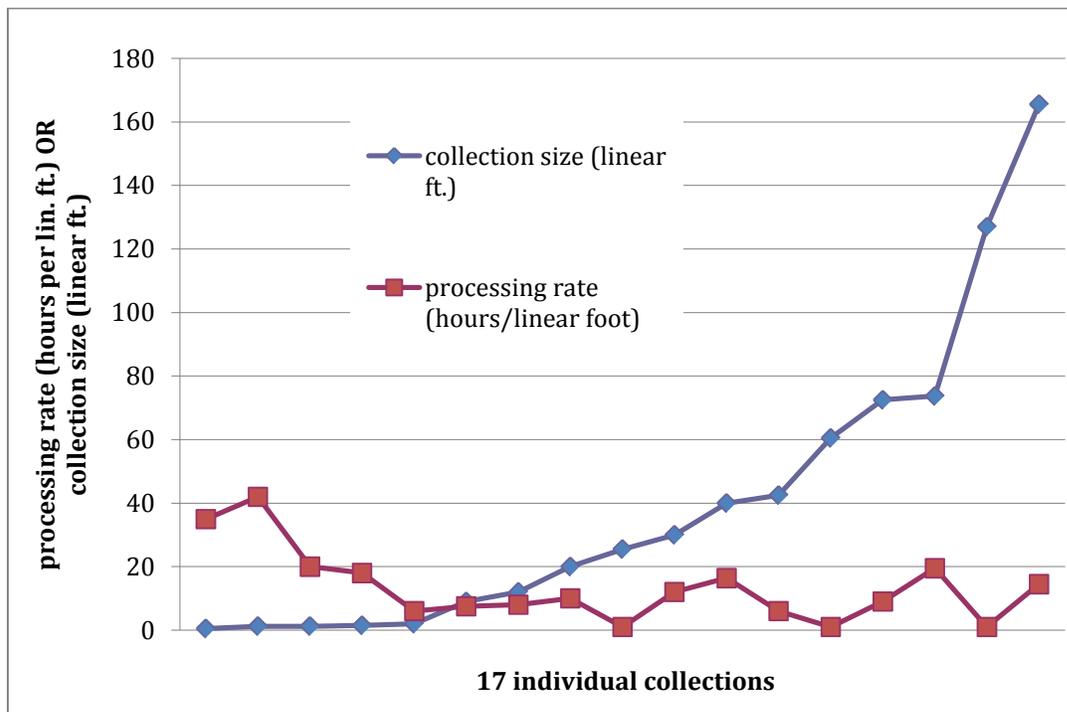
3. Processing rates

The average processing rate for these 17 collections was 13 hours per linear foot; the median, 10; the lowest, 1 hour per linear foot; and the highest, 42. In the couple of cases where repositories reported two processing rates for one collection, (e.g., one for the papers portion and another for photographic material), these have been averaged to derive one figure per collection. In what follows, most numbers have been rounded up to the nearest whole number.

For the purpose of this study, processing rates (in terms of hours per linear foot) were correlated with the following: the size and character of the collections; pre-processing condition of collections; and processing levels applied to collections. Of the three different correlations, the strongest correlation appears to be between processing rates and processing levels: the more intensive the description, arrangement, and/or preservation applied, the higher the processing rate.

This is discussed in further detail below (with the strongest correlations treated in Section 3.3).

3.1. Processing rate correlated with size and character of collection

The following chart shows the size of each of the 17 collections matched with its corresponding processing rate, combined on one scale for rough comparison:

