Original Order in Digital Archives

JANE ZHANG

RÉSUMÉ Cet article fait état des résultats d'une étude de deux ans qui visait à établir la validité du principe de l'ordre original lorsque des documents électroniques sont transférés des systèmes de gestion de l'information et traités comme archives numériques. En se servant d'un cadre de connaissances établi à partir des écrits en archivistique, l'auteure a mené des études de cas de trois projets d'archives numériques afin d'explorer les définitions, les buts et les représentations de l'ordre original tels qu'elle les a perçus dans les archives numériques. Les résultats de cette étude montrent une variété de structures et de liens entre les documents créés électroniquement, puis identifiés et conservés en tant que tel dans les archives numériques. L'étude met en évidence les possibilités et les défis auxquels font face les archivistes en ce qui concerne la représentation de l'archivage numérique.

ABSTRACT This article reports on the findings of a two-year study designed to investigate the validity of the principle of original order when electronic records are transferred from recordkeeping systems and processed into digital archives. Guided by a knowledge framework identified from archival literature, the author conducted case studies of three digital archival projects to explore the definitions, purposes, and representations of original order witnessed in digital archives. The findings of the study demonstrate a variety of structures and relationships of records created electronically and identified and preserved as such in digital archives, and highlight the possibilities and challenges archivists face in digital archival representation.

Introduction

The principle of original order, originating from the nineteenth-century European paper-based records tradition, has been a guiding principle for archivists in the arrangement and description of archival records for more than a century. Modern information technologies have changed the way records are created, stored, and used. Whether the principle of original order can be adapted and applied to digital environments has been an open question in the archival community for the past two decades. The question of the adaptability of the principle of original order to the arrangement and description of elec-

ARCHIVARIA 74 (Fall 2012): 167–193

tronic records was raised in the early 1990s.[1] In the first decade of the twenty-first century, digitized and born-digital collections have grown dramatically, giving rise to increased interest in the application of archival principles to organizing digital collections. Within the archival community, archivists are thinking and experimenting. They believe that archival context and original order should be preserved as an important and valued component of archival practice but wonder how this might happen in digital archives.[2]

This article summarizes the findings of a two-year study on the concept of original order as applied to digital archives.[3] A comprehensive review of the literature on the principle of original order and its application identifies three major areas that researchers and practitioners generally touch upon when discussing original order in the twentieth century: its definitions, purposes,

1 For example, Fredric Miller observed in 1990 that original order preserves the physical arrangement of individual paper documents in file units, but whether the principle of original order "can be adapted to automated systems is very much an open question" (Fredric M. Miller, *Arranging and Describing Archives and Manuscripts* (Chicago, 1990), 27). Charles Dollar commented in 1992 that "[t]raditional arrangement and descriptive methods that have focused upon physical order cannot deal adequately with the changing form of documents generated by emerging information technologies" (Charles M. Dollar, *Archival Theory and Information Technologies: The Impact of Information Technologies on Archival Principles and Methods*, ed. Oddo Bucci (Ancona, Italy, 1992), 77). Heather MacNeil stated in 1994 that "original order has tended, … to be associated with physical arrangement. That association is no longer valid for most electronic records." (Heather MacNeil, "Archival Theory and Practice: Between Two Paradigms," *Archivaria* 37 (Spring 1994): 10).

2 For example, at the 2007 Society of American Archivists (SAA) Annual Meeting, several presenters discussed the problems uncovered in understanding the context when analog archival materials are digitized. They argued that "analog" techniques in the communication of context and original order can be improved to express archival context in a digital environment (Program Session 804, "Preserving Context and Original Order in a Digital World"). Case studies were also reported at the 2008 SAA Annual Meeting in which archivists discussed how they tried to apply traditional archival theories when working with electronic records and other born-digital materials (Program Session 203, "Getting Our Hands Dirty (and Liking It): Case Studies in Archiving Digital Manuscripts"). Published papers presented at these two sessions and other related publications include Abigail R. Griner, "Where's the Context? Enhancing Access to Digital Archives," *Provenance* XXVI (2008): 59–69; Michael Forstrom, "Managing Electronic Records in Manuscript Collections: A Case Study from the Beinecke Rare Book and Manuscript Library," *American Archivist* 72, no. 2 (Fall/Winter 2009): 460–77; Jeanne Kramer-Smyth, "SAA 2007 Session Proposal: Preserving Context and Original Order in a Digital World," *Spellbound Blog*, 28 September 2006, and "SAA2008: Preservation and Experimentation with Analog/Digital Hybrid Literary Collections (Session 203)," *Spellbound Blog*, 6 September 2008, http://www.spellboundblog.com (accessed 29 October 2010).

3 Jane Zhang, "The Principle of Original Order and the Organization and Representation of Digital Archives" (PhD diss., Simmons College, Boston, 2010). The author would like to express her sincere thanks to Simmons College, Wellcome Library, Library and Archives Canada, and the Persistent Digital Archives & Library System (PeDALS) for their academic, financial, and/or professional support for the completion of her dissertation research.

and representations.[4] The general knowledge framework of original order – how it is defined, what it is used for, and how it is represented in an archival system – provides an underlying logic that orients the study. The investigation addresses the following research questions in the three areas of original order:

- *Definitions*: Original order as an important aspect of records seems to be identifiable and definable in the paper records environment. Would that be the case with electronic records? What constitutes the "order" of records in digital archives? Can the traditional organization of records continue to exist in digital environments? How would the documentary relations of electronic records be identified and preserved?
- *Purposes*: Original order as an integrated part of the archival system serves some important archival functions, such as protecting context and facilitating access. Would that order continue to exist in digital archives to achieve the same purposes? Would there be any new "orders" of electronic records that have been transferred and preserved to serve new archival purposes in digital archives?
- *Representations*: A close relationship can be observed between original order and archival hierarchical representation in paper-based archives. Would that relationship also exist in digital archives? What role might the organization of electronic records play in digital archival representation? How should digital archives be described and represented to meet user needs?

