Archivaria
The Journal of the Association of Canadian Archivists

"Leveraging Technology to Facilitate Access"

Automated Description of the Mariposa Folk Festival's Born-Digital Performance Recordings

KATRINA COHEN-PALACIOS AND SARAH LAKE

Archivaria 98 (Fall 2024), pp. 72-101

Cite this article:

Cohen-Palacios, Katrina and Sarah Lake. "Leveraging Technology to Facilitate Access: Automated Description of the Mariposa Folk Festival's Born-Digital Performance Recordings." *Archivaria* 98 (Fall 2024): 72-101.
https://archivaria.ca/index.php/archivaria/article/view/13991

Leveraging Technology to Facilitate Access

Automated Description of the Mariposa Folk Festival's Born-Digital Performance Recordings

KATRINA COHEN-PALACIOS AND SARAH LAKE

ABSTRACT In recent years, the increasing volume of born-digital materials (i.e., those created digitally rather than digitized from analog originals) deposited in archives has fostered the development of new software-based tools and workflows for processing archivists. Archivists seeking practical guidance for preserving digital materials have a wealth of resources at their disposal, including many community-owned tools, workflows, and tutorials. This case study examines how archival standards and technological advances have influenced the semi-automated description of born-digital audio records through the lens of a recent project at the Clara Thomas Archives and Special Collections (CTASC) at York University Libraries (YUL). The Mariposa Folk Foundation Fonds, containing a large and growing collection of born-digital audio recordings, served as an opportunity to design and test a new software-aided descriptive workflow. The project leverages the programmable nature of born-digital materials in an attempt to streamline the time-consuming process for creating the item-level descriptions typically associated with sound recordings and born-digital records while also improving the discoverability of this material in the unmediated environment of online finding aids. This case study demonstrates how technology has influenced descriptive practices, with the advent of online finding aids providing increased access to archival descriptions, online databases permitting keyword searching, and tools to script metadata extracted from born-digital records enabling robust archival descriptions.

RÉSUMÉ Au cours des dernières années, le volume de documents nés numériques (ceux créés directement par des procédés numériques plutôt que par la numérisation) transférés vers des centres d'archives a mené au développement de nouveaux outils et procédés de travail pour les archivistes. Les archivistes désirant obtenir des conseils pratiques pour la préservation de documents numériques ont accès à une vaste gamme de ressources à leur disposition, incluant des outils créés par des gens du milieu, des procédés de travail et des tutoriels. Cette étude de cas examine comment les normes archivistiques et les avancées technologiques ont influencé la description semi-automatique de documents audio nés numériques à travers un projet récent à la Clara Thomas Archives and Special Collections (CTASC) de la Bibliothèque de l'Université York (YUL). Le fonds Mariposa Folk Foundation, qui contient un volume important et croissant de documents nés numériques, sert d'opportunité pour la conception et le test d'un nouveau logiciel de procédés de description assisté. Le projet exploite la nature programmable des documents nés numériques dans une tentative d'améliorer le temps consacré à la création de descriptions à la pièce typiquement associées aux enregistrements audio, tout en améliorant la découvrabilité de ce type de documents dans l'environnement ouvert d'instruments de recherche en ligne. Cette étude de cas démontre comment la technologie, avec l'avènement d'instruments de recherche en ligne, influence les pratiques descriptives. Cette technologie permet la recherche par mot-clé dans les bases de données en ligne et l'utilisation d'outils scriptant les métadonnées extraites des documents nés numériques, offrant ainsi des descriptions archivistiques robustes.

Introduction

In recent years, the increasing volume of born-digital materials (i.e., those created digitally rather than digitized from analog originals) deposited in archives has fostered the development of new software-based tools and workflows for processing archivists. Archivists seeking practical guidance for preserving digital materials have a wealth of resources at their disposal, including many community-owned tools, workflows, and tutorials.[1] The increasing availability of command-line tools, scripts, and both open-source and proprietary digital preservation applications have allowed archivists to partially automate several standard preservation activities for born-digital archives, such as recording and auditing fixity; identifying, characterizing, and validating file formats; and creating copies in access and preservation formats. However, less attention has been given to the ways in which software can aid archivists in other aspects of archival processing, such as description. While there is some evidence of practitioners using software and scripts to create semiautomated archival description workflows, this practice is not yet well documented, and more research is needed to explore the potential impact of this approach. This case study examines how archival standards and technological advances have influenced the semiautomated description of born-digital audio records, through the lens of a recent project at the Clara Thomas Archives and Special Collections (CTASC) at York University Libraries (YUL). The Mariposa Folk Foundation Fonds, which contains a large and growing collection of born-digital audio recordings, served as an opportunity to design and test a new software-aided descriptive workflow. The project leverages the programmable nature of born-digital materials in an attempt to streamline the time-consuming process of creating item-level descriptions for sound recordings while also improving the discoverability of these records in the unmediated environment of online finding aids.

The Clara Thomas Archives

Founded in 1970, CTASC is the custodian and steward of 8,000 metres of archival records and more than 100,000 items in special collections. The department supports the research and learning activities of the university's 52,000 under-

1 For example, Community Owned digital Preservation Tool Registry (COPTR) (website), last modified November 3, 2021, http://coptr.digipres.org/index.php/Main_Page.

graduate and graduate students and its 1,500 faculty in addition to responding to reference inquiries from the university's administrative staff, national and international researchers, and the general public. It is staffed by five full-time, permanent employees: the university archivist and head of the department, two archivists, an archives technician, and a cataloguer for special collections. Part-time student assistants and part-time archivists are also hired on a contractual basis to facilitate and support daily operations. Archivists often also rely on YUL colleagues, notably the digital assets librarian and the IT department, to operationalize the preservation of digitized and born-digital records. Through such collaborations, CTASC is able to preserve and provide access to the born-digital performance recordings of the Mariposa Folk Foundation (MFF).

The Mariposa Folk Foundation Fonds

This fonds was initially acquired in 2007 and consisted of approximately 120 metres – or 300 bankers boxes – containing the records documenting the history of the internationally recognized Canadian music festival. Once conservatively valued at $1.5 million, this large fonds donation offers an opportunity to study North American folk music during the second half of the 20th century through the festival's documentation of its history.[2]

Ruth McVeigh and her husband Dr. Casey Jones founded the festival in 1961. It has grown to become one of the longest-running music festivals in Canadian history and is known for its impact on the Canadian music scene. Many notable artists launched their careers at MFF, including Gordon Lightfoot, Joni Mitchell, Bruce Cockburn, Ian and Sylvia Tyson, and Murray McLauchlan.[3] Since MFF's inception, its organizers have shown an interest in preserving its history through efforts such as the publication of *For What Time I Am in This World: Stories from*

2 York University Media Relations, "York University Acquires Mariposa Folk Fest Archives," news release, *News@York*, June 6, 2007, http://news.yorku.ca/2007/06/06/york-university-acquires-mariposa-folk-fest-archives/.

3 York University, "Mariposa Archives are Safe, says Orillia Paper," *YFile*, June 8, 2007, http://yfile.news.yorku.ca/2007/06/08/mariposa-archives-are-safe-says-orillia-paper/. The festival also nurtured the early years of North American folk music and showcased the artistry of many performers, including

> The Travellers . . . Bruce Cockburn, Arlo Guthrie, Richie Havens, Murray MacLachlan, Pete Seeger, James Taylor, Stompin' Tom Connors, Rita MacNeil, Tom Paxton, Colin Linden, Jane Siberry, Kate and Anna McGarrigle, Buffy Sainte-Marie, Emmylou Harris, Melissa Etheridge, Shirley Eikhard, Bonnie Raitt, John Hiatt, Lyle Lovett, Loreena McKennitt, Holly Near, Tracy Chapman, Moxie Fruvous and the Barenaked Ladies.

H.J. Kirchhoff, "Mariposa's Future Safe, but Past Fading: Before 35 Years of History Can Be Properly Stored, Someone Must Pay for Cataloguing," *Globe and Mail*, March 30, 1996.