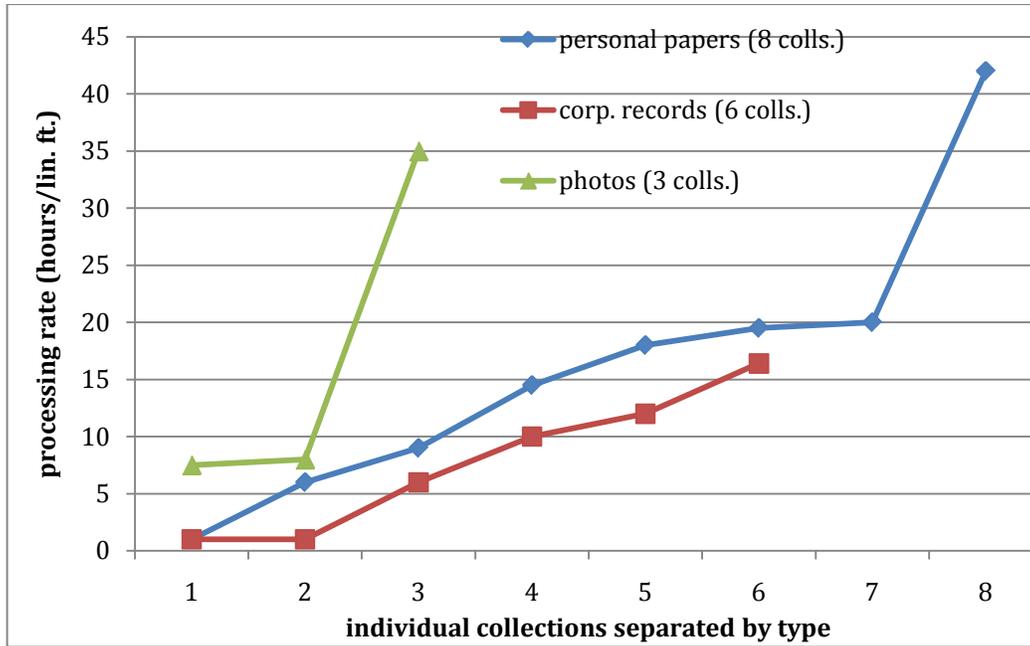**Figure 2: processing rates with collection sizes**

Collection size by itself clearly is not predictive of processing rates, though we might say there is a ramping-up factor that applies to the smallest collections as well as the largest. For the highest processing rates reported, i.e., the seven collections reporting rates above the average of 13 hours per linear foot, four were, in fact, the smallest collections, and all of them were collections for which one or more of the following conditions was true:

- archivist processing their first collection
- predominance of photographic content requiring preservation work and/or item-level description
- predominance of fragile material
- change of processing staff mid-project

The 17 processed collections may be further characterized as of one of three types: corporate records, personal papers, and collections with large photographic components. This simplification is based on collection title and creator combined with further information from interviews. The average processing rates for these three types were as follows:

- papers (8 collections, 1 with audiovisual content) = 16.3 hours/lin. ft.
- records (6 collections) = 7.7 hours/lin. ft.
- photographs (3 collections) = 16.8 hours/lin. ft.

The 17 collections separated into the three types are seen in the following chart mapped to their processing rates:

Figure 3: processing rates of 3 types of collection

Collections of personal papers and those with photographic content overall averaged higher processing rates than corporate records (50% of personal papers as opposed to 17% of corporate records had rates above the average). This finding is fairly intuitive: personal papers are often less well-ordered and -labeled than corporate records, and photographs tend to invite item-level processing.

3.2. Processing rate correlated with pre-processing conditions

Participants were asked to rate three aspects of their collections' condition before processing according to the scheme borrowed from the PACSCL project (see appendix A). Though the PACSCL ratings come with explanation, there is still ample rooom for local and contextual interpretation, a factor to be considered in comparing these ratings.

The following three tables with comment show these three pre-processing condition factors separated out and associated with processing rates across the group of 17 collections.

3.2.1 Original physical condition

original physical condition level (1=poor, 5=excellent)	percent of 17 collections reporting this level	average processing rate (hours/lin. ft.) for given rating	range of processing rates for given rating
1	0%	n/a	n/a
2	18%	21 hours	8-42 hours
3	41%	16 hours	8-35 hours
4	35%	6 hours	1-16 hours
5	6%	20 hours	20 hours

Table 1: original physical condition

The average reported physical condition rating was 3.45 out of 5, slightly above "medium". In the table above, there can be seen some association of poorer physical condition with most of the highest processing rates; most of the lowest processing rates are likewise associated with better physical condition, but in the middle, where most collections fall, it is hard to draw conclusions. Only one collection was reported to have been received in excellent original condition, and none were reported to be in poor condition.

