Journal of Western Archives

Volume 11 | Issue 1 Article 1

2020

Balancing the Art and Science of Archival Processing Metrics and Assessment

Cyndi Shein
University of Nevada, Las Vegas, c.shein@yahoo.com

Sarah R. Jones
University of Nevada, Las Vegas, sarah.jones1@unlv.edu

Tammi Kim
University of Nevada, Las Vegas, tammiyk@gmail.com

Karla Irwin
University of Nevada, Las Vegas, karla.irwin@unlv.edu

Balancing the Art and Science of Archival Processing Metrics and Assessment

Cover Page Footnote

The authors would like to thank James Cheng, UNLV Libraries' Library Data Analyst, for his counsel in how to appropriately present the analysis of their field data.

Balancing the Art and Science of Archival Processing Metrics and Assessment

Cyndi Shein
Sarah R. Jones
Tammi Kim
Karla Irwin

ABSTRACT

At the University of Nevada, Las Vegas (UNLV) University Libraries, archival processing metrics are used to support value propositions, project proposals, project management, and strategic planning. When making data-driven decisions, UNLV Special Collections Technical Services staff strive to balance the art and science of archival processing metrics—to critically assess their data and look beyond the numbers for additional information that brings meaning to the metrics. In this research paper, the authors review processing metrics across the profession and place their own archival processing field data within the context of more than three decades of professional practice. They report and explore UNLV's processing rates and the variables that potentially influence their rates. Learning from methods piloted during special projects, they reach toward a more streamlined, sustainable assessment practice. They also include a "work in progress"—an experimental framework that suggests core processing data points for UNLV with a second tier of optional data points that may add value to metrics in specific circumstances. The flexibility and extensibility of the framework give it the potential to serve as a model for other repositories.

Gathering and assessing archival processing metrics is both an art and a science. There is an exact science in numbers—in calculating averages, rates, and medians. Comparatively, there is an art to interpreting numbers and bringing meaning to the stories they tell. Recent years have witnessed an increasing expectation for archivists and librarians to quantify the value of their work and programs. In the absence of profession-wide archival processing metrics standards, archivists must find their own ways to develop the science of collecting data and master the art of interpreting it. The University of Nevada, Las Vegas (UNLV) University Libraries Special Collections and Archives assesses data about collection use, patron visits, online views of digital collections, and archival processing to inform their services and operations. The authors, archivists at UNLV Special Collections and Archives Technical Services (hereafter, UNLV), begin by reviewing three decades of processing metrics across the

profession. They then critically evaluate UNLV's processing data against the backdrop of professional practice and assess UNLV's metrics within the context of the variables that influence the data.

Over the past four years, UNLV staff piloted different processing assessment methods during special projects, collecting only four data points for early projects, and upwards of twenty data points for recent projects. They periodically assessed snapshots of their processing metrics to manage projects, demonstrating progress in making archival resources discoverable and accessible, and communicating the value of their work to stakeholders. While the metrics have been useful, the data sets were inconsistent, noisy, and difficult to assess as a whole. Looking to standardize and operationalize their assessment practices, UNLV aggregated processing metrics from their various projects, added what metrics were available from daily operations, and critically examined their data collection and assessment methods.

With an eye toward sustainability, UNLV saw the need for a streamlined assessment framework that would more efficiently return the most vital information. In the same way that archivists might strive for the "golden minimum" in their processing efforts, UNLV attempted to find the golden minimum in their assessment practices.[1] Discussions among UNLV staff revealed that each functional area had different assessment priorities. Data collection frameworks developed for projects were not universally useful, and processors were inventing data points to meet specific needs in their own areas (processing during accessioning, born-digital processing, oral history processing, manuscript processing, and special projects). UNLV needed to answer a number of questions. *What are we trying to learn from the data? What is the purpose of each data point? Which points are absolutely vital? Which points provide added value under certain conditions?* UNLV sought answers by reviewing professional literature and tools, scrutinizing their own field data, conducting an opinion poll, and mapping out an experimental framework of data points for themselves.[2] Their extensible framework suggests core processing data points designed to address fundamental processing assessment needs and includes a second tier of optional data points that add value to metrics in specific circumstances. The flexibility and extensibility of the framework give it the potential to serve as a model for other repositories.

Literature Review

Concerted efforts to analyze statistics related to special collections and archives date back more than half a century, and current literature reflects the ever-growing

1. The "golden minimum" is a concept that focuses on doing only what is necessary to achieve an objective—nothing more. Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *The American Archivist* 68, no. 2 (2005): 208-263, 240.

2. UNLV conducted a survey to gather professional opinions about archival processing metrics in autumn, 2019. Results of the poll will be submitted for publication in 2020.

"culture of assessment" across the library and archives profession.[3,4] The literature continues to demonstrate the importance of assessment. As summarized by Lisa Carter, "A culture of assessment focused on creating and communicating value provides direction, verifies strategies, and indicates improvements needed to effectively leverage our special collections for greatest impact."[5] Research suggests that assessment is most effective when undertaken holistically, combining metrics from front-facing services with metrics from behind-the-scenes operations. Griffin, Lewis, and Greenberg advocate for a "comprehensive, integrated assessment strategy."[6] Likewise, Daines and Brightenburg underscore the wisdom of assessing data from both technical services and user services to support evidence-based decisions that improve user services and inform collection management decisions. Carter notes that in addition to measuring special collections and archives' impact on teaching, learning, and research, "Calculating the costs associated with effective discovery, access, and engagement is just as critical to constructing the [value] proposition."[7]

Measuring the archival processing actions that lead to discovery, access, and engagement is a recurring topic in the literature. Processing productivity rates (quantifying the bytes or feet processed per hour/week/month) are at the heart of this assessment. Professional discourse consistently affirms the significance of understanding processing productivity in relation to budgets and resource allocation.[8] An abundance of literature, primarily written from within the context of

3. Christian Dupont and Elizabeth Yakel, "'What's So Special about Special Collections?' Or, Assessing the Value Special Collections Bring to Academic Libraries," *Evidence Based Library and Information Practice* 8, no. 2 (2013): 12.

4. J. Gordon Daines, III and Cindy R. Brightenburg, "Jumping In: Creating an Assessment program for the L. Tom Perry Special Collections Reading Room," *Journal of Western Archives* 10, no. 2 (2019). See literature review: 2-4.

5. Lisa R. Carter, "Articulating Value: Building a Culture of Assessment in Special Collections," *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 13, no. 2 (2012): 99.

6. Melanie Griffin, Barbara Lewis, and Mark I. Greenberg, "Data-Driven Decision Making: An Holistic Approach to Assessment in Special Collections Repositories," *Evidence Based Library and Information Practice* 8, no. 2 (2013): 225.

7. Carter, "Articulating Value," 91.

8. Studies on using processing metrics to inform priorities and demonstrate collection management costs include: W. N. Davis Jr., "Budgeting for Archival Processing," *The American Archivist* 43, no. 2 (Spring 1980): 209-211; Paul Ericksen and Robert Shuster, "Beneficial Shocks: The Place of Processing-Cost Analysis in Archival Administration," *The American Archivist* 58 (Winter 1995); Emily R. Novak Gustainis, "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, no. 2 (2012): 113-28; and Emily Walters, "Processing Large-Scale Architectural Collections," *Journal for the Society of North Carolina Archivists* 10, no. 1 (Fall 2012): 20-51.

grant work, emphasizes the importance of processing metrics.[9] Fewer articles, however, take a deep dive into the details of assessment, and readers are left wondering a number of things: *Which data points proved essential and how was the data used? How did processing rates estimated at the start of projects compare with the actual productivity rates? Did processing assessment methods developed during grant projects inform ongoing operations and, if so, were the methods scalable and sustainable?*

Although the literature provides models for estimating and/or measuring processing activities, the archival profession has no universally accepted processing productivity frameworks or assessment standards. Conceptual models based on tiered levels of collection complexity and processing intensity provide adaptable frameworks for estimating local processing rates for physical materials, but there are no such models for born-digital materials.[10] Some authors have written about capturing metrics for electronic materials, as well as the need for understanding the unique variables that affect the processing and management of these records; however, consensus on a standard set of data points has not yet been reached.[11] The extreme discrepancy in processing rates reported across the profession suggests that universal processing benchmarks are unrealistic. Since processing rates are the bedrock of processing assessment and no profession-wide standards exist, it is critical for repositories to establish local guidelines and benchmarks in order to improve the predictive power of their own time and cost projections.