Methodology

The research method used is the collective case study. Also called the multiple-case design, this method aims to gain greater insight into a research topic by concurrently studying multiple cases.[5] According to Robert Yin, the strength of a multiple-case design lies in its replication logic: when each case is carefully selected and examined and all cases generate similar results, the study provides more compelling and robust support for the initial propositions than a single case study.[6]

The focus of the study is on archival thinking and practice relating to digital archival organization and representation. For that purpose, the researcher selected digital archival collections that were originally generated and organized by records creators and later transferred to an archival institution and processed by archivists based on archival principles and research needs.

4 Ibid., 11–43.

5 Burke Johnson and Larry Christensen, *Educational Research: Quantitative, Qualitative, and Mixed Approaches*, 3rd ed. (Thousand Oaks, CA, 2008), 408.

6 Robert K. Yin, *Case Study Research: Design and Methods*, 4th ed. (Thousand Oaks, CA, 2009), 54.

The researcher sought examples of digital archival collections that had gone through a more or less formal, complete archival process: acquisition, accession, organization, and representation. The selection targeted digital archival projects that were well planned and sufficiently documented with a reasonable scope in terms of project duration, collection size, and stability of repository. To ensure the relevance of study cases, the focus of the selection was on institutions that have an emphatic mandate for collecting and preserving born-digital archival materials. Three digital archival projects were selected for this study:

- The Digital Curation[7] project at the Wellcome Library for the History and Understanding of Medicine, London, a digital initiative launched in 2005 to acquire and collect born-digital archival materials for its collections.[8]
- The Library and Archives Canada Trusted Digital Repository (LAC TDR), a three-year grant project (2007–2010) funded by the Treasury Board of Canada Secretariat to preserve electronic publications and government records.[9]
- The Persistent Digital Archives & Library System (PeDALS), a grant-funded multi-state research project led by Arizona State Library, Archives, and Public Records in partnership with state archives and libraries in Florida, New York, South Carolina, Wisconsin, New Mexico, and Alabama. Launched in 2007, the project focuses on capturing, preserving, and providing access to state and local government digital information.[10]

The Wellcome Library, as indicated in the will of its namesake benefactor Sir Henry Wellcome (1853–1936), holds "[c]ollections of books, manuscripts, archives, films and pictures on the history of medicine from the earliest times to the present day."[11] Its Archives and Manuscripts collection specializes in "medical and health-related archive material from antiquity to the present."[12] Since 1979, the Wellcome Library has been committed to the collection and preservation of contemporary materials:

7 For an introduction to digital curation, see Digital Curation Center, "What is Digital Curation," http://www.dcc.ac.uk/digital-curation/what-digital-curation (accessed 29 October 2010); Elizabeth Yakel, "Digital Curation," *OCLC Systems & Services* 23, no. 4 (2007): 335–40.

8 Wellcome Library, "Digital Curation at the Wellcome Library," http://library.wellcome.ac.uk/node288.html (accessed 29 October 2010).

9 Library and Archives Canada, "LAC Trusted Digital Repository," http://www.collectionscanada.gc.ca/digital-initiatives/012018-4000.01-e.html (accessed 29 October 2010), [2].

10 Persistent Digital Archives & Library System (PeDALS), "PeDALS Project Brochure," http://pedalspreservation.org/Resources/pedalsbrochure.pdf (accessed 29 October 2010).

11 Wellcome Library, http://library.wellcome.ac.uk/ (accessed 29 October 2010).

12 Wellcome Library, "Online Catalogue of Archives and Manuscripts," http://library.wellcome.ac.uk/node49.html (accessed 29 October 2010).

In 1979 the Wellcome Trust took the decision to extend the library's existing collecting policies more pro-actively into the acquisition of twentieth-century materials, by setting up the Contemporary Medical Archives Centre (now subsumed into Archives and Manuscripts). This had the aim of collecting and preserving records illuminating twentieth-century developments in medicine, biomedical science and health care.[13]

Compatible with its commitment to contemporary archival collections is the Wellcome Library's new pledge "to maintain its position as a major international research library in an **increasingly digital environment**" (emphasis as in the original), as highlighted in its *Library Strategy 2006–2009*.[14] The strategic plan is derived from an understanding that many of the library's donors today produce materials only in digital form. "These files, emails, datasets and images have no paper equivalent, they are 'born digital'. If our collections are to grow and remain relevant we must find ways to acquire, manage and make this material available."[15]

The Wellcome Library started to accept born-digital material in the 1990s but was short of suitably skilled staff to manage born-digital archives.[16] The first digital accession was received in November 2004.[17] In 2005, the Library launched the Digital Curation in Action Project to examine how to take born-digital material into its collections.[18] With the official acknowledgement in *Library Strategy 2006–2009* of the importance of digital material in the institution's future, the Wellcome Library appointed its first digital curator[19]

13 Adrian Steel and Lesley A. Hall, "Sir Henry Wellcome's Archival Legacy and the Contemporary Historian," *Contemporary British History* 17, no. 3 (August 2003): 103.

14 Wellcome Library, "Library Strategy 2006–2009," http://library.wellcome.ac.uk/assets/wtx034924.pdf (accessed 29 October 2010), 4.

15 Wellcome Library, "Digital Curation at the Wellcome Library."

16 Jenny Haynes and Dave Thompson, "Accession to Access: Born Digital Archives in the Wellcome Library," Future Proof V – International Scientific Archives Conference, Barcelona, 6–8 May 2009, http://files.me.com/fxroque/8swmac (accessed 29 October 2010), 16.

17 Natalie Walters, "Going Digital: The Case of the Wellcome Library," Digital Preservation Coalition Briefing Day – Digital Preservation Planning: Principles, Examples & the Future with Planets, British Library, July 2008, http://www.planets-project.eu/docs/presentations/natalie_walters.pdf (accessed 29 October 2010), 3.

18 Dave Thompson, "Going Digital, Experiences at the Wellcome Library," 29th Annual IATUL Conference, Auckland University of Technology, New Zealand, 21–24 April 2008, http://www.iatul.org/doclibrary/public/Conf_Proceedings/2008/DaveThompson.pdf (accessed 29 October 2010), 5.