3.2.2 Original physical access/arrangement

original physical access/arrangement level (1=poor, 5=excellent)	percent of 17 collections reporting this level	average processing rate (hours/lin. ft.) for given rating	range of processing rates for given rating
1	6%	6 hours	6 hours
2	29%	17 hours	1-42 hours
3	59%	14 hours	1-35 hours
4	6%	1 hour	1 hour
5	0%	n/a	n/a

Table 2: original physical access/arrangement

The original physical access to/arrangement of collection rating saw an average of 2.97 out of 5, slightly below "medium". One collection was given the lowest possible rating of "completely unarranged", but no collections were rated as excellent (full arrangement to item level) and only one as in "generally good order". 88% of collections fell squarely within "partial" to "rough" arrangement. As the associated processing rates range from the lowest to the highest within this majority, it is difficult to draw conclusions here.

3.2.3 Original housing condition

original housing condition level (1=poor, 5=excellent)	percent of 17 collections reporting this level	average processing rate (hours/lin. ft.) for given rating	range of processing rates for given rating
1	18%	9 hours	1-15 hours
2	18%	7 hours	1-20 hours
3	12%	6 hours	6 hours
4	47%	20 hours	8-42 hours
5	6%	9 hours	9 hours

Table 3**: original housing condition**

For original housing condition, the findings are a little counterintuitive. Processing rates overall rise somewhat with better original housing condition. Only one collection was reported received in excellent original housing, but almost half were received in good condition; this half correlates to the highest average processing rate. Replacing original housing may increase processing costs, but it does not seem to correlate to increased processing time.

In summary, these mid-20th century collections were reported overall to be in decent condition pre-processing. Preservation issues (largely related to format) tended to surface once the work began. The lowest processing rates (three collections at <5 hours per linear foot) correlate with good original physical condition, but not necessarily good original housing or good original arrangement. The bulk of processing rates fall between 5 and 20 hours per linear foot. Across this range, it is difficult to see clear correlations, with dips for both arrangement and housing conditions as well as peaks for both physical and housing conditions. The three collections with highest processing rates, however, correlate to somewhat poorer arrangement and physical condition, as might be expected.

3.3. Processing rate correlated with processing levels

Participants were asked to rate the level of processing they applied to their collections in three categories borrowed from Princeton University Archives (see appendix B). As with pre-processing condition ratings, explanation was provided for applying the ratings. The standards for these three elements, description, arrangement, and preservation action, were somewhat more cut and dried perhaps than those for original condition and the data therefore somewhat more reliable. Where participants listed two ratings for one collection (based on different treatments for different series, usually), the higher rating has been selected to retain the scale of the rating (rather than parsing it into decimalized shadings).

3.3.1 Description processing level

description level (1=collection level, 3=multi-level to files/items)	percent of 17 collections reporting this level	average processing rate (hours/lin. ft.) for given description level	range of processing rates for given description level
1	0%	n/a	n/a
2	24%	5 hours	1-10 hours
3	76%	16 hours	1-42 hours

Table 4: description level

The average of description processing levels reported was 2.7 out of 3. The description processing level correlates to processing rate as might be expected: all participants ranked their description between 2 and 3 on a scale from 1-3, and the only instances of a descriptive ranking of less than 3 correlate to processing rates under the average. Higher description inputs correlate here to higher processing rates.

3.3.2 Arrangement processing level

arrangement level (1=none, 4=folder level)	percent of 17 collections reporting this level	average processing rate (hours/lin. ft.) for given arrangement level	range of processing rates for given arrangement level
1	0%	n/a	n/a
2	0%	n/a	n/a
3	24%	7 hours	1-20 hours
4	76%	15 hours	1-42 hours

Table 5: arrangement level

Likewise, for the arrangement processing level, ranked from 1-4, the only instances of rankings under 4 line up with lower processing rates, with two exceptions: UC Davis, with a higher processing rate (20, rounded up average) but challenging photographic component in the Prokopovich papers, and UC Irvine, again with a slightly higher processing rate (15, rounded up average) but significant audiovisual component in the Rowland papers. Higher arrangement inputs tend to correlate to higher processing rates, but it is notable that this includes cases with challenging formats that were largely responsible for higher processing rates.