The complexities of establishing processing rate benchmarks are illustrated by the broad spectrum of reported productivity rates. Ericksen and Shuster compared

9. See Karen Temple Lynch and Thomas Lynch, "Rates of Processing Manuscripts and Archives," *The Midwestern Archivist* 7, no. 1 (1982): 25-34; Helen W. Slotkin and Karen T. Lynch, "An Analysis of Processing Procedures: The Adaptable Approach," *The American Archivist* 45, no. 2 (Spring 1982): 155-163; Richard W. Hite and Daniel J. Linke, "Teaming Up with Technology: Team Processing," *The Midwestern Archivist* 15, no. 2 (1990): 91-97; Gustainis, "Processing Workflow Analysis for Special Collections"; Adrienne Pruitt, "Processing by the Numbers: How Metrics Can Help With Project Planning," presented at the Mid-Atlantic Regional Archives Conference, Session S18, October 27, 2012, Richmond, Virginia; Walters, "Processing Large-Scale Architectural Collections"; Anne L. Foster, "Minimum Standards Processing and Photograph Collections," *Archival Issues* 30, no.2 (2006): 107-118; and Cheryl Oestreicher, "Personal Papers and MPLP: Strategies and Techniques," *Archivaria* 76 (Fall 2013): 93-110.

10. One model for estimating processing rates is found in *University of California Libraries, Guidelines for Efficient Archival Processing in the University of California Libraries*, September 18, 2012, https://libraries.universityofcalifornia.edu/groups/files/hosc/docs/_Efficient_Archival_Processing_Guidelines_v3-1.pdf (accessed February 24, 2020), 23.

11. See Laura Wilsey, Rebecca Skirvin, Peter Chan, and Gylnn Edwards, "Capturing and Processing Born-Digital Files in the STOP AIDS Project Records: A Case Study," *Journal of Western Archives* 4, no. 2 (2013); Lisa Calahan and Carol Kussmann for the University of Minnesota Electronic Records Task Force, Electronic Records Task Force Phase 2 Final Report, August 2017, http://hdl.handle.net/11299/189543 (accessed February 24, 2020); and Dorothy Waugh, Elizabeth Russey Roke, and Erika Farr, "Flexible Processing and Diverse Collections: A Tiered Approach to Delivering Born Digital Archives," *Archives and Records* 37, no. 1 (2016): 3-19.

studies published from 1976 to 1987 and extrapolated processing rates ranging from 1 to 40 hours per cubic foot.[12] Since then, in spite of developments that might be expected to normalize practices to some degree, processing rates continue to vary greatly.[13] More recently, the Northwest Archives Processing Initiative (NWAPI) project reported an average rate of 2.9 hours per linear foot; a North Carolina State University (NCSU) Libraries' project averaged approximately 3.5 hours per linear foot, with rates ranging from 0.28 to 15.43 hours per linear foot; and a University of California (UC) Libraries project reported an average processing rate of 13 hours per linear foot, with rates ranging from 1 to 42 hours per linear foot.[14] Continuous disparities in processing rates suggest that archivists must look beyond the numbers to understand the differences; they must examine the variables that influence the numbers.

Variables Influencing Processing Rates

The numbers alone do not tell the whole story. Disparities in the multitude of published archival processing rates can be attributed to many factors, including differences in data collection methods, collection characteristics, human factors, and the nature of the work being performed. The primary reason that processing rates across the profession are difficult to compare is that rates are based on different units of measure, different processing methods, different data points, and different data collection methods. For example, a Washington State University project roughly estimated rates retrospectively by extracting select data points from eight years of processing worksheets, whereas Harvard's Countway Library systematically tracked 49 data points throughout a three-year project.[15] Incongruities in methodologies are

12. Ericksen and Shuster, "Beneficial Shocks," 41.

13. Describing Archives: A Content Standard (DACS) was adopted by the Society of American Archivists in 2004. The first version of the Encoded Archival Description (EAD) standard was released in 1998. Open source content management applications have been available for more than a decade. Examples include Archon (first released in 2006) and Archivists Toolkit (released for beta testing in 2006), both of which were merged and superseded by ArchivesSpace in 2013.

14. NWAPI rates were reported in Janet Hauck, Rose Slinger Krause, and Kyna Herzinger, "MPLP Ten Years Later: the Adventure of Being among the First," *Provenance* 35, no. 2 (2018): 71-123, 87. Linda Sellars provided a spreadsheet of hourly processing rates for the NCSU project via email on September 3, 2019. UC rates were reported in Melanie Wisner, *Uncovering California's Environmental Collections: A Collaborative Approach* (CLIR UCEC), Processing Metrics Report, October 2011, https://www.clir.org/wp-content/uploads/sites/6/UCECfinalRev.pdf (accessed February 24, 2020), 9.

15. Terry Abraham, Stephen Balzarini, and Anne Frantilla, "What Is Backlog Is Prologue: A Measurement of Archival Processing," *The American Archivist* 48, no. 1 (Winter 1985). Abraham et al. gleaned limited data from eight years of processing worksheets as described on pages 34-36. Gustainis' project team collected 49 data points in total, including activities from acquisition to digitization. See pages 115-117 for a complete list of data points. Gustainis, "Processing Workflow Analysis for Special Collections."

further illustrated in the UC Libraries project, for which participants derived rates "by whatever tracking means they chose."[16] Variables related to what was measured and how it was measured are fundamental to understanding the wide range of processing rates that have been reported.

A second set of variables that contribute to differences in processing rates are introduced by the distinct characteristics inherent in each collection. Meissner and Greene's 2005 survey shows the profession's awareness of the potential impact of collection characteristics on processing rates; subsequent work demonstrates the actual impact.[17] UC project metrics showed that the better the physical condition of the materials, the fewer hours it took to process the collections, and that collections of personal papers and photographs took longer to process than corporate records.[18] This finding was intuitive and aligned with an earlier study that demonstrated the degree to which collection type influenced processing time.[19] Counterintuitive findings related to collection characteristics were reported by Countway Library, whose metrics showed "collections of moderate complexity clearly took less time to process than collections of low complexity." They attributed the slower processing rate to the fact that less complex collections were usually processed by less experienced staff.[20] This correlation between individual performance and processing rates raises another set of variables in the processing rate equation.

The third set of variables, human variables, are influential, but difficult to quantify. Human variables center on individuals' work ethics, skills, and levels of expertise. Anecdotes about the impact of individuals on processing rates are common, but few authors venture into this territory. Davis (California State Archives) broke down discrete processing tasks performed by different levels of personnel and charted their rates of output, and Gustainis (Countway Library) compared team processing to solo processing.[21,22] Studies acknowledge that measuring productivity

16. Wisner, *Uncovering California's Environmental Collections*, 8.

17. Survey respondents indicated the top five collection characteristics with the greatest impact on processing productivity were: existing level of organization of collection materials; physical condition of collection materials; overall collection size; structural complexity of collection; and heterogeneity of collection materials. Meissner and Greene, "More Product, Less Process," 261.

18. Wisner, *Uncovering California's Environmental Collections*, 10-15.

19. Abraham et al., "What Is Backlog Is Prologue," 37-39. Abraham et al. studied 328 accessions (totaling 1,341 cubic feet) and reported the following processing rates for different types of materials: photographs and oral histories at 47 hours per cubic foot; personal papers at 36 hours per cubic foot; and university archives at 16.8 hours per cubic foot.

20. Gustainis, "Processing Workflow Analysis for Special Collections," 120.

21. Davis, "Budgeting for Archival Processing," 211. The implication of the division of labor in the chart is that archivists' productivity rates are slower because they are performing more complex work than clerical staff.

22. Gustainis, "Processing Workflow Analysis for Special Collections," 120-121.

relative to a processor's skills and expertise is relevant.[23] Placing processing rates in the context of position types is fundamental to computing costs, since wages correspond with position types. The authors of this paper hesitate to define a person's skill level or potential productivity based on their position title; however, they concede that correlating processing rates with position type may be the most straightforward and diplomatic way to factor individual skills into the equation.

The final set of variables discussed in the literature is based on the nature of the work performed. Given the vast array of organizational cultures, staffing models, and resources at the disposal of repositories, it is fitting for processing approaches to vary across the profession. For decades, archivists have processed materials at different levels of granularity.[24] Gustainis and Wisner each accounted for levels of processing in their assessments. As might be expected, they each found the more intense or detailed the processing, the more hours of processing were required per foot. Wisner's results for the UC project showed average hours per foot more than doubled from one level to the next.[25] In addition to the complexities introduced by different levels of processing, another thorny variable surfaces when partially processed collections are processed further at a later date. As Dan Santamaria pointed out, "Iterative processing…adds a layer of complication to data collections because the same material may be addressed in different ways within short time periods."[26] The authors of this paper did not find any published processing rates that differentiated between new processing, re-processing, or iterative processing.

Challenges related to developing meaningful metrics are a constant refrain in the literature. Overall, the literature confirms that archival processing is flexible, nuanced, and subject to many influences. Consequently, processing rates legitimately vary from one environment to another, and even from one collection to another within the same environment. Given the multitude of variables that impact processing, working toward a universal benchmark for processing productivity rates

23. See Davis, "Budgeting for Archival Processing," 211; Ericksen and Shuster, "Beneficial Shocks," 44; and Gustainis, "Processing Workflow Analysis for Special Collections," 120-121.