19 According to Hilton and Thompson, the digital curator at the Wellcome Library works closely with archivists and provides technical support for them. "The processes for acquiring and managing digital material are being built on sound archival practice, and driven by the archivists, supported by one new appointment, a digital curator to provide technical support." Christopher Hilton and Dave Thompson, "Further Experiences in Collecting Born Digital Archives at the Wellcome Library," *Adriadne* 53 (October 2007), http://www.ariadne.ac.uk/issue53/hilton-thompson/ (accessed 29 October 2010), [1].

in 2006 and implemented Fedora as a test/experimental repository.[20] Digital objects have been ingested into Fedora in their native format. Using Fedora as a test repository has introduced to the library the basic concepts and processes of the acquisition and storage of born-digital material.[21] Building on the experience with Fedora, the library began to define requirements for a digital repository in 2007[22] and has recently selected the Tessella SDB system for the long-term preservation and management of its born-digital records.[23] The idea of "Accession to Access" has been developed as an end-to-end workflow for the next five years, to move and embed digital curation into everyday work and make dealing with digital material business as usual.[24]

The Library and Archives Canada Trusted Digital Repository (LAC TDR) is a three-year grant project that was launched in 2007 to build a trusted digital repository to preserve its digital collections. The LAC TDR is a new digital archival initiative launched following the merger of Canada's National Library and National Archives to form Library and Archives Canada in 2004. The 2004 Library and Archives Canada Act mandates LAC to preserve the documentary heritage of Canada for the benefit of present and future generations.[25] The legislature requires LAC to address its digital preservation capacity, including legal deposit of electronic publications, harvest of the Internet for web content of interest to Canadians, and transfer of federal government digital records of archival value. In 2005, LAC decided to adopt the concept of trusted digital repositories (TDR)[26] to build its own Trusted Digital Repository (LAC TDR) and embarked on a multi-year development project funded by the Treasury Board to establish a reliable and integrated digital preservation infrastructure through which the nation's digital documentary heritage could be identified, collected, managed, preserved, and accessed.[27]

20 Dave Thompson, "Fedora at the Wellcome Library, Progress and Work to Date," DPC Digital Repositories Meeting, 5 July 2006, http://www.dpconline.org/docs/events/06briefdigrepthompson.pdf (accessed 29 October 2010), 4.

21 Chris Hilton and Dave Thompson, "Collecting Born Digital Archives at the Wellcome Library," *Ariadne* 50 (January 2007), http://www.ariadne.ac.uk/issue50/hilton-thompson/ (accessed 29 October 2010), [2].

22 Hilton and Thompson, "Further Experiences in Collecting Born Digital Archives at the Wellcome Library," [1].

23 Tessella, "Press Release: Tessella Wins Wellcome Library Digital Repository Contract," 24 March 2009, http://www.tessella.com/2009/03/press-release-tessella-wins-wellcome-library-digital-repository-contract/ (accessed 29 October 2010).

24 Haynes and Thompson, "Accession to Access," 17.

25 Department of Justice Canada, "Library and Archives Canada Act," http://laws.justice.gc.ca/en/L-7.7/ (accessed 29 October 2010).

26 Research Libraries Group, "Trusted Digital Repositories: Attributes and Responsibilities: An RLG-OCLC Report," May 2002, http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf (accessed 29 October 2010).

27 Pam Armstrong, "Becoming a Trusted Digital Repository, Library and Archives Canada,"

The LAC TDR has been constructed on the basis of collaboration between LAC and government departments. To ensure the basic standardization of electronic records management systems, LAC worked with government departments and organizations to establish a records management metadata standard and application profile.[28] In 2008, LAC launched the Core Archival Metadata Project to identify the core metadata set for transferring records from departmental electronic records management systems to the TDR.[29] The LAC TDR was initially built to preserve electronic publications and government electronic records. It is based on the Open Archival Information System (OAIS) Reference Model,[30] and when fully developed will ensure "a set of trusted services that provide reliable and persistent access to" as well as "reliable storage and long-term preservation of the digital collections at LAC."[31]

The Persistent Digital Archives & Library System (PeDALS) is a multistate research project with principal funding from the Library of Congress National Digital Information and Infrastructure Preservation Program (NDIIPP). The project, led by Arizona State Library, Archives, and Public Records, started in early 2007 with five state archives and libraries partners (Arizona, Florida, New York, South Carolina, and Wisconsin) and expanded in early 2009 to include two new state partners (New Mexico and Alabama). PeDALS is a research project that aims to achieve four major goals: to describe the curatorial processes of acquisition, arrangement and description, storage, access, and preservation as business rules; to build a secure and inexpensive storage network to protect the authenticity and integrity of the documents; to build a community of best practice to promote resource sharing and to avoid costs of redundant work; and to remove barriers to preservation by keeping costs as low as possible.[32]

The technical goals of the project are to develop a curatorial rationale that supports an automated, integrated workflow for processing collections of digital publications and records; and to implement "digital stacks" using an inexpensive storage network that can preserve the authenticity and integrity of

Ontario Information Management Conference, Toronto, Ontario, 28 April 2009, http://www.verney.ca/opsim2009/presentations/783.pdf (accessed 29 October 2010), 7.

28 Library and Archives Canada, "Government of Canada Records Management Metadata Standard," "Government of Canada Records Management Application Profile," http://www.collectionscanada.gc.ca/government/products-services/007002-5000-e.html (accessed 29 October 2010).

29 Library and Archives Canada, Core Archival Metadata Project Team, "Core Archival Metadata Project – Final Report (DRAFT)," Internal Report, Draft 0.3, 17 April 2009, 6.

30 Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS),* January 2002, http://public.ccsds.org/publications/archive/650x0b1.pdf (accessed 29 October 2010).