3.3.3 Preservation action level

preservation level (1=none, 5=reboxed and refoldered; A/V, photos, folded items, and other formats fully addressed)	percent of 17 collections reporting this level	average processing rate (hours/lin. ft.) for given preservation level	range of processing rates for given preservation level
1	0%	n/a	n/a
2	0%	n/a	n/a
3	12%	4 hours	1-6 hours
4	88%	15 hours	1-42 hours
5	0%	n/a	n/a

Table 6: preservation action level

As an archival activity that is expected to be reduced in rigorous applications of MPLP, the preservation processing level metric, on a scale of 1-5, is interesting: participants recorded a uniform value of 4 except for a few values of 3 correlating to processing rates well under the average. Given that a few institutions reported two values to reflect parts of collections requiring different inputs of preservation activity (which were rounded up to derive one figure per collection), here, too, higher processing inputs still correlate with higher processing rates.

In summary for these processing level data points, the average description level tips toward file- or item-level description, the arrangement level average toward folder-level, and the preservation level verges on everything reboxed and refoldered with all formats accommodated. These are high levels of processing. In some cases folder- or item-level treatment was noted and justified in processing notes; the project-wide MPLP approach probably heightened awareness of "going over the top" with processing. Correspondingly higher processing rates are seen for the highest-ranked processing inputs.

3.4. Formats

An aggregation of environmental collections is very likely to include demanding formats and unexpected materials from the field. This profile was borne out in the UCEC project. Four of the seven repositories encountered formats that were unexpected in quantity and/or condition and which impacted processing rates: photographs, nitrate negatives, glass plate negatives, lantern slides, blueprints, and analog audiovisual material. The photographic and audiovisual materials were, in general, processed to the item-level, deemed necessary to provide meaningful access. The presence of nitrate negatives in one case required varieties of expert assistance and additional time. Staff members at UC Davis consulted with a

photographic specialist when a large number of slides was discovered in the Prokopovich papers; they were thereafter able to take an MPLP approach to the slides, a surprising and instructive outcome.

There was no separate tracking of processing actions or MPLP approaches specifically for these demanding formats. We seem to be in early days vis-à-vis the application of MPLP to special formats and the sharing of examples of such application; format-based best practices would be a boon to the profession going forward.

3.5. Other observations

A number of collections were reported to be simply so important that greater detail was warranted (for instance, in the case of collection creators being important faculty at the institution). Greene and Meissner warned that grant-funded projects can see higher processing rates associated with perceived collection importance; it is true that these environment-related collections were valued highly for their research potential. This is to some degree a catch-22 in our current grant-driven environment: the mandate to use MPLP as a tool to expose hidden collections expeditiously alongside the opportunity to process thoroughly with funds in hand. The catch may resolve itself if more processing and collection use data is shared; both repositories and funders may know better what to expect.

Most participants reported that they did not programmatically track processing rates prior to the CLIR UCEC project. In estimating resources and budgets needed to process particular collections, various participants indicated that they derived allocations based on work on previous collections processed by experienced archivists. In a few cases, repositories had documented benchmarks for budgeting purposes. This is not surprising, given how uncommon measured rates are. A few repositories reported that, due to underestimates of funding needed, some of their work was a donation in kind beyond what was budgeted.

To the related question, whether participants maintain a processing manual or not, six of the seven repositories reported maintaining a processing manual or consulting another repository's manual. The answers reflect an awareness of "best practices" alongside real-world circumstances that make codifying and carrying them out sometimes difficult. While a processing manual does not guarantee quality or quantity of work, engaging staff in discussion of processing decisions can aid consistency and therefore planning.

Selected participant profiles

We should note that three institutions, Humboldt State University, UC Davis, and UCLA, presented detailed case studies of their UCEC processing experiences at the

Society of California Archivists' 2011 Annual General Meeting in April, 2011, and these are available online from the UCEC wiki.