24. An OCLC survey showed 75% of respondents strategically employed different levels of processing via an "MPLP-style approach" in Jackie M. Dooley and Katherine Luce, *Taking Our Pulse: The OCLC Research Survey of Special Collections and Archives* (Dublin, Ohio: OCLC Research, 2010), 49. At least three decades of literature discuss levels of processing. See the levels of processing and five premises discussed by Slotkin and Lynch in "An Analysis of Processing Procedures"; the processing continuum in Megan Floyd Desnoyers' "When Is a Collection Processed?" *Midwestern Archivist* 7, no. 1 (1982): 5-23; the flexible approach proposed by Mark A. Greene and Dennis Meissner in "More Product, Less Process"; the practice of minimal processing reported in Stephanie Crowe and Karen Spilman's "MPLP @ 5: More Access, Less Backlog?" *Journal of Archival Organization* 8, no. 2 (2010): 110-33; and iterative processing methods described by Daniel A. Santamaria in *Extensible Processing for Archives and Special Collections: Reducing Processing Backlogs* (Chicago: ALA Neal-Schuman, 2015).

25. See Gustainis, "Processing Workflow Analysis for Special Collections," 120-121; and Wisner, *Uncovering California's Environmental Collections*, 10-15.

26. Santamaria, *Extensible Processing*, 113.

across the profession is not a worthwhile effort. Instead, a more fruitful endeavor would be to focus on standardizing methods for assessing archival processing. Archivists can increase profession-wide understanding of archival processing realities by publicly sharing their processing rates, data collection methods, and self-assessments. The remainder of this paper contributes to this professional dialog through UNLV's critical examination of their processing data and assessment methodologies.

Processing Metrics at UNLV Libraries Special Collections and Archives

The Special Collections and Archives (SCA) division of the UNLV University Libraries documents the history, culture, and environment of Las Vegas, the Southern Nevada region, the global gaming industry, and the University. SCA's holdings include over 12,000 cubic feet of archival collections, over 32,000 books and periodicals, 1,800 maps, and 4,000 oral history interviews. The division is comprised of five units: the Oral History Research Center, the Center for Gaming Research, Digital Collections, Public Services, and Technical Services. The Technical Services unit is comprised of five permanent full-time staff and eight part-time student employees. Temporary contract personnel are hired periodically for special projects.

UNLV's archival processing policies and procedures reflect the resource management strategies advocated by Meissner and Greene's theory of "More Product, Less Process." Given the absence of profession-wide guidelines, UNLV adopted the *Guidelines for Efficient Processing in the University of California Libraries* (hereafter, UC *Guidelines*) to predict the resources that are needed to complete the processing of each unique collection by helping to determine the golden minimum level of processing that will provide appropriate access.[27] The effective planning and prioritizing of ongoing activities and special projects rest upon frequent assessment of archival collection data, processing metrics, and patron use statistics.

When assessing processing activities, UNLV views "processing" as "activity required to gain intellectual control of records, papers, or collections, including accessioning, arrangement, culling, boxing, labeling, description, preservation and conservation."[28] This includes collection retrieval, storage, and location tracking that occurs in relation to processing as well as activities that make the materials discoverable online. UNLV does not consider a collection "processed" until it is discoverable online. This is consistent with the *Guidelines for Standardized Holdings Counts and Measures for Archival Repositories and Special Collections Libraries*

27. UNLV adopted the UC *Guidelines* because UNLV's Director of SCA (at the time) and the Head of SCA Technical Services had each effectively used the *Guidelines* at other repositories before coming to UNLV.

28. Society of American Archivists, *A Glossary of Archival and Records Terminology*, definition of "processing", https://www2.archivists.org/glossary/terms/p/processing (accessed December 5, 2019).

developed by the SAA-ACRL/RBMS Joint Task Force.[29] Also in alignment with these guidelines, UNLV reports holdings and processing rates in cubic feet, however, this paper also includes some legacy UNLV reports in linear feet.[30] The processing rate tables herein include extents in both cubic and linear feet to provide a comparison for readers who use linear feet. To simplify the dialog, the authors use the term "collections" hereafter to reference all formats and material types (corporate records, personal papers, oral histories, university archives, born-digital archives, photograph collections, and audiovisual materials).

From 2015 to 2016, UNLV conducted a collection survey to evaluate their archival holdings at the box level (at that time, 1,324 collections totaling over 6,500 cubic feet). The survey addressed the UNLV Libraries' strategic goal of identifying collection needs and detailing the services, strategies, and resources required to meet those needs. It detailed the processing work needed to make each hidden and under-described collection serviceable. Based on the charts in the UC *Guidelines*, surveyors suggested processing levels and estimated processing times, considering variables such as the existing level of access, collection characteristics, potential restrictions, research value, and the nature of the work they recommended. Rather than retiring the data as a snapshot of the collections at a point in time, staff maintain its relevance by continuing to record information there. Each survey form holds data recorded during the 2015-2016 collection survey, as well as current information related to accessions, processing operations, and special projects. As of this writing, each collection form has 84 data points, including processing rates and the level of processing that was performed. Collectively, the forms provide the most comprehensive source of information for local archival collection management.[31]

29. SAA-ACRL/RBMS Joint Task Force on the Development of Guidelines for Holdings Counts Metrics (JTF -HCM), *Guidelines for Standardized Holdings Counts and Measures for Archival Repositories and Special Collections Libraries*, Appendix C, http://www.ala.org/acrl/sites/ala.org.acrl/files/content/standards/holdings_counts_2019.pdf (accessed December 5, 2019).

30. In 2019, the American Library Association (ALA) and the Society of American Archivists (SAA) endorsed cubic feet as the standard unit of measure for archival collections in the *Guidelines for Standardized Holdings Counts and Measures for Archival Repositories and Special Collections Libraries*. Previously, UNLV had reported collection extents only in linear feet. In 2016, in anticipation of this change, UNLV developed the Rebel Archives Calculator, https://www.library.unlv.edu/speccol/rebel_archives_calculator/, which simultaneously calculates linear and cubic feet, and then re-measured all their physical collections.

31. For survey methodology, data points, and examples of past and current forms, see UNLV Archival Survey Documentation at https://drive.google.com/drive/folders/0B25SSpmxsyXOLVp3UjZSWFJMRoo (accessed February 24, 2020). As of this writing, UNLV is exploring alternatives to Google Forms with the hope of moving toward a single software with the potential to replace their use of Google Forms, Trello, and Excel for workflow and collection management.

Assessment of Special Projects

UNLV's archival assessment methods are developing as operations mature. As with many repositories, UNLV began tracking processing metrics within the context of special projects. They began collecting basic processing metrics (4 data points) during an architectural records project and gradually adding data points with each project. Table 1 compares processing metrics across projects. Table 1 is followed by information that impacted the metrics of each project, such as the length of the project, dates of the collections, levels of processing performed, and the size and experience level of the project team. Data reported herein reflect the processors' hours only—metrics do not include time supervisors, project managers, or curators dedicated to planning or overseeing processing.

Table 1. Processing rates for UNLV special projects

Project name	Number of collections	Extent: cubic feet	Extent: linear feet	Extent: GB	Total processing hours	Range of hours/ cubic foot	Average hours/ cubic foot	Median hours/ cubic foot
Architectural records	14	1,214	1,405	0.098	2,725	0.88 to 17.54	3.79	2.46
America's Great Gamble	4	454	422	N/A	4,037	3.97 to 18.60	9.71	8.13
Archival backlog elimination	391	5,221	5,247	149	8,640	0.11 to 206.25	9.45	3.52
Legacy oral histories	1,249	N/A	N/A	2,540	1,534	N/A	N/A	N/A
Project totals to date	1,658	6,890	7,074	2,689	16,936	0.11 to 206.25	7.65	3.52

Architectural Records Processing Project

Architectural records processing was performed by paid graduate student interns from the UNLV School of Architecture who were experts in architecture but novices in archival processing. The metrics reflect the work of 4 individuals over 2 years. Their processing rates were the fastest rates of all UNLV projects to date. They processed 14 collections dating from 1931 to 2010 (1,214 cubic feet) at an average rate of 3.79 hours per cubic foot and a median rate of 2.46 hours per cubic foot. The interns processed collections intensively at Level 4 or 5 (the set or drawing level) and performed labor-intensive conservation treatment—both of which slow processing rates. The project manager speculated that the lightning-fast processing rates were sparked by a near perfect storm of positive variables. The human variables included: the interns' subject knowledge enabled them to identify reproduction methods, media, and types of drawings on sight and to quickly describe materials; they learned basic archival theory and technologies quickly; and working collaboratively in pairs lent efficiencies. Some variables related to collection characteristics also expedited processing: there were no restricted materials to separate and the text blocks on drawings provided a ready reference for description. Because the data collected were not parsed into specific actions, UNLV could not quantify the degree of influence that each variable wielded; however, the collective impact of the variables is evident in the exceptionally fast processing rate.