Mariposa (1977), by Bill Usher and Linda Page-Harpa, which recounts the festival's first 11 years. More importantly, they also held onto the festival's early operational documents for safekeeping. Unfortunately, these archival records were stored in an old building at King and Dufferin in Toronto,[4] and there were fears that the possibility of the overhead sprinklers or the building's main sewer line in the storage space drowning the records would one day become an inevitability.[5] These fears necessitated finding a new home for the MFF archives. While it might have been ideal to provide access to the records at the festival's home grounds in Orillia, Ontario, no local facilities had the capacity or the resources to properly care for the material at the time.[6] The MFF board of directors then spent over a decade looking for a home where these records could be preserved, and they soon learned that approaching institutions to take on the stewardship of the archives without having a preliminary inventory of the material was challenging. A large portion of the material lacked documentation, making it difficult to quickly identify and evaluate the contents of boxes. A festival volunteer describing the state of the archives said there was "no year, no artists, nothing, like the technician just took it off the recorder and threw it in the box."[7] After reviewing proposals from several locations, the MFF board decided to donate the records to York University, as it "had a renowned music program and an archival department within their library that is world-class."[8]

In the early days following the acquisition, CTASC dedicated its processing of the fonds to gaining intellectual and physical control of the enormous mountain of material. An archives technician, part-time archivists, and student assistants sifted through the material to gain a quick grasp of its contents.[9] The material was then intellectually arranged into 12 preliminary groupings, which would later form series based on functions (operational, financial, press and publicity, artist submissions, reference and resource, and memorabilia and ephemera)

4 Michael Hill, *Mariposa Folk Festival: A History* (Toronto: Dundurn Press, 2017), 176.

5 Kirchhoff, "Mariposa's Future Safe, but Past Fading."

6 Hill, *Mariposa Folk Festival*, 176.

7 Kirchhoff, "Mariposa's Future Safe, but Past Fading."

8 Hill, *Mariposa Folk Festival*, 176. See also Kirchhoff, "Mariposa's Future Safe, but Past Fading."

9 Eric Fillion, "Curating for Change: An Interview with Anna St. Onge," episode 3, *Curating for Change: The Work That Music Festivals Do in the World* (podcast), 29:00, October 6, 2022, http://www.whatmusicfestivalsdo.ca/resources/curating-for-change-an-interview-with-anna-st-onge-episode-3/.

and formats (programs and newsletters, photographs, posters, moving-image recordings, and audio recordings). This process also helped the archivists locate archival gems to showcase during the festival's upcoming 50th anniversary. With the celebration only a few years away, departmental priorities shifted from preliminary processing of the fonds to preparation of the records for these celebratory outreach initiatives. One such activity included completing grant applications for federal funding from the Archival Community Digitization Program, which would allow the archives to hire contract metadata librarians and graduate music students in 2009.[10] These librarians and students digitized, described, and sought permission to upload 600 items online for public access. Some of these digitized items were also included in an online exhibit titled Mariposa: Celebrating Canadian Folk Music.[11] The exhibit highlights selected festival programs, promotional materials, photographs, and live performance recordings to share the history of the festival's first 20 years.[12]

Today, the annual festival continues to donate records created during its operational activities to CTASC on an almost annual basis. Previously, these donations had typically consisted of recordings – on VHS and MiniDV videocassettes, DVDs, audio reels, and audio cassettes – capturing live performances from the previous year's festival. In 2010, these donations switched to a digital format, and over the past decade, thanks to the hard work of festival audio engineer Sy Potma, the festival has deposited approximately 600 gigabytes (GB) of born-digital records for perpetual preservation. With regular accruals expected for as long as the festival continues, the ongoing resource requirements for describing these digital recordings and making them available are considerable. The continuous donations also present an opportunity to rethink approaches to archival description, including how one might expedite the process by leveraging the born-digital nature of these recordings.

10 York University, "Exhibit Organized by York Archivist Looks at 50 Years of Mariposa," YFile, July 12, 2011, http://yfile.news.yorku.ca/2011/07/12/exhibit-organized-by-york-archivist-looks-at-50-years-of-mariposa/.

11 See Anna St. Onge, "Mariposa: Celebrating Canadian Folk Music," York University Libraries, accessed June 23, 2022, http://archives.library.yorku.ca/exhibits/show/mariposa/.

12 York University, "Mariposa Digitization Project featured on CBC's Metro Morning," News from the Clara Thomas Archives & Special Collections, July 5, 2010, http://ctasc.blog.yorku.ca/2010/07/05/news_mdp/; CBC Arts, "Mariposa Looks Back at 50 Years of Folk Fest," CBC News, July 6, 2010, https://www.cbc.ca/news/entertainment/mariposa-looks-back-at-50-years-of-folk-fest-1.929269.

Literature Review

While born-digital materials are increasingly present in archival collections, leading archivists to adapt usual processes for processing and description, methods for such processes remain relatively ad hoc.

Descriptive Standards

Established in 1990, the *Rules for Archival Description* (RAD) is the Canadian archival descriptive standard and provides a multi-level, top-down framework to reflect the context of archival records.[13] With its last revision occurring in 2008, RAD presents many challenges for archivists describing digital records. In particular, chapter 9, which deals with records in electronic form, has been criticized for lacking the necessary granularity to record basic information about digital objects, such as extent, in a consistent way.[14] A 2016 survey by the Archives Association of Ontario (AAO) shows that Canadian archivists are well aware of these challenges; most participants identified RAD's handling of digital materials as one of its largest vulnerabilities over time.[15] This gap in descriptive standards for digital archives is not a uniquely Canadian problem, as US-based digital archivists have noted similar limitations with *Describing Archives: A Content Standard* (DACS).[16] In the absence of formal standards, some practitioners have turned instead to community-driven initiatives for guidance on describing born-digital materials. Examples of Born Digital Description in Finding Aids, a project by the US-based Digital Library Federation's Born-Digital Access Working Group, represents a snapshot of descriptive practices relating to born-digital archival

13 For an overview of the standard's development, challenges, and attempted revisions, see Richard Dancy, "RAD Past, Present, and Future," *Archivaria* 74 (Fall 2012), 7–41, https://archivaria.ca/index.php/archivaria/article/view/13406.

14 For a breakdown of the issues with RAD's chapter on electronic records, see Kat Timms, "The Devil Is in the Details: Describing Born-Digital Records Using the Rules for Archival Description" (paper presented at SAA Conference, New Orleans, LA, August 2013), http://files.archivists.org/conference/nola2013/materials/701-TimmsA.pdf.

15 Danielle Robichaud, "Rules for Archival Description (RAD) Survey Results," Archives Association of Ontario, June 14, 2016, http://www.aao-archivists.ca/News/4074646.

16 Annalise Berdini, Charles Macquarie, Shira Peltzman, and Kate Tasker, "Describing Digital: The Design and Creation of a Born-Digital Archival Description Standard at the University of California Libraries," *Journal of Western Archives* 9, no. 1 (2018): 1–24, http://doi.org/10.26077/e90b-2e7b.

materials in 2021 and 2022.[17] The project's website provides examples illustrating how born-digital materials are described in finding aids as well as a list of existing online descriptive guidelines from various kinds of institutions. The emergence of these ad hoc community resources shows that there is a pressing need in the field for common practices in describing born-digital materials.

Software and Scripting Archival Description

In recent years, archivists have developed innovative approaches to partially automate the processing of born-digital archival materials. A typical digital preservation workflow includes the use of software applications or custom scripts to perform preservation actions such as verifying fixity and transcoding to preservation formats upon digital objects. Some institutions have gone a step further and have incorporated similar technology into their descriptive practices. For example, the authors of the *UC Guidelines for Born-Digital Archival Description* suggest using software like Directory List & Print Pro to automatically generate file listings to supplement description, but they emphasize that these should not serve as the only description, because filenames can be vague, inaccurate, and misleading.[18] Similarly, the Canadian Centre for Architecture (CCA) Digital Archives Processing Manual recommends, "'Let the bits describe themselves': Machine-actionable information such as extent, file formats, and filesystem metadata should be captured automatically, never calculated by hand, and rarely transcribed."[19] The CCA GitHub repository also features custom Python scripts to auto-populate the scope and content note fields in processing spreadsheets.[20] Web archiving is another area where archivists have leveraged software to automate aspects of archival description. The Library of Congress, for instance,

17 The project lists the MFF Fonds finding aid as an example of born-digital description. See Born Digital Description Project, "Examples of Born Digital Description in Finding Aids," Born Digital Description in Finding Aids, accessed June 23, 2022, http://borndigitaldescriptioninfindingaids.wordpress.com/.

18 Annalise Berdini, Charles Macquarie, Shira Peltzman, and Kate Tasker, *UC Guidelines for Born-Digital Archival Description* (UC Office of the President: University of California Systemwide Libraries: 2017), 31, https://escholarship.org/uc/item/9cg222jc.

19 Canadian Centre for Architecture, "Digital Archives Manual: Description of Born-Digital Archives," GitHub, September 25, 2019, http://github.com/CCA-Public/digital-archives-manual/blob/master/guides/description.md#principles-and-practical-guidelines-for-description-of-born-digital-archives.