Bancroft Library, UC Berkeley
Two collections processed by the Bancroft, the George Marshall Papers and the Edgar Wayburn Papers, 42 and 72 linear feet respectively, were both received in good condition. Processing rates of 9 and 6 hours per linear foot respectively were achieved. A very experienced staff archivist was assigned to the UCEC project part-time, processing at folder-level with part-time help for container listings. Bancroft uses well-established homegrown software in lieu of the Archivists' Toolkit. A previous (2008) Mellon-funded collection survey had provided processing plans and MARC records. In conversation, David de Lorenzo, Associate Director and Head of Technical Services, noted that MPLP beneficially allows calling collections "done" that might previously have languished on backlog lists. He also notes that till now they have used the Beinecke Library's standards to estimate grant costs; and that beginning to track processing metrics (i.e. in an online spreadsheet/database) has the added benefit of reinforcing the "data-ness" of current archival workflow for processors.

Langson Library, Special Collections and Archives, UC Irvine
Irvine processed the F. Sherwood Rowland Papers, 166 linear feet, received in poor condition, without order, and few existing folders (and followed, after processing began, by an unanticipated donation including VHS tapes). The overall processing rate was 14.5 hours per linear foot. Series were processed at different levels of detail, depending on their research value and the work required to make them minimally usable. Item-level sorting was necessary to place random papers into identified sub-series, and processors had to supply and label folders, usually at the sub-series level. The audiovisual material, with high research value, was reformatted and described at the item level. The extreme disorder of the papers and the need to migrate legacy audiovisual formats proved to be time sinks. That said, data entry into the Archivists' Toolkit went quickly. Irvine maintains a detailed cost matrix for processing, but the matrix didn't account for doing a minimal level of processing for a collection with no existing order. Extrapolating from other rates, however, the matrix predicted a rate of 15 hours per linear foot. Then-archivist Dawn Schmitz describes the collection as "minimally processed, with caveats".

Tomás Rivera Library, Special Collections & Archives, UC Riverside
Riverside has to date processed six collections totaling 39 linear feet, received in decent shape but containing a variety of photographic formats, including nitrate negatives; each collection was different, but the average processing rate was 22 hours per linear foot. MPLP was said to be "eye-opening" for them and currently more of an aspiration than a practice; item-level processing had been the norm. Dealing with nitrate negatives and lantern slides skewed their predicted processing rates (from MPLP), as did a distracting communal processing space and training new staff on the use of the Archivists' Toolkit. Nonetheless, work has been completed on time and a seventh, larger, collection added to the project.

<u>Charles E. Young Research Library, Department of Special Collections, UCLA</u>
UCLA processed three collections (Unocal, Pinal Dome, and Richard Lillard papers), a total of 213 linear feet, with a consistent processing rate of one-two hours per linear foot, in part by avoiding deep preservation action and deaccessioning. They used the first of the collections to establish a baseline processing rate and to adopt an appropriate hybrid MPLP approach; MPLP was well-received by Special Collections staff who attended the training. They have since been able to approach a couple of large backlog collections with fresh perspective. For these three collections, photographs were described to the folder level; series abstracts were created; their MS Access database was converted to EAD; and research was conducted for the front matter. UCLA's context includes the Center for Primary Research and Training which "...integrate[s] special collections materials more fully into the teaching and research mission of the university" by having graduate students process collections in depth. The CFPRT offers an interesting alternative source of processing data.

Conclusions

- "Processing" includes arrangement and description universally, other activities less consistently--there is room for more precision
- Size of collection is not a factor in predicting processing rates
- Type of collection, as determined from creator and format, is weakly predictive of processing rates
- Pre-processing conditions for these 20th century collections correlated at the lowest and the highest processing rates but inconclusively in the middle range
- Processing inputs of description, arrangement, and preservation clearly correlate with processing rates--an especially reliable data point since all participants included these activities in defining "processing"
- MPLP played a role for all participants (if only one of raising awareness), but overall, collections were described, arranged, and preserved to a high level
- Photographic and audiovisual content correlates on multiple planes to higher processing rates (description and preservation particularly)