America's Great Gamble Processing Project

The processing team for *America's Great Gamble*, a federally funded grant project, included three individuals with processing experience and one subject expert. The processing team recorded 20 data points. The metrics reflect the work of 4 individuals over 18 months. They collaboratively processed 4 collections dating from 1811 to 2017 (422 linear feet) at an average rate of 11.05 hours and a median rate of 8.95 hours per linear foot.[32] During this project, the UC *Guidelines* proved reasonably accurate. To establish timelines for the grant proposal, the project manager used numbers from the low- to mid-range of the estimates in the UC Processing Rates chart.[33] In reality, processing the four collections took the team 0.7 to 2.3 hours more per linear foot than they had anticipated—differences with significant impact when multiplied by 50 to 150 linear feet (the size ranges of the collections). The actual processing rates were within the hourly ranges predicted in the UC *Guidelines*, but came in at the top of the

32. The grant was written prior to UNLV's transition to measuring holdings in cubic feet, therefore the proposal and outcomes were reported in linear feet. Additions to some of these collections have been received since the completion of the grant project. The numbers herein reflect the grant project metrics only.

33. See chart in *University of California Libraries, Guidelines for Efficient Archival Processing*, 22-23.

range.[34] The manager noted two unquantified variables that pushed their rates toward the higher end of the UC *Guidelines*' estimates: when the team was ahead of schedule, they elected to selectively perform more intensive processing than originally planned; and after processing was completed, the team had to intensively review two of the collections to identify and separate unanticipated restrictions.

Archival Backlog Elimination Project

As of this writing, UNLV is in the final year of a project to eliminate their backlog of unprocessed and under-described archival materials. The main goal of this project is to improve access to over 1,300 physical archival collections (totaling over 6,500 cubic feet) and over 4,000 oral history interviews that were previously hidden, unprocessed, or inadequately processed.[35] The processing metrics from the manuscript/records and photograph component of the project reflect the work of 28 individuals (working sequentially, not simultaneously) over 2 years. They recorded 22 data points. When the backlog elimination project efforts were supplemented by a team of inexperienced undergraduate students, who processed over 144 cubic feet of historic menus, the data points were simplified to reflect the less nuanced work they performed. As of this writing, project staff have processed 391 legacy collections with materials dating from 1817 to 2017. These collections were processed at various levels of detail at an average rate of 9.45 hours per cubic foot and a median rate of 3.52 hours per cubic foot. Frequent assessment of processing data significantly assists in project management and leveraging the unique skills of individuals.

Legacy Oral History Project

In August of 2017, UNLV began collecting processing metrics for 4,000 legacy oral histories (dating from 1972 to 2017), another component of the archival backlog elimination project. At that time, the metrics focused on processing tasks associated with born-digital interviews stored on external legacy media. Undergraduate students processed the digital files; they captured 9 data points, including time scanning donor agreements and biographical paperwork, conducting virus scans, copying files from media to a networked server, embedding metadata in digital files, and creating access copies. In December of 2017, they began creating finding aids, and captured 11 data points related to that work. The processing rates of the 12 employees (undergraduate students and archival processors) who have digitally processed or created finding aids

34. The project was funded by the National Historical Publications and Records Commission. For additional information see Cyndi Shein, Hannah Robinson, and Hana Gutierrez, "Agility in the Archives: Translating Agile Methods to Archival Processing Projects," *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 19, no. 2 (2019): 94-120. For details on processing rates, see University of Nevada, Las Vegas, *Final Narrative Report* (award number NA16-RH-50190-16), https://drive.google.com/file/d/0B06zBpcGMoHSbkZHQnNoUlJPTmRqRGhQWTk4OTRKeDVMcXlF/view (accessed February 24, 2020).

35. UNLV's archival backlog elimination project includes goals outside the scope of this paper, such as processing some born-digital collection materials and assessing UNLV's audiovisual archives.

for oral histories over the past two years were remarkably consistent.

As of this writing, staff have processed 1,249 interviews (2,540 gigabytes) at an average rate of 1.2 hours per interview and a median rate of 1.1 hours per interview. Digital file processing averaged 0.6 hours per gigabyte, with a range of 0.4 to 2.1 hours per gigabyte. The manager attributes this consistency to thoroughly trained staff, clear instructions, and the uniform content and format of the oral histories. Basic data points for creating oral history finding aids closely mirror metrics for physical processing; however, because oral histories are processed at the item level and are ultimately served in digital format, their treatment includes additional steps that do not parallel physical processing. Processing steps performed on the digital versions of the oral histories (such as converting files for preservation or access) are aligned with data points used to assess born-digital processing. The data points piloted for oral histories were a steppingstone for developing data points for processing born-digital manuscript and photograph collections.[36]

Evolving from Project-based to Program-wide Assessment

Special projects provided opportunities for UNLV to explore assessment methods and take incremental steps toward operationalizing the approaches they piloted. SCA Public Services has been collecting and reporting use statistics for more than a decade, but SCA Technical Services is still developing their assessment practices. Incorporating use and collection data into the division's decision-making process is a work in progress. Persuading staff to embrace assessment practices has been challenging in some cases. Collecting and assessing data is time consuming, and some staff argue that it detracts from time spent doing "their real jobs." Moving from project-oriented to program-wide assessment not only requires a culture change, it also requires a change in procedures. Metrics were recorded differently as the program matured. UNLV faced significant challenges when assembling and evaluating the inconsistent data it had collected from 2015 to 2019. This highlighted the importance of standardizing their data collection methods. UNLV re-measured all collections to obtain extents in both linear and cubic feet, and investigated data anomalies to isolate and exclude untrustworthy data from their metrics. UNLV began to identify the lowest common denominators across their projects to determine which data points were vital to ongoing operations (beyond project work). To meet professional standards and to ensure that data are interoperable across the different domains of their program, UNLV also worked to standardize their terminology and units of measure.

36. Oral histories were not part of the 2015-2016 collections survey, and there were no predictions of how much time it would take to complete them. Digital extents reported represent the post-processing total in gigabytes and count the number of unique files acquired (not derivatives). In 2019, UNLV began tracking time on accessioning new oral history interviews, but that data is not available yet. Providing access to oral histories also involves additional time-intensive, cross-divisional actions that have not been tracked. UNLV staff has not tracked time preparing legacy audio cassettes for outsourced digitization (shipping and creating metadata), quality control of digital files produced by the vendor, or actions that facilitate online delivery of transcripts.

Processing Born-digital Material

In 2015, UNLV began actively managing the born-digital material in their collections. However, estimating the time required for born-digital processing actions proved extremely challenging due to a number of variables, including the lack of professional standards and frameworks. UNLV established consistency for local born-digital accessioning and processing procedures and began tracking basic actions in April 2018. As of this writing, staff are collecting metrics for 22 data points, including time spent on ingest actions (such as virus scans, generating checksums, and creating metadata), appraisal, digital processing actions (such as file conversion and batch file renaming), and online public description (a finding aid). Thus far, they have collected metrics for 38 collections (557.55 GB) with an average processing rate of 0.45 hours per gigabyte. However, there is not yet enough data to accurately forecast processing times. The time required to perform born-digital processing steps varies widely based on the file format, complexity of the file arrangement, potential for restrictions, age of the collection, and level of effort. For example, a collection with a 2.12 gigabyte total took 67.5 hours to process, primarily because the collection contained a large variety of complex design files from the 1990s stored on 267 pieces of external media, and necessitated highly intensive, item-level processing. On the other hand, a different collection of 86.4 gigabytes was processed in only 7.5 hours. Because this collection contained uniform image file types, no restrictions or personally identifiable information concerns, and a logical folder arrangement, it required minimal effort. Going forward, UNLV may track the level of processing effort performed on born-digital collections to see how that corresponds to processing time. Given the quickly evolving technologies, strategies, and practices employed to manage born-digital archives, this may continue to be an area where processing estimates prove elusive.

Accessioning and Processing Practices

Processing planning and prioritization begins at the point of accessioning. At the time of accessioning, all incoming collections are surveyed, physically and/or digitally stabilized, and described at the collection level in a published finding aid. During accessioning, some collections are identified as candidates for "accessioning as processing," meaning that baseline physical and intellectual work is all they will receive unless there are compelling reasons to process them in more detail in the future.[37] For all other collections, however, the accessioning archivist performs baseline processing, publishes a collection-level finding aid, and makes recommendations for further processing that might be performed if and when that collection rises to the top of the processing priority queue. These recommendations include the collection's processing priority as determined by curators, the level of processing to be performed, and an estimate of how many hours it will take to achieve the recommended level of processing.