20 Canadian Centre for Architecture, "CCA Processing Scripts: Objects-dir-to-sip.py," GitHub, April 11, 2018, http://github.com/CCA-Public/cca-scripts/blob/master/processing_scripts/objects-dir-to-sip.py.

uses Python scripts with application programming interfaces (APIs) to automate the description of its web archives.[21] Although some archival processing manuals and project documentation show evidence of archivists using scripting and applications to partially automate archival description, there is still an important gap in scholarly literature documenting these practices.

While using scripts and software to partially automate archival description can be a powerful tool to increase access to born-digital archives, for many archivists, a lack of technological infrastructure and expertise represents an important barrier to implementation. In his proposal for a computer-assisted approach to describing web archives, Gregory Wiedeman notes that librarians and archivists working with centralized IT departments commonly have to request server access and privileges to install and test software.[22] These restrictions, driven by the need to mitigate risks in the current information security landscape, limit "both the types of tools [archivists] can use and their ability to explore and learn."[23]

Discoverability of Born-Digital Records

Integrating technical metadata into archival descriptions can provide helpful information for researchers, but deciding which metadata is appropriate to include presents another challenge for archivists. In his 2016 study of the application of descriptive practices to born-digital materials, John Langdon notes, "While the literature recognizes the limitations of descriptive standards and the need for technical metadata for the management of digital archives, one aspect which is not fully explored in the literature is the question of precisely what additional information needs to be given to the user."[24] An entire record of all the preservation actions performed upon the objects could be included in the finding aid, but this might not be needed or understood by the user.[25] The *UC Guidelines for Born-Digital Archival Description* suggest that it is up to

21 Grace Thomas, "More Web Archives, Less Process," *The Signal* (blog), Library of Congress, August 3, 2018, https://blogs.loc.gov/thesignal/2018/08/more-web-archives-less-process/.

22 Gregory Wiedeman, "Describing Web Archives: A Computer-Assisted Approach," *Journal of Contemporary Archival Studies* 6 (2019), 15, http://elischolar.library.yale.edu/jcas/vol6/iss1/31.

23 Wiedeman, 15.

24 John Langdon, "Describing the Digital: The Archival Cataloguing of Born-Digital Personal Papers," *Archives and Records* 37, no. 1 (2016): 37–52, 44, http://doi.org/10.1080/23257962.2016.1139494.

25 Langdon, 45.

the archivist to determine what level of descriptive and technical metadata is appropriate to include in a finding aid so as to not overwhelm and confuse researchers.[26] The authors advise archivists to keep all the metadata somewhere else and provide access to it upon request.[27] Richard Pearce-Moses and Susan E. Davis, in *New Skills for a Digital Era*, argue that "in a digital world, hierarchies often disappear, and item-level metadata predominates," in contrast to traditional archival description, where data is aggregated and hierarchical.[28] This is also surfaced by Magnus Berg, who explains the multitude of challenges and misunderstanding involved in presenting hierarchical descriptions in an online research environment.[29]

In their foundational article on the more product, less process approach – its ubiquity evidenced by the common use of the shorthand MPLP – Greene and Meissner advocate for a minimal approach to processing in response to the problem of growing backlogs of unprocessed and therefore inaccessible collections.[30] In a follow-up opinion piece, Greene applies that same logic to archival description, arguing that in the interest of expediting access for users, most archival materials made available online should not be described with item-level metadata, which he refers to as a "boutique" digitization practice.[31] Responses to MPLP and its application to digitization are mixed, with some archivists arguing that a boutique approach can be the most appropriate choice for some materials.[32] No matter which side of the MPLP debate one falls on, the problem it attempts to mitigate, the ever-increasing backlog that plagues most archives,

26 Berdini, Macquarie, Peltzman, and Tasker, *UC Guidelines for Born-Digital Archival Description*, 5.

27 Berdini, Macquarie, Peltzman, and Tasker, 6.

28 Richard Pearce-Moses and Susan E. Davis, "Knowledge and Skills Inventory," in *New Skills for a Digital Era: A Colloquium Sponsored by National Archives and Records Administration, Society of American Archivists, Arizona State Library, Archives and Public Records*, ed. Richard Pearce-Moses and Susan E. Davis (Chicago: Society of American Archivists, 2008), 6, http://files.archivists.org/pubs/proceedings/NewSkillsForADigitalEra.pdf.

29 Magnus Berg, "A 'Major Technological Challenge': Multi-Level Description and Online Archival Databases," *Emerging Library & Information Perspectives* 4, no. 1 (2021): 62–87, https://doi.org/10.5206/elip.v4i1.12529.

30 Mark Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *American Archivist* 68, no. 2 (2005): 208–63, 208, http://doi.org/10.17723/aarc.68.2.c741823776k65863.

31 Mark Greene, "MPLP: It's Not Just for Processing Anymore," *American Archivist* 73, no. 1 (2010): 175–203, 193–94, http://doi.org/10.17723/aarc.73.1.m577353w31675348.

32 Shan C. Sutton, "Balancing Boutique-Level Quality and Large-Scale Production: The Impact of 'More Product, Less Process' on Digitization in Archives and Special Collections," *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 13, no. 1 (2012): 50–63, 54, http://doi.org/10.5860/rbm.13.1.369.

is well documented. To make matters worse, the sheer volume of information being created in our digital age also poses a problem of scale for the archivists tasked with making those materials accessible to researchers. As evidenced by the processing manuals referenced in this section, some archivists have already started to tackle this problem by using technology to partially automate the processing and description of born-digital records, but this practice is not yet widespread. As archives are flooded with an increasing quantity of born-digital materials, further investigation into strategies to maximize efficiency and optimize the use of archivists' time is urgently needed.

Describing the MFF's Born-Digital Recordings

With technological advances and the implementation of archival standards, the presentation and use of archival finding aids at CTASC and many other repositories across the country has changed over time. Researchers now have more access than ever before to robust descriptions in archival finding aids to assist their searches for relevant material. The development of online finding aids has perhaps been the most impactful advancement in the past few decades, as it gives researchers from across the globe immediate access to file inventories and descriptive records at any time of the day. Prior to the existence of finding aids on institutional websites, researchers needed to consult books such as the *Guide to the Fonds d'Archives and Collections in the Holdings of the York University Archives*[33] to learn about the availability of archival holdings at individual institutions. Such books often contained only fonds-level descriptions, and researchers needed to consult with archivists and travel to reading rooms to review paper file inventories or the institutions' off-line, proprietary descriptive databases. Unfortunately, this meant that only a privileged few had the opportunity to discover whether an institution held material that matched their research interests. It was thanks to the advent of the Encoded Archival Description (EAD) XML-based standard, created in 1998, that archivists could increasingly share, exchange, and circulate archival resources with researchers through online databases.

33 Barbara L. Craig and Peter D. James, *Guide to the Fonds d'Archives and Collections in the Holdings of the York University Archives* (Toronto: ECW Press, 1995).

Descriptive Software and Practice at CTASC

CTASC's first efforts to make its descriptions available through the Internet occurred in early 1997, a year prior to the introduction of EAD.[34] The department exported fonds-level descriptions from its Inmagic database, generated machine-readable cataloguing (MARC) format records, and uploaded them to the library's catalogue, which ran on SirsiDynix software. In 1998, CTASC adopted the EAD standard and spent two years developing a customized platform, called the EAD encoded system (EES), to provide access to hierarchical descriptions of its holdings from the fonds to the file level.[35] This system borrowed the base code of the initial Archeion software application developed by the AAO.[36] EES generated finding aids in the form of PDF and HTML outputs for the CTASC website, as well as "an EAD-based record for uploading to Archeion which would in turn generate a MARC record for uploading to the Archives Canada portal, [and] . . . a customized MARC record that could be uploaded to the library catalogue that could in turn link back to the full finding aid on the CTASC server."[37] Just as one might imagine, uploading descriptions to a variety of platforms helped researchers discover the availability of these archival records. Without the automation provided by EES, this would have been a rather labour-intensive endeavour.

34 Suzanne Dubeau, "Clara Thomas Archives and Special Collections: A Retrospective: 1998–2018" (unpublished report, June 9, 2023), 17.

35 Suzanne Dubeau, "York University Adopts EAD Standard," *Off the Record* 15, no. 3 (1998): 4–5, http://aao-archivists.ca/otr.

36 Early in the first decade of the 21st century, a period of rapid growth dedicated to promoting and displaying finding aids online, there were both individual and collective efforts by institutions to develop national and provincial registries of fonds. This centralization allowed researchers to use a single information portal rather than scouring through each individual archival institution's website. During this massive endeavour, "York's university archivist Ken Haworth and [Suzanne Dubeau] were heavily involved in laying the groundwork for Ontario via the newly established Automated Union List of Ontario (AULO) special committee of the [AAO]. . . . In June 2000 the fruits of all this effort resulted in the launch of Archeion . . . which feeds into the Canadian archival search port now known as ArchivesCanada." Dubeau, "Clara Thomas Archives and Special Collections," 18.