31 Library and Archives Canada, "LAC Trusted Digital Repository."

32 Persistent Digital Archives & Library System (PeDALS), "PeDALS Project Brochure," http://pedalspreservation.org/Resources/pedalsbrochure.pdf (accessed 29 October 2010), 2.

the collections.[33] The project uses Microsoft BizTalk middleware[34] to automate the flow of digital information as it moves from government offices to secure archival storage, and uses LOCKSS software to maintain copies of the documents in separate physical locations and to provide automatic integrity and error checking.[35]

Based on the belief that traditional archival curatorial processes will not scale to the ever-increasing volume of government records and publications, PeDALS has developed as its curatorial rationale the principle that project librarians/archivists work with rules, not with records. The PeDALS system is based on the Open Archival Information System (OAIS) Reference Model and contains eight high-level activities: negotiating a schema for a submission information package (SIP) with each office of origin; writing rules to validate the records; writing rules to create a standard archival information package (AIP); writing rules for accessioning and describing the records; writing rules to ingest the AIP into the LOCKSS system; writing rules to create a dissemination information package (DIP) from the AIP; writing rules to publish DIPs to the Web; and monitoring to make sure the processes are running correctly, making adjustments if necessary.[36]

The researcher collected data from the three selected study cases by examining project documentation and conducting onsite interviews. Project documentation provided written evidence about the organizational structure of the digital archives. Onsite interviews collected oral and observational evidence from archival or digital curation professionals regarding their experiences and reflections responding to the research questions in this study – their conscious or subconscious efforts to make use of original structure of records in the organization and representation of digital archival materials. The interviews were conducted at the sites of the three digital archival projects in November and December 2009.

The interviews were semi-structured and included detailed discussions with project archivists and/or digital curators at each site. The discussions focused on the three research areas of the study to understand the organizational structures of digital records before they are transferred to the archives, the purposes the organizational structures serve while and after digital records are transferred, and the processes and systems used to represent the organizational structures of records in digital archives. A total of eight hours

33 Persistent Digital Archives & Library System (PeDALS), "About PeDALS," http://www.pedalspreservation.org/ (accessed 29 October 2010).

34 Microsoft BizTalk Server, http://www.microsoft.com/biztalk/en/us/ (accessed 29 October 2010).

35 Persistent Digital Archives & Library System (PeDALS), "PeDALS Project Brochure."

36 Persistent Digital Archives & Library System (PeDALS), "PeDALS Curatorial Rationale," http://pedalspreservation.org/About/rationale.aspx (accessed 29 October 2010).

of interviews were recorded and transcribed. Interview transcripts were coded and processed using a system designed by the researcher, following a three-step qualitative data analysis process recommended by Miles and Huberman.[37] Although this is qualitative research and the coding system was designed and implemented by the researcher, the coding procedures were closely followed to generate findings that could be replicated by the same process.

Digital Archives from Creation to Access

The construction of digital archives starts, as expected, with the creation of records in digital form in the creating/keeping system, followed by the transfer of records of enduring value and their processing, preservation, and description in the archival management system, and continues until records are made available for use in the archival public access system. A generic process from creation to access is generally understood, but specific practices in each step may vary from case to case. Understanding how records in digital form are created, transferred, processed, preserved, described, and accessed is helpful for the identification and interpretation of original order in digital archives.

The Wellcome Case

The Digital Curation Project at the Wellcome Library works with individual donors as well as non-profit, volunteer-based societies or charity centres to collect born-digital materials. Individuals and records creators at small non-profit organizations use personal computers to create, manage, and organize their digital materials, mostly word-processed documents, spreadsheets, presentations, e-mails, and digital photographs. They rely on electronic file directory and electronic mail folder systems to retain electronic files or digital objects in order. Individuals make personal, non-uniform decisions about file organization and format. Sometimes people voluntarily or involuntarily resort to letting the operating systems of their personal computers organize their electronic documents for them. In cases where digital collections have a disorganized internal structure, the archivist relies on the knowledge about the creator or creating agency and the understanding of transferred material to re-establish an internal structure for the records.

The appraisal process of personal digital papers stored in the computer involves a close review of the folder structure and title description of electronic files. In the transfer process, selected materials are copied from a

37 Matthew B. Miles and A.M. Huberman, *Qualitative Data Analysis: An Expanded Sourcebook*, 2nd ed. (Thousand Oaks, CA, 1994), 57–72.

donor's hard drive onto portable carrier media, such as CDs, DVDs, or floppy disks. The library prefers to use its own portable USB hard drives and assigns validation keys to ensure that the material arrives intact. Files transferred are downloaded to a designated drive in the Windows system after checking for viruses. The processing drive is a secure, restricted area accessible only to the project staff. The drive serves as a temporary storage space and a designated processing area for the archivist to do arrangement work. The arrangement work involves the identification and verification of documentary relationships of digital records. After processing, digital records are deposited in the digital preservation system for long-term storage.

The Wellcome Library uses Tessella's digital archiving and preservation solution, Safety Deposit Box (SDB), which was developed in partnership with The National Archives (UK) for long-term storage of its digital materials.[38] The core technology of SDB has been designed to help memory institutions such as libraries and archives confront the challenge of preserving material created and stored in digital formats. The system incorporates a unique approach to the preservation of digital objects, and the key feature of its active preservation technology is its tight control of the migration process.[39] Migration is the conversion of digital objects from one technology to another and is "a reliable way to preserve the intellectual content of simple digital objects (e.g., page-based documents)."[40] To facilitate file conversion, preservation and accessibility, multi-page files are unpacked and saved to a single new file with its associated metadata, thus adding one more level of document relations to the hierarchical control of records.

The description of digital collections takes place in CALM, the system adopted by the library to manage its archival collections.[41] CALM is used by archivists to create and store descriptive data, and by users to search and request archival material online. An interface is then created to link the archival description in CALM with the digital records stored in SDB. The archivist first creates an accession record and a multi-level description, normally providing the provenance and series information. While the collection- and series-level description is created by the archivist, the file- and

38 Tessella Technology & Consulting, "Safety Deposit Box," http://www.digital-preservation.com/solution/safety-deposit-box/customer-list/ (accessed 29 October 2010).

39 Tessella Technology & Consulting, "The National Archives: Digital Archiving," http://www.digital-preservation.com/wp-content/uploads/TNA.pdf (accessed 29 October 2010).