37. University of California Libraries, *Guidelines for Efficient Archival Processing*, 13.

The practice of tracking accessioning metrics has evolved over time at UNLV. Between September 1, 2017 and August 21, 2019, UNLV tracked 8 data points associated with accessioning activities: extent (physical and digital), level of effort (per the UC *Guidelines*), dates accessioning work was started and completed, total hours spent on accessioning tasks, whether the work performed was considered "accessioning as processing," and staff names. Metrics were recorded by permanent staff, graduate students, and undergraduate students over the two-year period. The following tasks were included in the total hours spent on accessioning: logistics (moving collection materials, building boxes, and shelving boxes), physical arrangement (refoldering and rehousing), and description (collection-level finding aid).

Accessioning data was recorded for 69 collections with materials dating from 1861 to 2017, totaling 232 cubic feet. These collections were simultaneously accessioned and processed, primarily at the collection level, at an average rate of 4.3 hours per cubic foot and a median rate of 3.0 hours per cubic foot. The average processing rate for this subset was lower than the average overall rates for minimal and collection-level processing at UNLV. (See Table 2 for baseline processing/ accessioning rates and Table 4 for overall Level 1 processing rates.) Of these 69 accessions, approximately half were minimally processed during accessioning. The minimally processed accessions were either accruals to previously processed collections that were intellectually integrated into existing collection descriptions, or they were new collections that were under 5 cubic feet and were already well-organized and did not present major conservation issues.

Table 2. Processing rates for baseline processing performed during accessioning

Number of collections	Extent: cubic feet	Extent: linear feet	Extent: GB	Total processing hours	Range of hours/ cubic foot	Average hours/ cubic foot	Median hours/ cubic foot
69	232	230	74.5	297.15	0.035 to 5.12	4.3	3.0

These new accessioning procedures offer potential to integrate UNLV's accessioning and processing workflows. To date, outside of project work, UNLV has only sporadically collected processing metrics. Having recently filled their full-time processing archivist position, UNLV looks forward to testing data collecting methods

during daily operations. In August of 2019, UNLV began to track more granular metrics during accessioning, utilizing many of the same data points that are recorded during processing. Although data points related to some processing actions (such as creating an ArchivesSpace inventory and physical arrangement) are not always relevant to accessioning, many processing actions are now routinely performed during accessioning (such as collection review, rehousing, and basic description). Tracking accessioning data points consistently with processing data points will enable staff to more accurately compare metrics across the program. Over time, UNLV will observe how baseline processing affects the overall processing rates of collections that are slated for further processing. Baseline processing averages 4.3 hours per cubic foot—what will the average be if those collections receive further processing? And how might iterative/extensible processing rates compare to UNLV's processing average of 8.74 hours per cubic foot?

Assessment of UNLV Field Data and Methods

As the major backlog elimination project winds down and UNLV returns to business as usual, it is essential to map out a sustainable path forward. It is time to distinguish data that served immediate project needs versus vital data that must be continuously collected to support ongoing operations. Future data collection efforts will be streamlined to focus on ongoing needs. This is an opportunity for UNLV to assess and refine their methodologies. Challenges encountered during their recent field data assessment will inform future data collection methods.

UNLV Field Data Collection and Analysis: Methodology

The processing field data discussed in this section refer to corporate records, personal papers, university archives, and visual materials (see Table 3 through Table 9). They do not include oral histories or born-digital materials (which are managed separately). Although the metrics include a small representation of one FTE archivist's processing rates prior to their retirement, the field data were primarily generated by project staff. Significantly, the metrics do not include project supervisors' time spent on planning, reviewing work, or contributing to processing in other ways. UNLV's processing metrics were collected at different levels of granularity during the projects described herein. Processors recorded their own time in individual or group spreadsheets as directed by project managers. When they completed a collection, processors updated the collection survey Google form for the collection (which feeds into a central sheet that tracks the size and status of all collections). Processors entered the total hours they spent processing the collection, level of processing, and processing notes in separate data fields. Finally, they marked the collection completed. They updated the extent field only when the overall extent changed due to rehousing, appraisal, or additions.

UNLV began assessing their processing rates by exporting the aggregated data from the collections Google forms and examining information on all their archival holdings—1,461 collections. UNLV's first step was to eliminate collections that lacked

complete data. Since processing rates are calculated by dividing time (hours) by extent (feet), those data points were essential. UNLV eliminated 821 collections (56%) because those collections were missing data for extent or time. The second step was to calculate processing rates and closely examine anomalies that might indicate inaccuracies. The most common reason for elimination from the data pool at this stage was when the extent reflected the *entire* collection but, as evident in the free text processing note, the time referred only to processing a *portion* of the collection. Processing legacy collections sometimes involved adding an accrual or completing processing on a collection that had been partially processed prior to the project. In situations where staff only processed a portion of the collection, there was not a separate field/data point for them to indicate how much of the collection they processed. This proved problematic when calculating processing rates. UNLV manually reviewed the processing notes and eliminated over 75 additional collections where they suspected the extent did not accurately reflect the extent processed. Although still subject to human error, UNLV made every effort to ensure the remaining data were reliable. Ultimately, they reduced the pool to 563 collections and analyzed it from several angles by sorting and filtering it in Excel.

Observations

In their data assessment, UNLV found as wide a variety of rates as those reported in the literature. UNLV struggled to determine whether their median processing rate or average processing rate best represented the reality that could serve as a benchmark going forward (see Table 3). Using the overall median to forecast processing rates would result in dramatic underestimations of time needed to process a collection. On the other hand, the higher (slower) productivity rates represented by the overall average includes extreme highs (206 hours per foot) and lows (0.04 hours per foot), which are exceptional and can sometimes be attributed to specific variables.

Table 3. Overall processing rates at UNLV

Number of collections	Extent: cubic feet	Extent: linear feet	Total processing hours	Range of hours/ cubic foot	Average hours/ cubic foot	Median hours/ cubic foot
563	7,437.5	8,137.28	13,495	0.04 to 206	8.74	2.70

Dramatic extremes demanded a closer look, so UNLV staff examined subsets of the data in an attempt to quantify the impact of select variables upon local processing rates. They attempted to assess many variables, but data only fully supported assessment in the following areas: levels of processing, collection size, learning curves, and repurposing existing electronic inventories.

Variable: Levels of Processing

One variable that strongly impacts all of UNLV's processing rates is the level of processing performed. UNLV uses the UC *Guidelines* for practical guidance in applying the golden minimum theory expressed in "More Product, Less Process". UNLV measured their actual processing rates against the hourly estimates in the UC *Guidelines* (see Table 4). It should be noted that this is a rough comparison, since UNLV employs cubic feet and the UC estimates are supplied in linear feet. In addition to the five levels of processing (minimal to highly intensive) defined by the UC *Guidelines*, UNLV parsed their processing efforts into a sixth category—description only—to explore the supposition that description (at any level) is extremely fast when no physical work is required. This supposition was not supported by the data.

When parsed by processing levels, UNLV's processing rates raise questions. For example, why were the average rates for Level 1 slower than for Level 2? UNLV's average rate for Level 1 was much higher/slower (5.8 hours per cubic foot) than the UC estimate (1-3 hours per linear foot). UNLV's averages for Levels 2, 3, and 4 fit roughly within the lower end of the ranges of the UC charts. UNLV's Level 5 average processing rate (10.3 hours per cubic foot) was lower/faster than the UC estimate (14 to 22+ hours per linear foot). The imbalance between Level 1 and Level 2 processing rates at UNLV can in part be attributed to some staff's misinterpretations of the processing levels. Spot-checking revealed that at least one processor routinely recorded their work as Level 1, even when they performed series-level or box-level processing (which should have been recorded as Level 2). UNLV concluded that the conceptual frameworks in the UC *Guidelines* proved to be a reasonable starting point for calculating processing rates in the UNLV environment, but it needed to investigate other variables to see if any of the unexpected rates exhibited patterns that offer insights into the overall rates.

Variable: Collection Size

Small collections. Generally, the highest/slowest rates in the overall data set corresponded with the smallest collections. UNLV found that the processing times they predicted consistently fell short of the time actually needed to process small collections, particularly collections that needed minimal or no physical processing. Using the UC *Guidelines*' estimate of 1-3 hours per linear foot for a collection-level finding aid, UNLV frequently estimated 1 hour would be enough time to process a small collection or a collection that was adequately processed physically but needed a DACS-compliant finding aid. This consistently led to underestimations in work plans.