There are, of course, issues to raise about the value of self-reported, non-standards-based data:

- Terminology and units of measure were not vetted with participants
- The two borrowed ranking systems (PACSCL and Princeton) required further explanation and were somewhat subjective
- Different experience levels with AT and MPLP un-evened the playing field

- Staff time may have included multi-tasking and/or work on more than one collection simultaneously

Nonetheless, the CLIR UCEC metrics sub-project has been a valuable experiment. With very little investment upfront, it has introduced consortial data collection to colleagues across the state of California; it gathered baseline numbers and impressions with which to take the process further in future; and it joined in community with other projects doing innovative processing across the nation.

Looking forward, at ground level, it is clear that photographic, geographic, and audiovisual material formats call for more focused MPLP and processing metrics case studies. At a higher level, commitment to early planning will produce useful numbers across future consortial projects--not an onerous burden if the tools currently available are given a trial.

There are larger issues about gathering data that lie beyond the UCEC project. How we define MPLP or "minimal processing" is crucial to sharing meaningful data. As a framework for approaching processing, "minimal" is relative to local practices; the case study method may prove a more useful way of building up a shared body of knowledge about processing approaches in context than a comparison of numbers.

The processing rate metric is just one along a spectrum of measurables, and processing effectiveness, gauged by user satisfaction, lies at the other end. The CHoM metrics tool, employed to pool data and thereby inform decision-making, is intended to lead us to "more fully evaluate the relationship between appropriate-level processing and user satisfaction." The Archival Metrics project run by the University of Michigan, the University of North Carolina-Chapel Hill, and the University of Toronto, has developed toolkits to assess users' research needs and repository effectiveness, promoting a "culture of assessment in the archival domain". What we might best aim for is an assessment cycle. As we learn more from efforts to assess user satisfaction, we will know better where to spend and measure our time beyond processing, for example in subject analysis, provision of crowdsourcing tools, and varieties of publicity.

Given that the data for the UCEC project is keyed to the hours-per-linear-foot metric, it is worth asking whether this is the right or even a good metric to track. It is certainly not the only right metric, but it seems a good one. The effort involved in deriving such a figure offers some value in self-awareness. Archival sensitivity about individual performance may mean that a higher-level, derived figure like dollar cost per linear foot should be the reported metric. The data gathered here suggest that hours-per-linear-foot might be most useful if broken out further by format. It may be useful ultimately to look across data for subsets of different subject areas and/or types of collection.

For most archival shops, a processing rate ultimately will be used to calculate per-collection dollar costs, whether for grant requests, reporting up the funding chain,

or to notify donors of collection processing costs. Processing rate is surely a useful component of that final figure and is especially achievable with the tools currently in development. It is trackable by processors themselves as work is done rather than being dependent on post-processing review; with the aid of more precise definitions (something like a data dictionary), it can achieve more granular accuracy. It is harder to quantify administrative time spent on processing projects, but ideally this should be factored in for the numbers to be comprehensive.

Until we can claim no hidden collections and probably thereafter, metrics can illuminate key intersections of time, staffing, space, and supplies with funding. Though it feels like it when in the midst, the metrics are not an end in and of themselves; they are a means of advancing the conversation with colleagues about return on investment in an era of conflicting demands and limitations.

Recommendations for employing metrics

- Know your and your funder's goals vis à vis metrics
- If there are other consortia/institutions working within the same grant program, consult with them on best practices early
- Know your repositories/departments and their processing culture
- Encourage the use of surveys and processing plans to preview difficult formats and preservation issues in the collections to be processed
- If challenging formats emerge, consult experts, and share resulting practices
- Know the curricula of the schools from which potential newly-graduated project processing archivists will come, if possible
- Tailor a version of MD (or another tool as may emerge) to the project in hand, consulting with the tool's creators
- Include an introduction to metrics and MD (or other tool/database) training along with MPLP processing or other training
- Insure that the metrics scheme employed does not resemble individual performance evaluation
- Derive a tailor-made-to-project vocabulary or data dictionary if needed, drawing on national standards as available
- Encourage the use of standard, well-supported tools for EAD and MARC creation with appropriate training
- Insure that metrics-related questions have a forum and a point person for the benefit of all participants
- Reach out to participants periodically to check on progress and catch uncertainties and non-conforming data early
- Ask processors to blog or otherwise record processing experiences quickly, frequently, and informally (e.g., weekly posts)
- Plan to share your metrics; know who you will be sharing your metrics with and what benefit you expect from the sharing