37 Dubeau, "Clara Thomas Archives and Special Collections," 18.

Nearly a decade after this initial push to share finding aids in the unmediated online environment, in 2010, the preliminary finding aid of the MFF Fonds, consisting of a file inventory for the artist submission file series, was made available to researchers through EES. The processing of the remaining records continued as a work in progress over the years. Draft descriptions for these boxes were made available in the reading room in the form of printed-out spreadsheets and notes from preliminary processing. A decade later, an undergraduate public history student on placement focused on curating a display case to celebrate the festival's 60th anniversary.[38] Her work initiated the migration of the MFF file inventory in EES and of the printed descriptive spreadsheets to Access to Memory (AtoM) in early 2020.[39] Among many features, including a standard compliant metadata schema, AtoM provided increased searchability and an ability to interlink related records. AtoM also offered the ability to bulk upload file inventories using comma-separated values (CSV) files. It is through this functionality that the preliminary MFF file inventories for posters, programs, and photographs were uploaded to the database. As well, the 10 accessions of born-digital files were each given descriptive records to alert researchers to the material's availability. A few months later, progress on increasing the discoverability of the MFF Fonds came to an abrupt halt. The COVID-19 pandemic limited access to physical holdings, and the remote work environment meant the reallocation of resources that were typically dedicated to processing analog records. Work in the remote environment during the pandemic's early days pivoted to digital platforms and digital records.

Using Technology to Facilitate Robust Description

Recent research on information-seeking behaviour shows that the way users access archival materials is changing. Online finding aids have changed the way archival records are discovered, and users increasingly rely on keyword searching

38 This student also translated her exhibit into French to create a Wikipedia article for francophones. See Wikipedia contributors, "Festival folklorique Mariposa," *Wikipedia*, last updated June 30, 2023, http://fr.wikipedia.org/wiki/Festival_folklorique_Mariposa.

39 In recognition of the precarity of CTASC's online finding aids in an ageing custom software environment, the initial migration of finding aids from EES to AtoM occurred in 2014. However, the technical and resource debt stalled the migration of several finding aids, including that of the MFF Fonds. As of 2017, only 100 finding aids had been published in AtoM; fortunately, this process was accelerated in 2018, when it became possible to update entire finding aids (usually more than 100 records) through spreadsheets, using the export/import function. As of June 2023, 500 finding aids of more than 700 fonds are now available in AtoM.

due to the ubiquity of search engines like Google.[40] Winters and Prescott argue that Google-style searching is "particularly unsuited for establishing the scope of a dataset or digital archive, as it encourages researchers to look for what they know to be there rather than to seek the unknown or to identify gaps and absences."[41] In an unmediated search environment, archivists no longer occupy the role of mediator and must often rely on robust descriptive practices with multiple access points to connect researchers with records. This shift will undoubtedly continue to shape descriptive practices.

In the case of the MFF, if a researcher was searching for a specific performance from the festival and their keyword search for a specific artist was unsuccessful, they might contact the archives for further assistance. An archivist answering the reference question might also be unsuccessful if they overlooked the born-digital files due to the disconnect between file names (i.e., session titles) and the contents of those files (the performances of several artists). Ultimately, CTASC decided to create item-level descriptions to help avoid the unnecessary layer of complexity by relying on festival programs for contextual information. Yeo argues that "most preservation and access strategies for digital records require item-level metadata, captured alongside or within the records."[42] Proponents of an MPLP approach to archival processing might argue that item-level description is not necessary to make born-digital or digitized records discoverable,[43] and while this may be applicable to groupings of photographs or textual materials, audiovisual records pose a particular challenge for an MPLP approach. Due to the temporal nature of audio recordings, a researcher cannot quickly glance through a record to locate material of interest as they would with a folder of papers or photographs. So, while most archival processing guides advocate for collection-level or series-level description in general, moving-image and sound collections are exceptional in that they are typically described at the item level.[44] Leab

40 Jane Winters and Andrew Prescott, "Negotiating the Born-Digital: A Problem of Search," *Archives and Manuscripts* 47, no. 3 (2019): 391–403, http://doi.org/10.1080/01576895.2019.1640753.

41 Winters and Prescott, 393.

42 Geoffrey Yeo, "Continuing Debates about Description," in *Currents of Archival Thinking*, 2nd ed., ed. Heather MacNeil and Terry Eastwood (Santa Barbara, CA: Libraries Unlimited, 2016), 183.

43 Greene, "MPLP," 193–94.

44 Anthony Cocciolo, *Moving Image and Sound Collections for Archivists* (Chicago: Society of American Archivists, 2018), 27.

Martin highlights the importance of creating item-level access points for moving images in particular, arguing that "because moving images represent a body of 'non-browsable' entities, access points serve an essential function in assisting users to locate the materials they need. The description is a surrogate for the item; the access points link the user to the surrogate."[45] This "non-browsable" quality also influenced the decision to create name access points for each performing entity in the MFF Fonds.

As CTASC had a decade of recordings to describe in detail and knew that future accruals would be coming from festival organizers, investing resources in partial automation to expedite processing seemed like a reasonable solution. However, with neither the technical expertise to create or modify Python scripts nor a familiarity with the command-line interface (CLI), CTASC could not implement instructions from published digital description guides that involved custom scripts – such as those outlined in the CCA's Digital Archives Processing Manual.[46] CTASC identified OpenRefine, a "free, open-source tool for working with messy data,"[47] as an alternative with a shallower learning curve. Open Refine is commonly used in libraries and archives for cleaning, migrating, and extending metadata.[48] It was possible to exploit the graphical user interface (GUI) in OpenRefine, which created a bridge between the CLI and experimentation with Python, GREL (General Refine Expression Language), and other programming languages that used the transformation functions.[49] For archivists,

45 Abigail Leab Martin, ed., *AMIA Compendium of Moving Image Cataloguing Practice* (Beverly Hills, CA: Association of Moving Image Archivists, 2001), 180.

46 For examples, see Canadian Centre for Architecture, "Entering File-Level Description in Processing Spreadsheets," in "Digital Archives Manual: Description of Born-Digital Archives," GitHub, September 25, 2019, http://github.com/CCA-Public/digital-archives-manual/blob/master/guides/description.md#entering-file-level-description-in-processing-spreadsheets; Kelli Bogan, "Adobe Bridge, OpenRefine, & Ruby: Extracting Large Quantities of Metadata and Outputting to an XML File," *bloggERS!: The Blog of SAA's Electronic Records Section*, October 2, 2018, http://saaers.wordpress.com/2018/10/02/adobe-bridge-openrefine-ruby-extracting-large-quantities-of-metadata-and-outputting-to-an-xml-file/.

47 OpenRefine (website), accessed June 22, 2022, http://openrefine.org/.

48 Library Carpentry, "OpenRefine: Summary and Setup," Library Carpentry, accessed June 22, 2022, http://librarycarpentry.org/lc-open-refine/; Library Juice Academy, "Using OpenRefine for Library Metadata," Library Juice Academy: Online Professional Development for Librarians, accessed June 22, 2022, http://libraryjuiceacademy.com/shop/course/130-using-openrefine-library-metadata/.

49 As Greg Bak points out, "Digital archiving required then, and requires now, the ability to control the machine at deeper levels than that of a common user. This requires an ability to think in terms of the machine's component parts and to control the machine in its own language – taking us back to Hugh Taylor's dream of archivists who

OpenRefine was already assisting with partial automation of legacy finding aid migration and digitization projects.[50] It was also facilitating the integration of reconciliation services for name authorities.[51]

With no previously created guide or instructions that readily met the requirements for the MFF's born-digital records, CTASC needed to build its own. This involved breaking down each step of the descriptive process to partially automate the creation of robust descriptions of the festival's live performances. This predetermined set of operations could then be reapplied to the 10 accessions in the backlog and all donations in the near future. To further break down the workflow into sustainable and manageable chunks, CTASC began using an already familiar

know 'the language of the computer like [their] native tongue.'" Greg Bak, "How Soon Is Now? Writings on Digital Archiving in Canada from the 1980s to 2011," *American Archivist* 79, no. 2 (2016): 283–319, 309, https://doi.org/10.17723/0360-9081-79.2.283.