Table 4. UNLV processing rate realities by level compared to UC *Guidelines*' estimates

Processing level	Number of collections	Extent: cubic feet	Extent: linear feet	Total processing hours	Range of hours/ cubic foot	Average hours/ cubic foot	Median hours/ cubic foot	UC Guidelines range hours/ linear foot
Description only (finding aid creation or editing), no physical work	157	261.33	294.6	465	0.13 to 206.25	22.74	7.27	N/A
Level 1 minimal effort; collection level	100	370.04	404.16	443	0.04 to 50	5.83	2.55	1 to 3
Level 2 low effort; series or subseries level	167	1,088.9	1,329	2,091	0.11 to 42.11	4.18	2.37	2 to 8
Level 3 moderate effort; file level (expedited)	83	1,086.3	1,048.4	2,422	0.54 to 21.74	4.26	2.84	5 to 14
Level 4 intensive effort; folder level	36	3,148.2	3,097	3,401	0.26 to 25	5.11	2.52	9 to 21
Level 5 highly intensive effort; item level	20	1,482.8	1,963.8	4,673	1.37 to 46.15	10.33	5.31	14 to 22+

To support more accurate processing work plans and proposals, UNLV analyzed data about collections with extents of less than 2 cubic feet. This subset of data included 322 collections with materials dating from 1817 to 2017, totaling 223.59 cubic feet. These collections were processed at various levels of detail at an average rate of

16.7 hours per cubic foot and a median rate of 5.8 hours per cubic foot.[38] The average rate for this subset was significantly higher than the average overall processing rates. Although processors can set some required fields to auto-populate in ArchivesSpace, human intelligence and actions are still needed for many of the steps that lead to accurate collection-level description. Steps such as reviewing collection documentation, creating subject headings and authorized names, writing front matter, and creating administrative files cannot be automated. At a minimum, processors also need to house, label, and shelve materials, and track container locations. In addition to processors' hours, staff from other divisions contribute an additional average of 1.25 hours per collection to make the description discoverable online.[39] Given current workflows and technologies at UNLV, the UC *Guidelines*' prediction of 1-3 hours per foot for Level 1 processing (minimal, collection-level description) does not apply to small collections. Because some extreme highs in individual rates for small collections cause the average to be unusually high (16.7 hours per cubic foot), UNLV looked to the median of 5.8 hours as the baseline time required for processing small collections. Every collection, regardless of size, consumes a minimum of 6 hours of processing staff time in the current UNLV environment.

Large collections. For comparison, UNLV isolated the larger collections in the data set, which ranged in size from 51 to 2,094 cubic feet. This subset of data included 25 collections with materials dating from 1828 to 2010, totaling 5,191.5 cubic feet. These collections were processed at various levels of detail at an average rate of 2.26 hours per cubic foot and a median rate of 1.4 hours per cubic foot. The average rate for this subset was significantly lower than the average overall processing rates. The collections in this subset were processed in less than 4 hours per cubic foot, with one exception. The historic menu collection (144 cubic feet) clocked in at 9.38 hours per cubic foot—more than twice the rate of any of the other large collections. The menu collection was a prime example of how a statistical outlier is often a sign of multiple variables acting concurrently to impact the processing rate. Identifiable variables for this collection included:

- Novice processor skill level: the collection was processed by undergraduate students who were first-time processors with no archival knowledge or experience.
- Highly intensive processing: the materials were in such physical chaos that item-level sorting was required to achieve any semblance of order.

38. For the 322 collections here, staff performed minimal to low processing, an average level of 1.4 on the untitled levels of control chart in University of California Libraries, *Guidelines for Efficient Archival Processing*, 15-16.

39. The Discovery Services department creates or updates MARC records in the local catalog and WorldCat, and the Web Application Development Services department uploads finding aids (PDFs) to UNLV Libraries' website.

- Intellectual complexity: the collection content included multiple languages and obscure geographic locations that required unusually intensive item-level research.
- Conservation issues: fragile menus from the 1800s required encapsulation.

Even with the negative influence of these variables, the processing rate for the menu collection (9.38 hours per cubic foot) was faster than the range of 14-22+ hours per linear foot predicted for item-level processing by the UC *Guidelines*. Though novice, the processing team learned quickly and was well coordinated, which offset the negative variables to some degree. To summarize, UNLV found their processing rates for large collections were significantly lower than their overall processing rates (see Table 5). In contrast, average rates for small collections were more than twice as high as the overall rates (see Table 6). This lends credence to anecdotal reports that processing small collections requires a proportionately high investment of human resources.

Table 5. Processing rates for collections over 50 cubic feet

Number of collections	Extent: cubic feet	Extent: linear feet	Total processing hours	Range of hours/cubic foot	Average hours/ cubic foot	Median hours/ cubic foot
25	5,191.5	5,343.6	7,901	0.13 to 9.38	2.26	1.84

Table 6. Processing rates for collections under 2 cubic feet

Number of collections	Extent: cubic feet	Extent: linear feet	Total processing hours	Range of hours/cubic foot	Average hours/cubic foot	Median hours/cubic foot
322	223.59	497.67	1,122	0.28 to 250	16.72	5.88

To gauge how much of an impact small collections had on their overall processing rates, UNLV then removed the small collections (less than 2 cubic feet) from the larger data pool and recalculated processing rates by level *without* the small collections. Removing the small collections from the data set lowered the ranges, averages, and medians for all levels, cutting the averages nearly in half for some levels. Without the small collections, Level 1 average rates remained higher than Level 2 average rates, but only slightly. Without the small collections, Level 1 median rates dropped below Level 2 median rates, implying that collection size did contribute to the counterintuitive imbalance between the overall rates for Level 1 and 2.

Table 7. Processing rates by level with and without small collections

	Range of hours/cubic foot OVERALL	Average hours/cubic foot OVERALL	Median hours/cubic foot OVERALL	Range of hours/cubic foot WITHOUT SMALL	Average hours/cubic foot WITHOUT SMALL	Median hours/cubic foot WITHOUT SMALL
Level 1	0.40 to 50	5.95	2.6	0.79 to 0.93	2.54	0.15
Level 2	0.11 to 42.11	4.18	2.37	0.11 to 17.5	2.37	1.67
Level 3	0.54 to 21.7	4.26	2.84	0.79 to 7.23	3.12	2.64
Level 4	0.26 to 25	5.11	2.52	0.26 to 16.67	3.28	2.35
Level 5	1.37 to 46.10	10.3	5.3	1.37 to 12.19	4.39	3.77

Variable: Learning Curves for Staff

Next, in an effort to quantify the impact of processors' skill levels on processing rates, UNLV explored the influence of the human variable. Although position title or type might be a fairly trustworthy method of calibrating processing speed in some situations, position title is irrelevant in an environment where all processors

(students, interns, and paraprofessionals) begin without any processing experience. UNLV was not able to parse their data based on skill level. However, they were able to clearly identify collections that demonstrate the effect of a learning curve on processing rates by isolating collections that were processed during training exercises. This subset of data included 48 collections with materials dating from 1846 to 2017, totaling 78.5 cubic feet. These collections were processed at various levels of detail at an average rate of 9.79 hours per cubic foot and a median rate of 7.26 hours per cubic foot. The average rate for this subset was only slightly higher than the average overall processing rates (8.74 hours per foot) but the trainee median rate of 7.26 hours per cubic foot was more than twice as high as the overall median rate of 2.7 hours per cubic foot. Some of the collections used in training were processed at Level 1 and reported higher than average processing rates. Also of note, three of the small training collections (average extent of 0.21 cubic feet) averaged processing rates of 20.13 hours per cubic foot—another example of how concurrent variables (novice processors and small size) can create extreme rates. Having an approximate figure for trainee rates will help UNLV select collections that can be completed by novices within specific time periods, particularly when they host MLIS internships with short timeframes.

Table 8. Processing rates for training exercises

Number of collections	Extent: cubic feet	Extent: linear feet	Total processing hours	Range of hours/cubic foot	Average hours/cubic foot	Median hours/cubic foot
48	78.5	108.5	431	0.62 to 46.2	9.79	7.26

Variable: Repurposing Existing Electronic Description

As might be expected, some of the fastest processing rates resulted from efficiencies aided by existing electronic inventories found in legacy files or provided by donors. This subset of data included 10 collections with materials dating from 1950 to 2007, totaling 803.59 cubic feet. These collections were processed at various levels of detail at an average rate of 1.56 hours per cubic foot and a median rate of 1.38 hours per cubic foot. The average rate for this subset was significantly lower than the average overall processing rates. These rates can be referenced during conversations with donors who expect their collections to be processed quickly and/or to a specific

level of detail. The rates demonstrate how processing can be accelerated when donors are willing and able to create electronic inventories of the materials they donate.[40]

Table 9. Processing rates for collections with pre-existing electronic inventories

Number of collections	Extent: cubic feet	Extent: linear feet	Total processing hours	Range of hours/cubic foot	Average hours/cubic foot	Median hours/cubic foot
10	803.59	850.73	1,036	0.59 to 3.66	1.56	1.38

Variables That Eluded Capture

During data analysis, UNLV initially attempted to correlate processing rates with additional variables such as processor skill level, presence of restricted materials or conservation issues, and the homogeneity of the materials. Although processor notes suggested the influence of these variables, UNLV data failed to provide hard data to substantiate anecdotal claims.