- Plan for a wrap-up report, meeting, call, or other summary activity to demonstrate value, reward participants, and publicize outcomes
- Statistics can be important in the field; acquire experience working with them

Appendix A: Condition ratings
Adapted from PASCSL Consortial Survey Initiative:
http://www.pacsclsurvey.org/documents/ratings.pdf

Physical condition rating
5. Excellent: little damage with no further deterioration expected
4. Very good: little damage, with some further deterioration possible, due to the mixed quality of the material
3. Good: expected deterioration with some further deterioration possible
2. Fair: somewhat worse than expected deterioration, with some further deterioration possible
1. Poor: significant damage/deterioration that makes collection difficult to use

Physical access rating
5. Full arrangement to item level in series and, as appropriate, subseries
4. Arrangement in series to file level. There is generally good order within the files
3. Rough arrangement by date, document type, function, source, or other characteristic; papers not thoroughly screened, but have been unfoldered and flatted; series not fully established.
2. Partial or superficial arrangement and/or non-standard housing and labeling.
1. Totally unarranged; many, sometimes most, documents not yet removed from envelopes, unfoldered, and flattened

Quality of housing rating
5. Collection housed completely in acid-free boxes and folders in good condition. Boxes and folders have reasonable amount of material in them. Boxes and folders are correct size and type for the materials they house.
4. Collection housed partially in acid-free boxes and folders in good condition. Most boxes and folders have reasonable amount of material in them. Most boxes and folders are correct size and type for the materials they house.
3. Collection housed in non-archival boxes and folders, but they are in good condition. Most boxes and folders have reasonable amount of material in them. Most boxes and folders are correct size and type for the materials they house.
2. Collection housed in non-archival boxes and folders. Significant number of boxes and folders might have unreasonable amount of material in them or are not the correct size and type for the materials they house.
1. Collection housed in non-archival boxes. Majority of material is not in folders and/or boxes are overstuffed or understuffed.

Appendix B: Processing levels
Adapted from Princeton University Archives' Processing Levels Summary

Description
3. Multi-level description to file- or item-levels
2. Multi-level description to series or subseries
1. Collection-level description

Arrangement
4. Arranged at folder level
3. Arrangement at series or subseries level - includes arrangement of folders within series or subseries
2. Arrangement at series or subseries level - no arrangement of folders within series or subseries
1. None

Preservation
5. Reboxed and refoldered. A/V, photos, folded items, or other formats are fully addressed.
4. Reboxed and refoldered. Minimal treatment of items within folders
3. Reboxed only. Treatment of exceptional items within folders (those obviously noticeable -- moldy, etc.)
2. Reboxed only.
1. None

Appendix C: Bibliography

Abraham, T., Balzarini, S. E., & Frantilla, A. (1985). What Is Backlog Is Prologue: A

Measurement of Archival Processing. *The American Archivist*, *48*(1), 31.

Beinecke Rare Book and Manuscript Library (1997). Archival Processing Manual.

Retrieved from

http://www.library.yale.edu/beinecke/manuscript/process/index.html

Canadian Council of Archives. (2011). National Archival Development Program:

NADP Applicant's Guidebook. Retrieved from

http://www.cdncouncilarchives.ca/NADP_11-12/ApplicantGuidebook11-

12_Y6.0_EN.pdf

Conway, M. O. & Proffitt, M. (2011). Taking Stock and Making Hay: Archival

Collections Assessment. Dublin, Ohio: OCLC Research. Retrieved from

http://www.oclc.org/research/publications/library/2011/2011-07.pdf

Cox, R. (2010). Maximal Processing, or, Archivist on a Pale Horse. *Journal of Archival

Organization*, *8*(2), 134-148.