50 For examples of legacy finding aid migration, see Cassie Schmitt, "Clean Up: Dates and OpenRefine," *Chaos —> Order: Four Archivists' Battles with Masses of Legacy Description* (blog), April 24, 2014, http://icantiemyownshoes.wordpress.com/2014/04/24/clean-up-dates-and-openrefine/; Bentley Historical Library Curation Team, "Normalizing Dates with OpenRefine," *ArchivesSpace-Archivematica-DSpace Workflow Integration project* (blog), Bentley Historical Library, June 19, 2015, http://archival-integration.blogspot.com/2015/06/normalizing-dates-with-openrefine.html; Graham Jevon, "Clean. Migrate. Validate. Enhance. Processing Archival Metadata with Open Refine," *Digital Scholarship Blog*, British Library, April 21, 2020, http://blogs.bl.uk/digital-scholarship/2020/04/clean-migrate-validate-enhance-processing-archival-metadata-with-open-refine.html; and, finally, Ruth Kitchin Tillman, "Learning Cell Cross in OpenRefine," *Ruth Kitchin Tillman*, March 26, 2020, http://ruthtillman.com/talk/cell-cross-webinar-2020-03/.

Kitchin Tillman's tutorial demonstrates how OpenRefine can assist with linking finding aid URLs to MARC records in the library catalogue. For an alternative, see Alicia Detelich, "Large-Scale Date Normalization in ArchivesSpace with Python, MySQL, and Timetwister," *Code4Lib* 44 (May 2019), https://journal.code4lib.org/articles/14443; and Bria Parker, "Managing Archival Metadata (is Maddening)" (presentation, University of Maryland, College Park, MD, April 21, 2017), https://doi.org/10.13016/M22Z66.

For examples of digitization projects, see Association of Canadian Archivists, "Working with Archival Texts as Data: Post-OCR Error Correction with OpenRefine," Association of Canadian Archivists, May 11, 2021, http://archivists.ca/event-4246843; and Sherman Centre for Digital Scholarship, "Pre-Processing Digitized Texts," Sherman Centre for Digital Scholarship, accessed September 6, 2023, https://scds.ca/product/pre-processing-digitized-texts/. Without OpenRefine, archivists would need to follow a more technologically advanced tutorial, such as Laura Turner O'Hara, "Cleaning OCR'd Text with Regular Expressions," *Programming Historian*, February 1, 2021, http://programminghistorian.org/en/lessons/cleaning-ocrd-text-with-regular-expressions.

51 See Stephanie Bolliger, Brigitte Brüderlin, Michael Gasser, Julia Lyskawa, Petra Maier, and Lothar Schmitt, *How to Link Your Institution's Collections to Wikidata?: A Short Manual to a Semi-Automatic Way of Using the "Archives At" Property* (P485) (Zurich: ETH Library; Central Library of Zurich; Swiss National Library, 2020), http://doi.org/10.3929/ethz-b-000393724. Without OpenRefine, the alternative would include using Python scripts to reconcile name authorities. See Eric Hanson, "VIAF and DBpedia Reconciliation Python Scripts," GitHub, December 12, 2022, https://github.com/ehanson8/linked-data-reconciliation-Python-scripts.

data manipulation software – Excel – as part of the semiautomated process.[52] This enabled the archivists to avoid becoming overwhelmed and fumbling in OpenRefine while still acquiring the prerequisite knowledge and skill set to leverage the software's functionality.

This first phase was supported in 2022 by two public history placement students, whose 240 hours of work contributed toward building and testing the workflows used to build semiautomated descriptions of Mariposa's born-digital records.[53] This placement was designed to further develop the digital description workflow drafted for the Home Made Visible (HMV) project. HMV was a nationwide project, led by the Regent Park Film Festival in partnership with Charles Street Video and York University Libraries, to digitize donated home movies of BIPOC families in Canada. Like the MFF Fonds, this donation consisted of a homogenous group of archival material (home movies), donated in several batches, and required item-level descriptions to facilitate access to the records. The workflow for describing these 268 digitized home movies required designing an extent template based on archivist and researcher needs (an intellectual challenge due to the previously discussed limitations of RAD's chapters on electronic records); manually extracting technical metadata from each digital file; and manually collating the information to capture extent.[54] The MFF workflow also leveraged the work of YUL's digital scholarship and metadata librarian, David Montgomery, who spent countless hours researching more than 900 digitized recordings in the Music Gallery Fonds and capturing performer names to create robust descriptions with name authorities and scope notes. For the MFF, the goal was to consolidate these two workflows and enrich them by developing an automated extent and transforming name access points into scope and content notes. To achieve this, an archivist from CTASC, in collaboration

52 Normally, CTASC archivists follow Artefactual's recommendation to draft descriptions in LibreOffice Calc; however, MS Teams and the online shareability of Microsoft Office 365's version of Excel offered an online environment for supervising students working remotely.

53 For one student's perspective, see Sundus Saba, "My Encounter with Digital Records at the Clara Thomas Archives and Special Collections," *In the Field: The ACA Blog* (blog), May 5, 2022, http://archivists.ca/Blog/12768228. For more information about the course, see York University, "HIST 4840 | Public History Placement," Department of History, accessed June 22, 2022, http://www.yorku.ca/laps/hist/undergrad/placement-course/.

54 See Katrina Cohen-Palacios, "Home Made Visible: Partnering with a Film Festival to Preserve IBPOC Home Movies" (presentation, Archives Association of Ontario Conference: Building Bridges, Connecting Communities, October 2020), 14–16, http://hdl.handle.net/10315/38531.

with students during their placements in 2022, processed three of the 10 MFF born-digital accessions. First, the archivist processed the accession and documented the steps taken to describe the record, and then, the students tested and refined the workflows with their proper accessions.[55]

This work primarily involved manipulating metadata and focused predominantly on three fields: (1) extent and medium, (2) scope and content notes, and (3) name access points. In the first field, technical metadata on the extent and medium were compiled and integrated to display pertinent information for researchers. These metadata were extracted as CSV files from MediaInfo, an open-source GUI application used to pull technical metadata from video and audio files; the metadata could then be displayed and exported in different formats.[56] The metadata extracted from the files included information about the quality of the sound recording (e.g., sampling and bit rate), file format, duration, and file size. In the second field, name access points were used to create scope notes. It was important to include these as AtoM search results do not display name access points but only scope notes. The third and final manipulation of metadata leveraged linked open data to reconcile, identify, and match name authorities in Wikidata, Virtual International Authority File (VIAF), and other services.

The challenge with using Excel to automate these tasks is that it can be time-consuming to repeat the process, even after the steps have been determined and the instructions documented. The final stage of the placement involved translating the Excel functions and formulas into OpenRefine operations. Using OpenRefine, one can export the multiset process of operations as a single JavaScript Object Notation (JSON) text and apply it in OpenRefine to future accession descriptions. By the end of this initial pilot, three of the 10

55 To create name authorities, performer names and other important information was copied and pasted from the program (PDF file) to the row for the associated file in the spreadsheet. The CONCAT function was used to string text to form the following scope note: "Item is an audio recording of the [session column] session performed at the [stage column] stage on [date column] and may include performance(s) by [name access points column]." The name access column was then used to create a name authority spreadsheet for importing the text to columns and executing VSTACK, SORT, and conditional formatting functions. Associated metadata was then pulled from sources such as the Virtual International Authority File (VIAF) and Wikidata. The population of the extent field was automated through an XLOOKUP formula that used CONCAT, TEXTJOIN, and mathematical operations and matched the two spreadsheets (AtoM and MediaInfo). The find and replace function was also used to mend data throughout all three descriptive elements.

56 MediaArea, "MediaInfo," MediaArea, accessed June 22, 2022, http://mediaarea.net/en/MediaInfo.

accessions had been processed, resulting in the creation and addition of 353 item-level descriptions of born-digital sound recordings and 184 name authorities to the MFF Fonds finding aid.

Navigating Standards to Automate Robust Description

With a large and ever-growing backlog of born-digital records, CTASC needed to expedite the process of creating item-level descriptions of digital audio recordings. This backlog created the illusion that the born-digital performance recordings had been forgotten for over a decade. In fact, it had taken nearly five years to identify and acquire the technological skills necessary to process these records in a sustainable manner. This period also saw technological advances, including the development of associated tools and software applications that made entry-level automation possible. Successive similar processing projects – the previously described HMV collection and the digitized recordings from the Music Gallery Fonds – facilitated a scaffolded and gradual interpretation of descriptive standards in the lead-up to tackling the Mariposa recordings. Such a sustainable approach required CTASC to invest a massive amount of time to learn how to use and adapt to each software's functionality, limitations, and languages, including GREL and regular expressions (regex).[57] This specific processing project required approximately 80 hours from an archivist and 240 hours from two placement students (120 hours each).[58] This investment of time and resources is estimated to have saved over 700 hours in processing the seven remaining accessions. Needless to say, achieving this time savings was not without its challenges, particularly when navigating the gaps in (and the flexi-

57 The introduction of ChatGPT and other large language model computing platforms has further reduced barriers to archivists' ability to embed such technologies into their descriptive processes. Instead of spending hours learning and experimenting with GREL, an archivist can now ask ChatGPT to generate and explain GREL and regex. For an example, see Sandra Fauconnier, "Playing a bit with ChatGPT to make my life easier – I think it can help (?) users generate regex, GREL …?" OpenRefine community forum, Open Refine, accessed August 8, 2023, http://forum.openrefine.org/t/playing-a-bit-with-chatgpt-to-make-my-life-easier-i-think-it-can-help-users-generate-regex-grel/369.