Skill level. It is logical to presume that skill level affects how quickly a person can process a collection, but UNLV found skill level difficult to quantify based on their data. As mentioned, UNLV metrics were primarily derived from projects staffed by contract employees who came in with no experience. UNLV attempted to measure the impact of the growing skills of individual processors (novice, competent, experienced, or expert) over time. However, determining the dates individuals moved from one skill level to the next and manually filtering over 500 collections based on those dates was inexact and too labor-intensive to perform.

Restricted materials. Processors noted that item-level review and separation of restricted materials noticeably impeded processing progress. Although UNLV's metrics included a data point to indicate the presence of restricted materials, there was no data point to indicate the time devoted specifically to reviewing and separating restricted materials. Additionally, reviewing and separating restrictions actions are typically integrated with arrangement and description actions, which would have made it difficult to accurately record restriction actions as a separate data point.

40. On the rare occasions when donors are willing and able to create description to accompany their donations, UNLV provides an Excel template to guide donor description and facilitate metadata transformation.

Conservation. Generally, basic conservation actions are also integrated with arrangement actions during processing, but some collections required extreme conservation measures. Collections with severe conservation issues (such as mold) were routed to the department's Preservation Lab, where labor-intensive, item-level treatment was performed. The work was performed by several employees whose hours were not included in UNLV's recorded processing rates. Even so, hourly rates for these collections were relatively high. The Preservation Lab tracks work at the item level, not the collection level, and further analysis would be required to draw reliable conclusions about the effect of severe conservation issues on processing rates.

Homogeneity. The final variable that processors cited as having an influence on their processing rates was the homogeneity of materials, as found in university or architectural records. Uniformity of material types appeared to accelerate processing rates; however, processors noted the homogenous collections with low processing rates were also well-organized and had clearly labeled folders. This implied that several variables concurrently influenced the processing speed for these collections, and any conclusions here would be speculative.

Moving Forward

Beginning in July of 2020, UNLV plans to collect metrics in their post-project world, and within two years' time, it hopes to have gathered enough data to reasonably compare processing rates for daily operations with the project rates reported herein. There are reasons that some of the extremely low rates reported herein may not be repeatable once the projects are completed. Project staff have been dedicated exclusively to processing, a focus that is thought (anecdotally) to accelerate processing. Processing rates for some collections (such as the menu collection) may falsely appear lower because supervisors did not include their own hours when reporting the total processing times of each collection. Furthermore, although UNLV staff painstakingly reviewed the final data set looking for signs that a collection had been partially processed prior to the project work, they may not have had sufficient information to eliminate all untrustworthy data. On the other hand, there are also reasons that processing rates may be lower in UNLV's post-project world. The project rates focused on processing legacy collections, the majority of which were accompanied by myriad, often extraordinary, challenges that slowed processing and resulted in high rates. UNLV's current accessioning procedures are designed to mitigate such challenges by performing baseline processing at the time of accessioning. All things considered, UNLV is confident that their final data set of 563 collections provided the size and diversity required to obtain averages and medians that will reliably inform plans and priorities for ongoing operations.[41]

41. UNLV's data set is substantial in comparison to processing metrics shared in recent case studies. For example, Gustainis analyzed data from 36 collections in "Processing Workflow Analysis for Special Collections"; Walters analyzed data from 30 collections in "Processing Large-Scale Architectural Collections"; and Wisner analyzed data from 17 collections in *Uncovering California's Environmental Collections*.

UNLV's processing metrics enabled them to answer a lingering question: *Do the benchmarks in the UC Guidelines support reliable processing forecasts?* In short, the answer is, *Yes*. UNLV tested and found the charts in the UC *Guidelines* extremely useful when interpreted through professional judgement. UNLV plans to continue using the chart of levels of effort/control to support processing planning.[42] As demonstrated in Table 4, when UNLV compared their processing field data to the processing time range estimates in the UC *Guidelines*, the UC chart proved fairly accurate for Level 2, 3, and 4.[43] UNLV's processing rates for Level 5 (item-level processing) and collections over 50 cubic feet were generally lower/faster than the UC expectations. UNLV's processing rates for small collections and collections that required intellectual description only (no physical work) were far higher than both the UNLV overall average and the UC *Guidelines* estimates. For local purposes, UNLV plans to adjust the UC chart for Level 1 and 5, employ cubic rather than linear feet, and continue using a slightly modified version of the chart to benchmark their processing rates. Once UNLV has gathered a year or two of post-project metrics, they will reassess these benchmarks and revise them as needed. In UNLV's current environment, every collection requires 6 hours to process (regardless of collection size); therefore, UNLV will not use a processing rate chart for collections under 2 cubic feet. Instead, they will plan for an average of 6 hours for every small collection, adding more hours to forecasts if the material is in poor physical condition or has other significant barriers to access.[44]

UNLV has been earnestly collecting data in some form since 2016 and staff were disappointed when data analysis in 2019 revealed flaws in their data collection methods. UNLV was forced to abandon over half their data because it was incomplete or misleading. The biggest preventable loss was caused by their failure to provide a central data collecting framework that isolated the extent processed from the entire collection extent and did not specify the nature of the processing work that was performed. As mentioned, data from hundreds of collections had to be thrown out because those calculations were based on the *entire* extent of each collection, when only a *portion* of the collection was processed during a project. Recording the level of processing was extremely valuable, but it only provided part of the equation. Going forward, they will test a framework aimed to consistently capture extent processed and the nature of the work performed.

42. See the untitled chart for levels of control in University of California Libraries, *Guidelines for Efficient Archival Processing*, 15-16.

43. Chart titled "Average processing rate (hours per linear foot) given level of processing effort and condition of the materials" in University of California Libraries, *Guidelines for Efficient Archival Processing*, 23.

44. "Barriers to access include any characteristic of a collection that might impede user access and require staff attention before a collection may be used productively for research. Barriers to access include disorganization, poor housing, poor description, preservation issues, the presence of special media or other fragile materials, etc." in University of California Libraries, *Guidelines for Efficient Archival Processing*, 23.

Proposed Data Framework

In their quest for the golden minimum of processing assessment, UNLV is attempting to standardize methods for recording data, identify data points that are vital to ongoing operations, create an extensible framework to collect added value data for special needs, and cease collecting data that are unnecessary or can be derived from other sources. In their attempt to centralize all information about collections through survey forms, UNLV intermingled processing metrics with holdings metrics and collection details, which unnecessarily complicated processing data. UNLV plans to pilot an experimental framework of "core" and "added value" data points.

Core Data Points: Serving Ongoing Needs

The UNLV Libraries asks the Special Collections and Archives (SCA) division for very little data. It asks only for the number and volume of archival collections acquired every six months (for fiscal year and calendar year reports). This information is gleaned from accession records. Taking a proactive stance, UNLV's SCA Technical Services collected additional data that they thought would support their operations. This is in line with Carter's belief that "Even if we do not undertake assessment activities for the sake of our parent organizations, we should develop metrics and employ them in the interest of improving our own practices."[45] After assessing all the processing metrics they collected, UNLV determined that much of it was excessive. For UNLV to inform strategic planning, operational priorities, grant proposals, progress reports, and value propositions at the most fundamental level, a bare bones processing metrics framework would need only five data points:

- Collection name or identifier
- Total processing hours
- Extent processed in cubic feet
- Extent processed in gigabytes
- Date processing was completed

Based on their experience, UNLV's SCA Technical Services plans to reach a little beyond these bare essentials. It is cutting several of their existing data points that detail processing actions about arrangement and description (which are often too intertwined to distinguish), and replacing those with a single data point, "Nature of the processing performed". UNLV believes this data point will enable it to assess disparities in their processing rates more accurately. They plan to experiment with a framework of 12 core processing data points (and one free text field) that are relevant across their operations (see Table 10).

45. Carter, "Articulating Value," 93.

Table 10. Core data points and purposes for collecting them

Data point	Purpose
Collection identifier	Essential for merging processing data into larger data sets and aggregating processing metrics from different functional areas of the department.
Collection name	Not absolutely essential but serves as a secondary identifier. Proven essential in cases where the collection identifier was missing or incorrect.
Extent processed (cubic feet for physical materials processed)	Essential for calculating processing rates and quantifying the amount of work that was performed. Necessary for annual reports and funding reports. Extent here indicates the pre-processing size of the materials to reflect the extent of all the work performed, including review of materials that were de-accessioned. The final collection description shows the post-processing extent.
Extent processed (linear feet of physical materials processed)	Not essential internally, but helpful in comparing local processing rates to those of other institutions. Since the Rebel Archives Calculator generates linear and cubic feet simultaneously, including linear feet does not require additional effort. Follow same procedure as for cubic feet.
Extent processed (gigabytes for digital materials processed)	Essential for calculating processing rates and quantifying the volume of resources made accessible. Necessary for annual reports and funding reports. Extent here indicates the volume of files prior to processing to reflect the extent of all the work performed, including review of materials that were de-accessioned. The final collection description shows the post-processing extent.
Total processing time to the nearest half hour	Essential for calculating processing rates and quantifying the volume of resources made accessible. Necessary for annual reports and funding reports.
Processing status	Essential for determining priorities and strategic plans because it shows what has or has not been processed.
Accession date	Essential for indicating how long it takes UNLV to make collections discoverable and accessible by calculating the time elapsed between the accession date and the date processing was completed. Average times help manage expectations of donors, administrators, and other stakeholders.