40 Ross Harvey, *Digital Curation: A How-to-Do-It Manual* (New York, 2010), 166.

41 According to Malcolm Howitt, Managing Director of Axiell CALM Limited, CALM stands for "Collection management for Archives, Libraries and Museums" (Malcolm Howitt, 26 April 2012, e-mail message to author). CALM is listed as a UK-based archives management system in Lisa Spiro, "Archival Management Software: A Report for the Council on Library and Information Resources," January 2009, 58. http://www.clir.org/pubs/reports/spiro/spiro2009.html/spiro/spiro_Jan13.pdf (accessed 27 April 2012).

item-level description relies on what has been supplied by the creator. The granularity of file- and item-level description thus varies depending on how individuals organize their files and what metadata are supplied by the creator or system.

The library has a catalogue system that provides access to bibliographic information about books, journals, and other library material. It also has archives and manuscripts collections, image collections, and film collections. Each of the collections (books and journals, archives and manuscripts, images) has a separate catalogue and search function. The Archives and Manuscripts catalogue provides online access to archival and manuscript collection descriptions. The descriptions can be accessed by keyword search of any text, featured search (reference number, title, data, name, language, and subject), as well as description-level search. The Archives and Manuscripts catalogue search function is limited by the levels of description provided for archival collections. In the paper records scenario, if users want to see a particular item, they click on an icon to generate a request to see the physical item. With digital material, they may click on the request button to gain immediate access to the material.

The process of handling digital archival collections at the Wellcome Library is modelled on the processes applied to paper archives. The workflow follows the same steps of transfer, accession, arrangement, description, storage, and access. By integrating digital records into the existing archival system, the library aims to build a hybrid system for paper and digital records. In other words, the same finding aid created for physical records will continue to be used for paper, born-digital, and digitized records that belong to that collection. The rationale behind the hybrid system is the belief that digital materials should not be described separately from paper records if they are in the same collection and, at the moment, traditional finding aids are still a useful tool for incorporating digital material into an existing collection. The ultimate goal of the institution, therefore, is to create hybrid collections – paper, born-digital, and digitized records from the same creating source that are all described in an integrated finding aid.

When working with digital material, the library follows two principles: the first is the archival principle of respecting provenance and original order; the second is to ensure long-term accessibility of the material. The library cannot accept long-term responsibility for digital material that it cannot work with. It makes changes if necessary, because there is no point in keeping material that is going to become inaccessible in a few years. In the Wellcome case, the change made to turn multi-page files to a single new file has pushed the control of original order down to the item level. The relationships among individual digital items have thus been made visible and significant.

The LAC TDR Case

The LAC TDR was initially built to ingest and preserve electronic publications and government electronic records. At this stage, government electronic records are defined as records derived from the Records, Document and Information Management System (RDIMS) currently in use by Canadian federal government agencies.[42] The RDIMS system involves four levels of organization. The first is the system itself. Within the system, the highest level of aggregation is called a "library." The library is the system-designated term for arbitrary divisions to segregate groups of records that have their own functional structures. Within the library, there are files and items. At the heart of the system's organization is the file classification code – the key metadata that maintain the internal relationship of records in the system, especially at the file and item levels.

RDIMS was designed without the capacity to identify and transfer electronic records along with their associated metadata. To overcome this shortcoming, LAC has developed a plug-in called eRecord Transfer Application (eRTA) for RDIMS to enable the transfer of electronic archival government records. The first step toward implementing the LAC TDR was the development of the Virtual Loading Dock (VLD) as an electronic record transfer tool to receive and unpack digital assets (that is, electronic records and their associated metadata). Digital assets are temporarily stored in the VLD until further appraisal before transfer to the TDR. The functionality of the tool includes search and selection of required records to be transferred, transfer of the metadata application profile, selection review, and record packaging.

The RDIMS system allows for the management of records according to disposition, thus enabling automation of records selection for transfer by the government department. This happens when the system is programmed to specify that each record selected is associated with a file classification code that is designated as archival under the disposition program. The role of the archivist is to check and confirm that no gross errors have been made in the classification code match that would result in the wrong records being transferred. Unless there are some compelling reasons for someone to go in and

42 The System (RDIMS) was selected in the mid-1990s to be used by Canadian government departments for the management of records, electronic documents, and information in response to the changing electronic records environment resulting from office automation. According to Robert Coffin, RDIMS is a vendor-developed suite of integrated software applications. The Treasury Board of Canada Secretariat purchased licences for all Government of Canada departments to implement when they were ready. Robert Coffin, "Information Management and e-Government in Canada" (PowerPoint presentation retrieved online, accessed 29 October 2010).

manipulate metadata, the transfer of electronic records to the LAC TDR is simply a mapping of the creator's metadata to the TDR system.

The LAC TDR takes only file- and item-level metadata, as identified in the archival core metadata set. This corresponds with the situation in RDIMS wherein the metadata created above the file and item levels are not consistent enough to be captured as a meaningful archival entity that can be detached from the file classification code. This is because there is no consistency in practice when departments create a "library" in the RDIMS system. A library may contain a series or a collection, or it can be used as a regional division. A large government organization may have separate libraries for separate regions. Although prefixes and suffixes can be attached to file codes as subdivisions to indicate library-level aggregates, there is no consistency as far as how prefixes and suffixes are constructed and what they represent. The lack of consistency makes it very difficult to automate the series-level representation in the digital archival system. It remains an open question how the series aggregates of records in the RDIMS system are understood so as to represent them in the archival description.

Public access in the case of the LAC TDR takes place in the MIKAN system.[43] MIKAN is a discovery tool that was developed to access physical archival records. It is an archival information system and, at the same time, part of the federated search mechanisms of the LAC collections. When used to display and access digital archival material, MIKAN takes descriptive metadata from two sources. The lower-level descriptive data, though not fully represented in the public access system, are linked from the TDR to MIKAN. The higher-level description data (i.e., fonds and series levels), on the other hand, are not associated with the TDR. They are manually created by the archivists in the MIKAN system in accordance with the archival description standard. The higher-level representation documented in the MIKAN system is a standard archival hierarchical description. The structure provides a context that ties individual files and items to the purposes for which they were created and to the persons responsible for their creation. It is this context that makes what the archivists collect and preserve archival. The finding aids thus generated are in compliance with the description standard at the higher-level aggregation.