58 Combined, this is the equivalent of eight weeks of full-time work for one individual. This does not include the time spent processing the Home Made Visible project and the digitized recordings of the Music Gallery Fonds or working on other similar projects.

bility of) Canadian archival descriptive standards.[59]

Building a Template to Describe the Extent of Digital Records

Heather MacNeil has argued that strategies for digital description "should be built on the foundation of descriptive principles and practices that have already been established."[60] Unfortunately, the last revision of the Rules for Archival Description (RAD) occurred in 2008, and since then, the landscape of born-digital records and the understanding of such material has evolved, making compliance to the standard quite difficult. In the case of the MFF Fonds, the added layer of audiovisual description complicates the matter even further – notably, due to the many possibilities for sound recordings in the extent and medium fields.[61] While RAD prescribes a standardized template to formulate the physical description of sound recordings, it does not provide a comprehensive example for born-digital records. RAD's chapter on electronic records includes diverse examples, with file format, material designation, and file size in varying sequences – for example, "6 photographs (tiff)," "1 sound recording (mp3)," "5 GB of photographs," "2 CD-ROMS (textual records)," and "1 zip disk (96 MB)."[62] There also appears to be a preference for describing electronic records in aggregates, based on the analog media carrier of the digital files, whereas the template for extent to document sound recordings provides the possibility to describe

59 Other decisions – such as keeping the digital file titles (session titles) as the archival descriptive titles, to mimic the transcription of donor-written analog folder titles – were easy. These creator-based titles also obscure discrepancies between anticipated (scheduled) programming and actual programming, which was subject to the realities of technical difficulties, weather cancellations, and other last-minute changes such as Gordon Lightfoot's unscheduled performance in 2012. Fortunately, audio engineer Sy Potma often captured such changes in the donation transfer notes. For an example of an unscheduled performance by a renowned performer, see Mariposa Folk Foundation, "The Gordon Lightfoot Connection," Mariposa Folk Festival, accessed June 22, 2022, http://mariposafolk.com/our-stories/the-gordon-lightfoot-connection/.

60 Heather MacNeil, "Metadata Strategies and Archival Description: Comparing Apples to Oranges," *Archivaria* 39 (Spring 1995): 22–32, 30, https://archivaria.ca/index.php/archivaria/article/view/12065.

61 There are eight possible characteristics to document in the extent field. The majority of these pertain to physical characteristics: physical composition, playing speed, recording method, groove characteristics, number of configuration of tracks, number of sound channels, and equalization and/or noise reduction systems. See rule 8.5C1 in Bureau of Canadian Archivists: Planning Committee on Descriptive Standards, "Sound Recordings," in *Rules for Archival Description* (Ottawa: Bureau of Canadian Archivists, 2008), 8-14, https://archivescanada.ca/wp-content/uploads/2022/08/RADComplete_July2008.pdf.

62 Bureau of Canadian Archivists, rules 9.5B1–9.5B3 in "Records in Electronic Form," *Rules for Archival Description*, 9-11 and 9-12.

at the item level. Simply put, the electronic records chapter provided minimal guidance on describing electronic audiovisual records at the item level for this case study. Complicating this further, the chapter on sound recordings, with a multitude of potential elements to describe the extent of audiovisual records, could be seen as overwhelming.[63]

Based on the incompatibility of the guidelines in these two chapters, CTASC then decided to build a standardized extent for the MFF recordings by selecting the technical metadata that was necessary and useful to researchers.[64] For analog audio recordings, this involved documenting the recording speed, which can assist in preparing machines for proper playback, including for digitization efforts, and can help in estimating the duration of the audio, which could in turn impact digitization costs as vendors typically bill at an hourly rate and/or based on the composition of the physical container. In the digital environment, information about file format and file size can help researchers and archivists determine whether they have the technological capacity to review files as well as whether the quality of a recording suits their needs. In deciding what to include in the MFF's extent, the CTASC archivist spent countless hours reviewing, comparing, and reflecting on the two format-specific chapters in RAD: sound recordings and moving images.

At the end of this process, they incorporated the following elements and order into the MFF extent template: # specific material designation (duration) : recording speed [sampling rate and bit depth], recording method [file format], recording technique [audio file format], number of channels ; and dimensions

63 As Richard Dancy highlights, "One of the good things about RAD is its wealth of categories for physical description at the item level for non-textual media, such as sound recordings, moving images, and so on. But in *RAD*, these are very much rooted in a world of analog objects. What does physical description mean in a digital world? . . . It is time to take a step back and rethink physical description in the digital age and to take a fresh look at RAD's so-called media chapters in that light." Richard Dancy, "Developing Archival Standards," *Archivaria* 78 (Fall 2014): 171–74, 172, https://archivaria.ca/index.php/archivaria/article/view/13503.

64 As Margaret Hedstrom writes, "The types of information needed to describe electronic records will differ from, and may exceed, that needed to describe records in paper formats, but the basic purpose of description remains much the same. Electronic records must have sufficient descriptive information to permit a user to learn that the record exists, identify and locate it, and determine the conditions under which it may be used. Once the record is located, the user must have sufficient descriptive information to write a command or instruct a computer to access the record or retrieve information from it." Margaret Hedstrom, "Descriptive Practices for Electronic Records: Deciding What Is Essential and Imagining What Is Possible," *Archivaria* 36 (Fall 1993): 53–63, 55, https://archivaria.ca/index.php/archivaria/article/view/11934.

[file size].[65] This formulation was based predominantly on RAD's extent for analog sound recordings, with a few modifications. Departures included the application of RAD 8.5B1 and 9.5B1, which refer to the creation of a specific material designation (SMD) if none of the suggested terms are deemed appropriate.[66] Rather than describing the temporary physical carrier that facilitated the donation (e.g., a USB drive), as suggested in the chapter on electronic records, CTASC described digital files at the item level using "1 audio file" for the SMD. Descriptions of temporary physical carriers are often not meaningful in the context of born-digital files – where the carriers serve simply as vehicles for donation, similar to the large plastic bins in which other donations make their way to institutional repositories – so it may not be useful to capture this information. Also, files are nowadays typically migrated off physical media for digital preservation, so even if they originally resided on that medium, that is not how users will likely be accessing them.[67] Historically, the Mariposa recordings were always transferred off the USBs to the department's shared drive, as the donor wanted the large storage USBs returned and YUL's digital preservation system did not exist at the time of the initial donation in 2009.[68] The decision to replace

65 For example, 1 audio file (42 min., 12 sec.) : 44.1 kHz, 16 bits, WAV, PCM, 2 channels ; 0.45 GB.

66 RAD 9.5B1 advises readers to "use terms found in subrule .5B in the relevant chapter of part I. If none of these terms is appropriate, use conventional terminology. If appropriate, add a qualifying word or phrase in parentheses to describe the type of computer or data file." Bureau of Canadian Archivists, "Records in Electronic Form," in *Rules for Archival Description*, 9-11.

The footnote for 8.5B1 also states, "This list of specific material designations is based on the physical carrier of the unit being described, i.e., the material or object which contains, supports or presents the sound recording. Institutions are encouraged to use the recommended list. However, if an institution chooses not to use this list it should, nevertheless, establish a policy regarding specific material designations." Bureau of Canadian Archivists, "Sound Recordings," in *Rules for Archival Description*, 8-12.

67 That said, one might want to include the USB drive in the extent if there are resource constraints that limit an archivist's ability to review the material or if the content cannot be migrated off the physical carrier. If there is time to review the material and there is a desire to describe the USB drive as a single unit, it might be helpful to create a list of physical extents similar to those used for series-level description. One could then add a restriction-on-access note mentioning that a USB drive is needed to access the material. If the born-digital record is described at the item level, the original arrangement on the USB drive could be captured through an arrangement note by including the original file directory path. This is similar to the case of the HMV project, where the majority of donations consisted of digitized home movies transferred as digital files without physical media carriers. However, as the media carrier could indicate the socioeconomic status of the donor, among many other things, including a reason for the quality of the recording, this context was preserved as a general note when possible. On the other hand, information about the transferring vehicle – a hard drive – was not added to the finding aid.