Crowe, S. (2010). MPLP @ 5: More access, less backlog? *Journal of Archival

Organization*, *8*(2), 110-133.

Ericksen, P. (1995). Beneficial Shocks: The Place of Processing-Cost Analysis in

Archival Administration. *The American Archivist*, *58*(1), 32.

Gorzalski, M. (2008). Minimal Processing: Its Context and Influence in the Archival

Community. *Journal of Archival Organization*, *6*(3), 186-200.

Greene, M. & Meissner, D. (2005). More Product, Less Process: Revamping

Traditional Archival Processing. *The American Archivist*, *68*(2), 208-263.

Greene, M. (2010). Perspective - MPLP: It's Not Just for Processing Anymore. *The*

American Archivist, *73*(1), 175.

Gustainis, E. (2010). Processing Metrics Collaborative: Database Development

Initiative. Retrieved from

https://wiki.med.harvard.edu/Countway/ArchivalCollaboratives/ProcessingM

etricsDatabase

Hackbart-Dean, P. & Slomba, E. (2009). Association of Research Libraries SPEC Kit

314, *Processing Decisions for Manuscripts & Archives.*

Lynch, K. T. & Lynch, T. E. (1982). Rates of Processing Manuscripts and Archives.

Midwestern Archivist, 7(1), 25-34.

Maher, W. J. (1982). Measurement and Analysis of Processing Costs in Academic

Archives. *College & Research Libraries*, 43(1), 59-67.

McCrea, D. (2006). Getting More for Less: Testing a New Processing Model at the

University of Montana. *The American Archivist*, 69(2), 284-290.

Meissner, D. & Greene, M. (2010). More Application While Less Appreciation: The

Adopters and Antagonists of MPLP. *Journal of Archival Organization*, 8:3-4, 174-

226.

Mengel, H. (n.d.). PACSCL Hidden Collections Processing Project Blog. Retrieved

from http://clir.pacscl.org/blog/

Mengel, H. (2010 January 13). End of Year Report: 2009. Retrieved from

http://clir.pacscl.org/2010/01/13/end-of-year-report-2009/

Mengel, H. (2011). Collaborating to Achieve Quality Work in a Minimal Processing

Environment. *RBM* 12(1), 41-50.

Mengel, H. (2011 January 7). Ringing in the New Year ... end of year report and

    future plans. Retrieved from http://clir.pacscl.org/2011/01/07/ringing-in-the-

    new-year-end-of-year-report-and-future-plans/

Northwest Archives Processing Initiative Phase II: A Final Narrative Report

    Submitted to The National Historical Publications and Records Commission For

    the grant period July 1, 2005-June 30, 2007 (2007). Retrieved from

    https://www.google.com

Patty, W. (2008). Metadata, Technology, and Processing a Backlog in a University

    Special Collections. *Journal of Archival Organization*, *6*(1-2), 102-120.

Regents of the University of California (2011). The CLIR Project Wiki - Home.

    Retrieved from https://wiki.ucop.edu/display/CLIR/The+CLIR+Project+Wiki+-

    +Home

Regents of the University of California (2011). Next-Generation Technical Services

    (NGTS). Retrieved from

    http://libraries.universityofcalifornia.edu/about/uls/ngts/index.html

Regents of the University of California (2011 April 21). SOPAG: NGTS

    Implementation Initiative Framework. Retrieved from

    http://libraries.universityofcalifornia.edu/sopag/ngtsframework.html

Slotkin, H. (1982). An Analysis of Processing Procedures: The Adaptable Approach.

    The American Archivist, *45*(2), 155-163.

University of Michigan, University of North Carolina-Chapel Hill & University of

    Toronto (2011). Archival Metrics | Promoting a Culture of Assessment in

Archives and Special Collections. Retrieved from

http://archivalmetrics.cms.si.umich.edu/

Walters, E. (n.d.). Changing the Landscape. Retrieved from

http://news.lib.ncsu.edu/changinglandscape/

Weideman, C. (2006). Accessioning as Processing. *The American Archivist, 69*(2),

274-283.