Data point	Purpose
Date processing completed	Essential for informing annual department reports on the number and volume of materials made accessible. Shows trends over time. The time elapsed between the accession date and the date processing was completed also indicates how long it takes UNLV to make a collection discoverable and accessible. Average times help manage expectations of donors, administrators, and other stakeholders
Processor(s) name(s)	Aids in creating subsets of data that support assessment of group projects. Supports individual's annual achievement reports.
Level of processing performed	Essential for accurate processing predictions and reporting. Tiers of processing granularity correlate with incremental increases in average hourly processing rates.
Nature of processing performed	Supports more accurate processing predictions and reporting. Nature of the work combined with level of processing provides a fuller picture of processing effort and accomplishment, which aids in assessing average hourly rates.
Processing notes	Free text field for processors to note variables that accelerated or slowed rates, explain work that was not completed, or provide information for future processors.

Operationalizing assessment requires information to be documented in a uniform manner that can be easily aggregated. To facilitate analysis, UNLV's proposed data framework limits free text options and relies primarily on numeric data and controlled values (see Appendix A). Moving forward, processing managers plan to periodically revisit the core data points to ensure their ongoing relevance across the program. Managers plan to collaboratively revise core data points very sparingly, however, changes to procedures will be made as often as needed. To maintain consistency in their emerging processing assessment methods, managers will strive to quickly communicate changes by updating manuals, templates, and staff training.

Added Value Data Points: Serving Special Needs

Building on the core data points, processing managers may apply a second layer of "added value" data points that apply to special formats, projects, or research needs.

Processing managers may independently add or adapt these secondary data points to meet specific needs within their areas of responsibility but have agreed to collaborate on any changes that cross domains. Staff may collect added value data periodically to inform or improve workflows, team management, or special projects. As of this writing, UNLV is considering collecting added value data for born-digital processing and digitization as preservation. Table 11 provides examples of a few added value data points that UNLV is currently considering.

Table 11. Potential added value data points and purposes for collecting them

Data point	Purpose
Performing appraisal and deaccessioning	Impacts processing rate based on the extent of materials that are being appraised and time spent consulting with curators and/or researching archival value
Restrictions review (separating or redacting digitally/physically)	Impacts processing rate significantly depending on scope/ size of collection. Average rates would be helpful for predicting wait times for users who request unprocessed/ minimally processed collections that contain sensitive materials.
Parse out processing activities into finer points, such as arrangement, digital file capture, file migration/ normalization, re-housing, retrieval and storage of boxes, etc.	Supports assessment of workflows and procedures. Helps identify bottlenecks and steps that need improvement. UNLV has identified needs for this information in the areas of born-digital processing and digitization as preservation.
Project name	May be used to identify team members and create subsets of data to assess the progress and priorities of group projects. Necessary for grant reporting.

The list of potential added value data points will be used selectively. The options will expand to meet situational needs, however, UNLV identified a few data points that it plans to strike from the list of options. Some data provided lessons that can be applied indefinitely and do not need repeating. Other data points proved extraneous. UNLV assessed a large enough sample of physical processing rates to confirm that a locally modified UC *Guidelines*-based chart will yield reasonable predictions for work

plans and funding proposals without routinely tracking processing activities at a granular level. Likewise, the metrics gathered for oral histories showed that there are consistent processing rates for these interviews due to their uniformity of content and processing level, therefore tracking metrics in this area is no longer necessary. Furthermore, there was data that proved difficult to interpret or had marginal value. For example, processors recorded time they spent tracking metrics. Those data points were not particularly enlightening or useful. Similarly, although processors recorded pre- and post-processing extents for the entire collection, those data points were available elsewhere. UNLV will rely on accession records for pre-processing extents, and final collection descriptions for the post-processing extent of the entire collection. Processors will record only the pre-processing extent of the portion of the collection they process.

Moving forward, as UNLV shifts from project-based to program-wide assessment and matures from improvisational to intentional data collection, it is critical for UNLV to conduct research into some of the more refined aspects of processing. Areas of further research may include iterative processing rates versus legacy processing rates; the impact of intensive conservation on processing rates; and born-digital processing rates. Once the archival backlog elimination project is complete, UNLV will begin tracking data starting with the core and added value data points in Table 10 and 11. By comparing iterative processing rates to legacy processing rates in the coming years, they hope to learn whether or not baseline processing during accessioning jumpstarts processing to a measurable degree. As previously mentioned, further analysis is also required for UNLV to draw reliable conclusions about the effect of intensive conservation issues on processing rates. UNLV plans to analyze specific collections more closely and continue to track time spent on traditional conservation work and digitization as a means of preservation to see how significantly intensive conservation affects overall processing rates. Finally, UNLV plans to continue tracking and refining metrics for born-digital processing to build a more substantial set of data for its research in this area. They will investigate the feasibility of predicting the time, effort, and labor needed to process these complex collections.

Conclusion

As UNLV Special Collections and Archives Technical Services' processing assessment methods have matured from one project to the next, the lessons learned during each project have propelled the program forward. Although their programmatic assessment of processing activities is still in the early stages, over the past four years, the team has collected, assessed, and used data in ways that affirm the value of metrics. At UNLV, processing metrics have been useful in mapping out work plans, timelines, and budgets that underpin projects and ongoing operations. Basing proposals and plans on past performance (rather than theoretical frameworks) adds a degree of reliability to financial and temporal predictions. Processing metrics have proven vital in demonstrating SCA Technical Services' achievements and

capabilities to UNLV administrators and external funding agencies. Graphs and charts generated from data help stakeholders visualize the return on their investment in the program. Hard data quantifies progress toward and completion of objectives, which builds confidence in the program and its people. A proven record of success increases funding opportunities. UNLV's use of processing data to demonstrate success, support long- and short-term planning, forecast budget and staffing needs, manage projects, and articulate value all echo the experiences reported by colleagues across the profession.

UNLV Special Collections and Archives Technical Services is still working to master the science of data collection and assessment. Their experience confirmed for them the importance of balancing the science of the numbers with the art of interpreting and contextualizing those numbers. By questioning counterintuitive data and statistical outliers, they were able to quantify the impact some variables had on their processing rates. They were able to measure the impact of variables related to assessment methodologies, collection characteristics (size), human factors (learning curves), and the nature of the work (processing levels). While they were able to gather empirical evidence on a handful of variables, there were many variables that escaped capture. In conclusion, because it can be difficult to isolate the degree of influence carried by every variable, it is essential to look beyond the numbers to understand how concurrent variables can impact processing outcomes. Where empirical evidence is lacking or counterintuitive, archivists should trust their own intuition. Archivists must consider all evidence, including anecdotal evidence, and use sound judgment when making "data-driven" decisions.

Appendix A. UNLV Processing Metrics Data Framework Pilot

Core data	Field format: data entry guidance
Collection identifier	Free text: XX-XXXXX
Collection name	Free text: Collection title proper
Extent (processed cubic feet for physical materials)	Numeric digits with decimals only: record to nearest hundredth (XX.XX)
Extent in linear feet (physical materials)	Numeric digits with decimals only: record to nearest hundredth (XX.XX)
Extent (processed gigabytes for digital materials)	Numeric digits with decimals only: record to nearest thousandth (XX.XXX)
Total processing time to the nearest half hour	Numeric digits with decimals only: record time in hours to the nearest half hour (X.XX)
Processing status	Select one from drop-down list of options: • Accessioning in progress • Processing in progress • Completed, but further processing recommended • Completed
Accession date	Numeric digits only: YYYY-MM-DD
Date processing completed	Numeric digits only: YYYY-MM-DD
Processor(s) name(s)	Free text: List names of processors in direct order separated by commas
Level of processing performed	Select one from drop-down list of options: • Description only (no physical work) • Level 1 (minimal—collection) • Level 2 (low—series or box) • Level 3 (Moderate—expedited file) • Level 4 (intensive—folder) • Level 5 (highly intensive—item)

Core data	Field format: data entry guidance
Nature of processing performed	Select one from drop-down list of options: • Quality assurance—revised or enhanced existing description (no physical work) • Finding aid only—created original description (no physical work • Addition—added physical material and updated description (must include extent of addition/accrual • Reprocessing—may include rehousing, conservation, description; may include undoing legacy work and re-doing it to current standards • Processing—work performed on previously untouched materials • Iterative/extensible processing—performing the next level of processing built upon quality baseline processing performed during accessioning
Notes	Optional free-text field: Note processing details such as impediments, accelerants, division of labor for group efforts, future processing recommendations, etc.