68 Each accession donated through the USB was given its own file directory folder, thus representing and

the USB with "1 audio file" – rather than the file format – as the SMD aligns with RAD's guidelines for describing moving images, which use video cassette as the SMD and place the recording mode or process (e.g., VHS, U-matic, or MiniDV) in the "other physical details" section of the extent. It also helps less technologically inclined users, who may not immediately recognize that terms such as MP3 or WAV refer to digital audio formats.[69] Another modification based on the moving-image chapter pertains to both the file format (analogous to recording method in the sound recording chapter or mode in the moving-image chapter) and the audio format (recording technique) as time-based media files are wrappers containing multiple files.[70] Recording speed was replaced by the sample rate and bit depth, as those parameters impact the quality of digital sound recordings. The number of channels was retained, as these are the same in both analog and digital formats. Finally, the file size was interpreted as the digital file's dimensions.[71]

Preconfiguring this approach to a standardized practice helped address the gaps in RAD. Ultimately, it helped facilitate the construction of automated scripts for the extent by breaking the descriptive elements into manageable chunks. In other words, CTASC needed to finalize the structure of the extent before determining the order of operations for automating the physical description of each recording.[72] As the extent field did not require further archival inter-

preserving the distinctiveness of the donation and the vertical arrangement of "order on storage." See Jinfang Niu, "Original Order in the Digital World," *Archives and Manuscripts*, 43, no. 1 (2015): 61–72, 64, http://doi.org/10.1080/01576895.2014.958863.

69 This decision would later also positively impact the use of SMD to automate the general material designation.

70 As moving images typically contain both a video file and sound file within a single record, the extent should include a description of the audio recording method. See RAD 7.5C1; 7.5C6c: "For video formats and optical discs, give the recording mode for the physical unit being described"; and 7.5C7b. Recording technique: "Give the type of sound track (e.g., composite optical track, magnetic stripe, analog, etc.)." Bureau of Canadian Archivists, *Rules for Archival Description*, 7-18, 7-19, 7-22.

71 In HMV, file size was placed within parentheses as it was initially interpreted as the equivalent to "film length." This decision has since been reversed to provide consistency with future descriptions of all digital formats by placing file size in the dimensions field.

72 Metadata was extracted from the digital recording using MediaInfo. It was then selected and rearranged by software and templated formulas. In Excel, this involved using both the CONCAT and TEXTJOIN functions to string cells and text together and the XLOOKUP function to search the MediaInfo spreadsheet and pull the metadata into the AtoM import archival description spreadsheet (or to search for an array and return the corresponding item). In OpenRefine, this process involved a combination of cell transformations using the following GREL functions: cell cross, replace, to number, to string, and multiplication.

vention following this intellectual process, it is considered the most automated phase of this case study.

Determining the Need to Duplicate Data in Scope and Content Notes

The next processing phase involved automating scope notes by repeating the information from the name access points. There were two reasons for this. The first involves RAD rule 21.0B, which states, "All access points must be apparent from the archival description to which they relate."[73] This could be interpreted to mean that name access points should not introduce new information but should rather provide structural support for information that already appears in the description (e.g., title, scope and content note, general note, etc.). Unlike the extent metadata, which was fully automated through technical metadata captured by MediaInfo, the scope and content descriptive metadata often have to be manually generated. In this case, students copied and pasted information from the festival program schedules as structured data into several columns, including performers as name access points. Then, another script was formulated to build the scope note, which read, "Item is an audio recording of the [session column] session performed at the [stage column] stage on [date column] and may include performance(s) by [name access points column]."[74] This approach created very structured scope notes compared to the narrative, free-flowing text traditionally associated with this field. One challenge with this rigid and regimented structure is that it did not account for additional information that could, and should, appear in the scope notes – like for example, an acknowledgement contextualizing Gordon Lightfoot's unscheduled appearance in 2012. In future, archivists might want to include an additional column for unstructured free text, to be appended to the end of the structured scope note. Another opportunity could involve working with the donor to embed artists' names in digital files.[75]

73 See Bureau of Canadian Archivists, "Choice of Access Points," in *Rules for Archival Description*, 21-3.

74 OpenRefine transformed the cell with the following expression: ("Item is an audio recording of the " + cells ['session'].value + " session performed at the " + cells['stage'].value + " stage on " + cells['eventDates'].value + " and may include performance(s) by " + cells['nameAccessPoints'].value + ."").replace("Item is an audio recording of the session,""Item is an audio recording").replace(" and may include performance(s) by .,"."").replace ("|,"," ").replace(/,([^,]*)$/, ," and $1").replace(" ,"" ").

75 Yeo suggests that, in a digital environment, records creators can share the responsibility of description by adding metadata to digital records at the moment of creation, thus lightening the load for archivists. See Yeo, "Continuing Debates about Description," 183.

Archivists could then extract this information from the born-digital files, thus saving time and reducing possible errors resulting from the process of copying and pasting performer names from the festival programs.

Simon Fraser University's AtoM data entry guidelines, perhaps in recognition of constraints on resources required to develop scope notes, offer an exception with alternative phrasing. They suggest using language such as, "For the list of correspondents, see Access Points."[76] However, if a researcher uses the search bar in AtoM, they will see only a snippet of relevant descriptive records in the user preview: the title, identifier, level of description, date, and scope note. So, if this alternative phrasing were used to circumvent the necessity to include name access points in the description, the researcher would not immediately see the connection to their search terms. This is because the donor named each recording after the relevant performance session, and such sessions could contain multiple performers; renaming the items' descriptive titles with long lists of performers could be problematic.

One could also decide to use only scope notes to avoid having to create name access points and their associated name authorities. Doing so would entail missing out on an opportunity to satisfy user expectations and to embrace the advantages of digital finding aids for exploring archival records through non-hierarchical methods. Creating name access points allows one to automatically draft scope notes with the names of the performers and avoid the duplication of labour involved in manually creating both fields. Zhang and Mauney found that, in response to changing user expectations, archives are moving away from traditional hierarchical finding aids, often offering more than one order and multiple access points (such as through categorization by place, name, or subject).[77] Research suggests that linked data and embedded digital objects in finding aids, by making hierarchical archival structures more obvious, can help users intuitively navigate archival descriptions online.[78] Yeo suggests that such granular description is even more necessary in the digital realm to harness the power of

76 Simon Fraser University, *GDL-39: SFU AtoM: Data Entry Guidelines*, version 4 ([Burnaby, BC]: SFU University Secretariat, Archives and Records Management Department, 2015), 24, http://www.sfu.ca/archives2/PDFs/GDL039.pdf.

77 Jane Zhang and Dayne Mauney, "When Archival Description Meets Digital Object Metadata: A Typological Study of Digital Archival Representation," *American Archivist* 76, no. 1 (2013): 174–95, http://doi.org/10.17723/aarc.76.1.121u85342062w155.

78 Berg, "A 'Major Technological Challenge.'"

relational systems (such as that in AtoM).[79] In this case, there is a shift toward perusing records in horizontal order through name access points and away from RAD's prescription of the top-down, vertically arranged, multi-level description represented in the finding aid.

Developing Frameworks for Horizontal Discovery through Name Access Points

Not only do name access points permit linkages that enable researchers to discover recordings of artists who performed at the festival over multiple years, but they also help intellectually reunite the festival's audio recordings by linking performances from the 1970s, preserved on audio reels, to those from the 21st century, saved as born-digital files. They also highlight material related to MFF artists across fonds preserved at CTASC. While one could anticipate finding related material of notable performers such as Buffy Sainte Marie, Gordon Lightfoot, and Jann Arden in other holdings, the name access points surfaced networks of relationships among various fonds. The most evident were connections between the MFF Fonds and the fonds of television producer and host of the weekly music series *Highway 10*, Jonathan Lynn, and television personality and broadcaster Dini Petty. Examples include records of Al Simmons, Valdy, and Skydiggers, who made the rounds on Lynn and Petty's television shows. Linkages were also occasionally formed with related music holdings, such as the fonds of the musician-run venue Music Gallery, which includes a recording of a performance of Ken Whiteley, and that of Women's Music Archives collector Ruth Dworin, which includes a recording of Nancy White.