The construction of a digital archival system like the LAC TDR is mainly driven by digital preservation needs and requirements. The TDR was not created to address the need for intellectual management in digital archives; it was created to deal with digital preservation challenges of individual objects. In

43 MIKAN "is not an acronym as such, but is apparently based on an Algonquin word meaning 'road,' 'path,' or 'discovery.'" Reference Services, Information and Research Services Division, Library and Archives Canada, 24 April 2012, e-mail message to author in response to online reference request.

other words, digital preservation precedes all other digital archival problems. The archival management of paper records mostly occurs at aggregate levels – appraisal, transfer, description, preservation, access. Electronic records, on the other hand, require a different approach. Digital preservation management must occur at the item level because different object types require different treatment. The TDR implementations to date have been driven by the library community. The *Metadata Encoding and Transmission Standard (METS)* and the *Metadata Object Description Schema (MODS)* have been adopted in order to structure received metadata within the TDR. *METS* and *MODS* are used as carriers of the metadata in order to build common grounds for digital preservation management of bibliographic and archival descriptive metadata. In the LAC TDR case, there has been a lot of effort to establish archival processes in the system, figure out how to move archival metadata to *MODS*, and try to make them work for managing digital archival records.

The automatic mapping of file- and item-level metadata to the MIKAN system has introduced something new to the archival description system. In the TDR digital preservation environment, electronic records are regarded as digital objects that require individual treatment. Although archivists in this project do not describe records below the series level, the automatic mapping of metadata has resulted in some kind of automatic description at the lower level. This is because, in the course of electronic records creation and use, descriptive metadata are accumulated at the file and item levels. Since these metadata are highly structured in the electronic records management system, it is possible in the future to repurpose the metadata for archival uses, especially for archival description and access.

The PeDALS Case

The PeDALS system is designed to process electronic records that are generated from similar business processes and are associated with existing metadata used by records creators to manage and access their records. The assumption is that the records creator has established an order or an access system so that records stored can be retrieved. After the records are transferred to the archives, archivists can exploit the received metadata instead of spending time recreating them. The system that has thus been built requires appropriate records sets that are large enough and consistent enough to allow business rules to function. The model is not immediately appropriate for processing less-structured records. The ideal scenario would be to store and index records in an electronic document management or recordkeeping system. Certain pilot records series generated during routine business by state government agencies might be suitable for transfer to the PeDALS system; for example, marriage certificates, e-mail records, death certificates, and litigation case files.

At the heart of the PeDALS system is the creation of business rules to extract metadata received from records creators and convert them into predefined core metadata that can be used for discovery, administration, and preservation purposes. To facilitate metadata mapping, the PeDALS project has established between thirty and forty core metadata elements that are considered common to all state government records. There may be more metadata than can be used in the future archival system. The PeDALS project maps only the metadata elements that are required in the core metadata and are needed to support the functions of archival discovery, administration, and preservation.

Mapping is a process of transforming and normalizing the received metadata based on the core. There are four classes of requirements for the core metadata: mandatory, preferred, desirable, and optional. The mandatory class dictates that the system will not function unless the mandatory element is present. For example, a title is a mandatory field. If a record is exported to the PeDALS system without a title, the system will stop working and will not be able to process the record. Therefore, the process of mapping records that do not have a title will involve the creation of a business rule to extract metadata from existing records and construct a descriptive title to be mapped to the core metadata.

The first step of metadata mapping is to negotiate with the office of origin to ascertain the existing metadata, the format and medium of submission records, the data structure, and the transfer cycle. It is the responsibility of the office of origin to prepare and create submission packages and transfer them to the PeDALS system. Each state has project archivists who are responsible for metadata mapping and preparing their records for processing. The PeDALS system receives both records and their associated metadata from the office of origin. The project hires a software programmer to work with archivists to translate the metadata requirements into machine language so that records can be processed automatically.

PeDALS uses BizTalk as the middleware technology to receive records and associated metadata from agency records systems, apply predefined business rules to each records series, and distribute them to the long-term storage system and the public access web server. The BizTalk middleware controls the processing of records as they move through the systems.[44] For long-term storage, PeDALS uses the LOCKSS technology to build its secure "digital stacks." LOCKSS, an acronym for Lots of Copies Keep Stuff Safe, is an international community initiative based at Stanford University Libraries. The LOCKSS technology provides libraries with digital preservation tools and technical support so that they can easily and inexpensively collect and preserve their own copies of authorized electronic material.[45]

44 Microsoft BizTalk Server, "Overview," http://www.microsoft.com/biztalk/en/us/overview.aspx (accessed 29 October 2010).

45 LOCKSS, "Home," http://www.lockss.org/lockss/Home (accessed 29 October 2010).

The general framework of the PeDALS system corresponds to the OAIS reference model. Records and associated metadata in agency record systems are transferred in the submission information package (SIP). The SIP is processed through the middleware (BizTalk), where the curatorial rationale has been translated into business rules to normalize received metadata. Received metadata, normalized metadata, and records are then packaged into the archival information package (AIP) and deposited in the LOCKSS system for secure storage. Records stored in the LOCKSS system have a lot of metadata that the public may not be interested in, so when the middleware generates and sends an AIP to storage, it also creates the dissemination information package (DIP), which formats records in a way that makes it easy for the public to view them through a web browser. The middleware also populates the administrative catalogue – an internal SQL database containing descriptive information about each record, with the core metadata elements that support required discovery and administrative functions.

The PeDALS system processes record items automatically by writing business rules to integrate original item-level metadata (e.g., names, dates, document forms) into single item titles. However, in order to place record items in context, the project archivists need to capture sufficient information for describing the provenance and series in the administrative catalogue. Some aggregate information not submitted by the office can be inferred from the transferred material. The archivists need to enter information specific to the accession that cannot be automated. They will spend some time generating series-level metadata, creating the scope note and the administrative history, and assigning some high-level subject headings to the series. The higher-level processing is a manual process for human beings to enter the aggregate-level metadata and description.

Contrary to their involvement in aggregate-level metadata, most of which are entered by archivists or cataloguers, human beings are never involved in generating item-level metadata in the PeDALS project. It is an automated process, which the project believes to be very important because it is transforming the way that electronic records archivists work in comparison with the ways paper records archivists work. Instead of working with individual records, archivists work with business rules expressed in computer software. The only time the archivists handle individual records is at the initial stage when they analyze sample records to write business rules. After that, they run the rules and make sure the rules are working appropriately. It is largely an automated process, with quality control procedures.

The series is at the heart of what archivists do in the PeDALS system because it is the highest level of an aggregate of records that are related and have context and meaning. It is mostly at the series level that the organization and description of records take place. Unlike a traditional provenance-based collection, series serve as the main entry in the PeDALS system. The search

interface of PeDALS maintains the traditional archival multi-level fields of provenance, series, files, and items. The design emphasizes the importance of hierarchical representation of archival material; however, the project does not present its collections in the format of "browsable" archival finding aids because it is assumed that most researchers are interested in non-hierarchical search access. The main search strategy in the PeDALS system is metadata-based keyword search across all predefined representation fields or within any fields specified by users.

Original Order in Digital Archives

The previous section describes three digital archival cases in which distinctive organizational structures are adopted by records creators to organize and access electronic materials. All three digital archival systems appear to rely on the original organizational structures of records to function, and they manage to maintain a certain level of the logical internal relationships among electronic records. However, a consistent notion of original order in digital archives does not emerge because of the differences between the three electronic recordkeeping scenarios.

When the organization of records is controlled by individuals, as shown in the Wellcome case, personal recordkeeping behaviours and the mechanisms of the information system play a critical role in determining how electronic records are created, organized, and stored. In a controlled electronic records management environment, as shown in the LAC TDR case, the organization of records is likely to be governed by regulated practices and by predefined file classification systems. In a routine business process-based, metadata-centric records environment, as shown in the PeDALS case, the organization of records can be realized by means of metadata generated to support a common business purpose and shared by multiple records creators.

How can we describe, characterize, and represent original order in digital archives in these three scenarios? How do we understand and interpret the preservation of original order as a valid concept as well as a legitimate practice in digital archives? The summary of the findings of the study attempts to address these questions. The key findings are highlighted in brief summary statements followed by discussion.

Key Finding 1

In spite of a general assertion that records, if created and stored digitally, are interconnected and share multiple logical relationships, some electronic records that find their way into the archives are tied to a single file structure and are organized in linear-like fashion, in ways familiar in the paper environment.

Archival researchers generally agree that the feature of physicality in traditional paper filing systems is missing in electronic recordkeeping environments. Heather MacNeil believes that the association of original order with the physical arrangement of records is no longer valid for most electronic records.[46] David Bearman points out that the original order physically imposed on paper filing is "neither necessary nor desirable for electronic records"[47] because automated electronic records systems "may involve many types of relationships."[48] In this vein, Terry Eastwood criticizes the old-fashioned "one-thing-one-entry approach" and argues that archivists should not "describe physical groupings of records in a fixed context and only in that context."[49] Similarly, Terry Cook comments that, in the digital environment, original order changes from maintaining the physical placement of records to "the conceptual intervention of software," and "orders reflect multiple uses in work processes rather than physical arrangement of recorded objects."[50] Geoffrey Yeo reinforces the point in a recent article:

> In digital environments, where juxtaposition of records is insignificant, even the limited contextual clues provided by physical ordering are largely missing, and we must reinterpret the principle of original order in terms of identifying multiple logical relationships among records, rather than in terms of their physical groupings.[51]

In spite of the general assertion that electronic records by default are interconnected, not physically grouped, sample records series examined in this study reveal that electronic records generated in a personal recordkeeping environment are mostly organized by applying traditional methods of physical groupings and are maintained in a single file structure. For many individuals or small organizations, personal computers with file folder directories are all they have to keep their electronic files. They are forced to create and maintain an order – an order that may not differ much from that in paper files because, structurally speaking, computer file folder directories work very much like electronic file cabinets. As file folders can be organized hierarchically, and within each level, individual files are maintained in a single file structure, less likely to be provided by the record creator than imposed by the software

46 MacNeil, "Archival Theory and Practice: Between Two Paradigms," 10.
47 David Bearman, "Item Level Control and Electronic Recordkeeping," *Archives and Museum Informatics* 10 (1996): 196.
48 David Bearman, "Record-Keeping Systems," *Archivaria* 36 (Autumn 1993): 18.
49 Terry Eastwood, "Putting the Parts of the Whole Together: Systematic Arrangement of Archives," *Archivaria* 50 (2000): 115–16.
50 Terry Cook, "Archival Science and Postmodernism: New Formulations for Old Concepts," *Archival Science* 1 (2001): 21–22.
51 Geoffrey Yeo, "Debates about Description," in *Currents of Archival Thinking*, ed. Terry Eastwood and Heather MacNeil (Santa Barbara, CA, 2010), 92.

system. Even in large organizations like government agencies, shared drives are commonly used to keep electronic records in file folder structures which, to a large degree, mirror the physical groupings of paper filing systems.

Electronic records retained on personal computers or shared drives are filed into folder structures, usually not guided by a classification scheme, which makes it harder to apply records disposition plans and identify archival records. Things are different in an electronic records management system, in which electronic records can be classified to facilitate systematic record disposition and transfer. By assigning classification codes to records, as shown in this study, a structure may have been imposed, because one file can only be assigned one code in order to make it unique. In other words, for the system to perform some essential records management functions, a file structure needs to be created to fix the relationship, so as to assert management and control.

To some extent, the function/activity-based hierarchical file structure developed in traditional paper filing systems has continued to be used by private and public records creators in the management of electronic records. Organizing records in a file structure based on activity categories and time sequence has been practised both in the file directories manually created in personal computers and shared drives and in the file classification schemes implemented in electronic records management systems. Similar to its counterpart in paper records, a single file structure may not be sufficient to capture the complexity of records creation and use. Upon transfer, records organized in one fixed order offer little flexibility when it comes to accessing them in digital archives.

Key Finding 2

Records and associated metadata generated in a metadata-centric recordkeeping environment are very likely to have multiple representations of original order because more complex relationships of electronic records are maintained in the system.