It should be noted that the insertion of name access points into the archival description automatically creates blank name authorities records upon ingest to AtoM. Without further intervention, these stub authority records contained only the authorized forms of the names, thus providing another opportunity for automation – or rather, an opportunity to leverage the power of reconciliation and linked data. After creating a list of the name access points in the archival descriptions, CTASC developed a new spreadsheet to populate the fields related to name authorities. Matching the name authorities to those in the VIAF, Wikidata, and Social Networks and Archival Contexts (SNAC) made it possible to pull dates of existence and authority record identifiers from these sources. Drafting the content for the history field (i.e., the administrative history or biographical sketch)

79 Yeo, "Continuing Debates about Description," 181.

required the most intervention due to its narrative nature. To save time, this information was copied from Wikipedia under its Creative Commons Attribution-ShareAlike 3.0 Unported License and attributed by citing the URL. CTASC hoped that the robust descriptions, with the additional contextual information of name authorities, would ultimately optimize the discoverability of archival records.

Investing time in creating name access points for item-level description also made it possible to surface the records related to BIPOC artists. As Olivia Wong states, "We don't know what's in our holdings because we don't have item-level descriptions," typically due to limited staff time and resources;[80] however, the iterative and non-final nature of finding aids provides opportunities for change as our collective relationships with archives evolve. The MFF is known for its inclusive representation of ethnocultural and Indigenous performers in its early years, and item-level description helps highlight the histories of these performers.[81] Between 1964 and 1967, the festival increasingly featured African American and Canadian blues performers as well as artists from First Nations across Canada. By 1972, MFF had formed an ethnic committee devoted to promoting the inclusion of musical artists and dance groups from Arabic, Turkish, Hungarian, Lebanese, Chinese, Mexican, Macedonian, and Portuguese communities.[82] The 1970s also saw the creation of the festival's "Native Peoples' area," which featured Cree storytellers, Inuit throat singers, artists from the Tłı̨chǫ Nation, and contemporary and traditional singer-songwriters and poets.[83] It should be noted that a high percentage of researchers accessing or requesting recordings from the MFF Fonds are focused on studying Indigenous artists. Berry notes that traditional archival concepts such as provenance, original order, and collective description can often obscure the histories of BIPOC communities by upholding the colonial foundations of institutional memory keeping.[84] She states,

80 Melissa J. Nelson, "13 | Olivia Wong, Toronto Metropolitan University Libraries," episode 13, in *Archives & Things* (podcast), April 19, 2023, http://melissajnelson.com/podcast/.

81 For example, in 1973, BIPOC performers represented nearly a quarter of artists in the MFF lineup. See Sija Tsai, "Public Policy and the Mariposa Folk Festival: Shared Ideals in the 1960s and 1970s," *MUSICultures* 38, no. 1 (2013): 152, http://journals.lib.unb.ca/index.php/MC/article/view/20207.

82 Tsai, "Public Policy and the Mariposa Folk Festival," 152.

83 Tsai, "Public Policy and the Mariposa Folk Festival," 152; Sija Tsai, "Mariposa Folk Festival: The Sounds, Sights, and Costs of a Fifty-Year Road Trip" (PhD diss., York University, 2013), 178, http://hdl.handle.net/10315/31317.

84 Dorothy Berry, "The House Archives Built," *Up//Root* (blog), June 22, 2021, http://www.uproot.space/features/the-house-archives-built.

> By loosening that structure in a digital setting with results organized from the researcher's query, as opposed to the received "original order," entirely new frameworks of receiving Black history are created. . . . By providing description at a different level through a different venue, targeted hidden collections projects can create artificial collections that highlight the experiences of specific marginalized groups, while not disturbing the traditional archival organization of the physical materials themselves.[85]

Cassin's experience researching master Atikamekw canoe builder and traditional artisan Cesar Newashish illustrates the potential of the MFF Fonds to share the stories of Indigenous artists.[86] Reconceptualizing the exploration of the MFF Fonds with the addition of name access points for MFF artists and facets for browsing using name authorities in AtoM has produced an online infrastructure that facilitates alternative access to and exploration of the festival's performance recordings.[87]

Conclusion

This case study illustrates the possibility Yeo once expressed: that "automated capture of descriptive information can be expected to offer the most promising economic solutions."[88] These tools, however, work at the item level and are "are unlikely to supply all the information – particularly broader contextual information – that archival institutions and users may require, but it seems almost certain that they will eventually perform the more mundane aspects of description and representation at the item level, thus allowing archivists to focus

85 Dorothy Berry, "Hide and Seek: Organizing Hidden Collections for *Umbra Search African American History*," *Los Angeles Archivist Collective* (blog), accessed June 22, 2023, http://www.laacollective.org/work/hide-and-seek-organizing-hidden-collections-for-umbra-search-african-american-history.

86 Stacy Allison-Cassin, "Research Libraries and Wikimedia: A Shared Commitment to Diversity, Open Knowledge, and Community Participation," Wikimedia Foundation, October 4, 2017, http://wikimediafoundation.org/news/2017/10/04/libraries-wikipedia-york-university-project/.

87 Artefactual Systems, "Add Access Points to Your Authority Record," AtoM, accessed June 22, 2022, http://www.accesstomemory.org/en/docs/2.7/user-manual/add-edit-content/authority-records/#authority-access-points.

88 Yeo, "Continuing Debates about Description," 183.

resources on aspects that require human understanding."[89] Despite the efficiencies promised in automating the processing activities, it should also be noted that these efficiencies do not necessarily resolve the problem of backlogs nor the underlying systemic issues that cause them. In their discussion of archival debt, which they say "amalgamates the myriad issues we now grapple with as a profession, including harmful or inadequate description, performative or competitive collecting, languishing backlogs, failure to recognize staff potential, shortsighted fund management, neglected constituencies, a lack of documentation, and poor project management," Cuellar et al. caution against attempting to pay off our debt with efficient technologies, noting that this "will not fix structural issues that are bigger than the departments we work in, the repositories we work in, and quite obviously bigger than our profession."[90]

The MFF Fonds is just one example in the evolving landscape of archival description in the digital era, where technologies such as online finding aids, keyword searching, and metadata scripting contribute to demands for increased access and robust descriptive practices. It also emphasizes how software can influence and impact the ability to create such descriptions.[91] That the success of this project relied heavily on interpretation of the RAD underscores the importance of descriptive standards to navigating intellectual challenges associated with ensuring standardized descriptions for a variety of formats. Also highlighted are the broader implications for archival practice, as technology has significantly influenced descriptive practices over the years. It has done so by facilitating increased access to archival descriptions through online finding aids and keyword searching as well as by permitting the scripting of metadata attached to born-digital records and creating robust archival descriptions to meet researcher expectations.

89 Yeo, "Continuing Debates about Description," 183.

90 Jillian Cuellar, Audra Eagle Yun, Jennifer Meehan, and Jessica Tai, "Defining Archival Debt: Building New Futures for Archives," *Journal of Contemporary Archival Studies* 10 (2023): 11.

91 Since the initial donation of the MFF born-digital recordings in 2007, new software applications have been invented, such as OpenRefine, used to manipulate data, in 2010, and Wikidata, used to extract crowdsourced linked open data, in 2012. Updates and upgrades to these software applications – which affect the automation of description, including the ability to bulk upload more than 100 descriptive records through AtoM's graphic user interface – first appeared in 2017, and Excel's XLOOKUP function to search arrays was introduced in 2021.

After completing the initial pilot project described in this article, CTASC described the remaining born-digital performances using this workflow. The homogenous nature of these donations – performance recordings using the same file format and requiring a similar descriptive pattern (i.e., scope note, name access point, and extent) – provided an entry point to building a semi-automated descriptive workflow. Recently, by working with smaller acquisitions, it has adapted the workflow to address the needs of a smaller, heterogeneous born-digital records acquisition consisting of textual, video, and audio files from the Egypt Migrations Collection. While CTASC's approach to born-digital records will likely evolve with each successive processing project, the department will likely still depend on the crucial guidance and interpretation of descriptive standards. It also depends on the continuing conversation within the profession to advance our collective capacity to process born-digital records. As Bak once stated, "Digital preservation in Canada has not been, is not now, and never will be the story of any one project or institution."[92]

BIOGRAPHY Katrina Cohen-Palacios is an archivist at the York University Libraries Clara Thomas Archives and Special Collections. She holds a Master of Information and a Master of Museum Studies from the University of Toronto's Faculty of Information and a Bachelor of Arts with a specialization in public history from Concordia University.

BIOGRAPHY Sarah Lake is the digital preservation librarian at Concordia University Library. Prior to her current position, she served as an archivist at Concordia University Records Management and Archives and in the library's special collections. She holds a Master of Information Studies from McGill University and a Bachelor of Fine Arts in film production from Concordia University.

92 Bak, "How Soon Is Now?," 308.