In a metadata-centric recordkeeping environment, such as the one described in the PeDALS case, records normally have no single fixed order and can be organized and manipulated in various ways. For example, records retained in a digital imaging or electronic document management system can be retrieved by different search criteria in accordance with its metadata schema. In an electronic records management system where records are organized according to a classification plan, records can also be associated with other metadata elements to make them more searchable. An e-mail management system is another example of a metadata-centric recordkeeping system in which e-mail messages can be sorted by a variety of attributes to display them in various orders. In these systems, everything is controlled by multi-faceted metadata: the number of orders that records can have depends on how sophisticated their supporting metadata schema is, and the order in which search results are

displayed can be predefined by the system or specified by a query formula. At the time of archival transfer, associated metadata can be mapped into the archival system together with records.

What is unique in a metadata-centric recordkeeping environment is that more complex relationships of electronic records are maintained in the system and, as a result, records and associated metadata generated from the system are very likely to have multiple representations of original order. Creator-supplied multi-field metadata elements support archival discovery, administration, and preservation. This is a new digital phenomenon that is different from physical recordkeeping and may create new opportunities for archivists to fulfill their long-cherished dream of representing archival material in multiple orders, all the while preserving records in their original dynamic context.

In circumstances where a specific order cannot be identified and records can be reorganized in more than one way, the concept of original order may refer to the context in which records are created, structured, accessed, and utilized. Because records are not arranged in a particular order, archivists may succeed in preserving original order to the extent of preserving the organizational context of records. Within the context, records are controlled by logical relationships expressed in the metadata schema and can be sorted, searched, and queried. To some extent, the concept can be extended to other circumstances in which the order of records may not be that important (e.g., documents in a paper file), or may not be meaningful to human eyes (e.g., data sets), or may be too chaotic to identify (e.g., desktop files). If nothing else, archivists should try to preserve the context of any order, or lack of order, so that users may be able to understand the records within that context.

Key Finding 3

Original order in the form of file structure and record metadata plays an important role in digital archival appraisal, acquisition, and processing. Creator-generated record metadata automatically mapped into the digital archival system makes automated archival processing possible at the item and/or file level.

As a series of organized activities to turn records into archives, archival appraisal, acquisition, and processing can be viewed as a continuum of recordkeeping actions imposed on a set of records transferred from the records management system to the archives system. This study indicates that original order in the form of file structure or record metadata plays an important role in performing these archival activities in digital archives. In the case of digital records acquired through a formal records scheduling and disposition plan built into an electronic records management system, archivists rely on the original file classification plan to decide what records to select for

mapping into the digital preservation system for permanent retention. The reasons for the acquisition of archival records are therefore expressed in original order – that is, by following the original file classification plan. In the case of records that do not have a central control system for scheduled disposition, archival appraisal and records disposition may take place at the series level or even higher, at the provenance level.

Archival transfer of digital material can be done manually (by copying electronic files onto transfer media) or automatically (by mapping of electronic records and associated metadata from system to system), as shown in this study. Archivists either follow the file structure or rely on the original metadata of digital material to complete the transfer in an effective and resource-conservative way. Digital archival processing involves the confirmation and validation of received metadata to be used to assist with file or item retrieval. Because of the increasing volume of digital records, archivists will not be able to get down to the lower-level description if they do not process records in an automatic way. It is no exaggeration to say that digital acquisition and processing are mainly about management and manipulation of received file structure or record metadata to ensure appropriate archival functions, including preserving and representing documentary relations of digital records.

Key Finding 4

Original order expressed in archival hierarchical representation continues to be an important tool to gain access to archival records as evidence of original creation and use. The introduction of item-level metadata in digital archives opens the door to direct access to information in digital archival records.

The main representation and access systems of the three cases are all based on the traditional hierarchical model following the order of provenance, series, file, and item. The Wellcome and LAC TDR cases have incorporated digital archival access into their existing archival catalogue/description systems. The PeDALS case has designed its own access system, which is a little light in descriptive detail owing to its automated process, but, like the other two cases, undoubtedly hierarchical.

All the archivists interviewed recognize the value of original order as a context-based access tool. There has been a lot of processing effort in the Wellcome case to ensure digital files are arranged in a logical order intended by records creators but disorganized by computer systems. Electronic records in the LAC TDR case are arranged in the order established in the original departmental file classification system. This case supports the notion that, to access records as evidence of original creation and use, the best order is the one that meets the archival principles of provenance and original order. In the PeDALS case, original order is considered a handy way to access records. To facilitate automatic processing, the digital archival system in this case works

under the assumption that an organizational structure has been established in the original records system by the office of origin to access and use its records. It is assumed that the structure will continue to be useful and effective and that mapping its underlying metadata to the archival system for reuse will save archivists a lot of time and resources.

However, as important as contextual access of archival records should be, archivists in this study have also pointed out that people may approach records for a variety of reasons. As they emphatically explain, not everyone comes to use archival records with a rigorous accountability or historical research agenda in mind. Some people may just want to find a picture of an old car from a collection, whereas others may need to conduct research on the automobile industry. A digital archival access system built on the traditional file structure model does not enable more than one order through which to access records. It is an indirect access system in which information can only be identified and discovered by means of knowing who the information creator is and where the information might possibly reside in the files. A digital archival system has the potential to enhance information discovery in archival records by enabling access options that the indirect access system of the traditional file structure model might not provide.

The manipulation of item-level metadata may hold potential in terms of opening the door to direct access to archival records. Original order is traditionally associated with paper filing systems. As an archivist involved in this study commented, access to digital records cannot be exactly the same as in paper records systems because automatic mapping of file- and item-level metadata has introduced new representation elements to the archival system. Received metadata can be highly structured and can be repurposed to achieve archival objectives, including direct access to information in digital archival records.

Key Finding 5

Digital preservation management cannot take place at the aggregate levels because digital objects require item-level control. The item-centric methodology in digital preservation has contributed to the shift of the archival expression of original order from the file level down to the item level.

The results of this study show that one of the main driving forces behind the use of item-level metadata to enhance information discovery is the archival community's push for solutions to long-term preservation of digital material. All three cases in this study were initiated to address the pressing concern of format obsolescence or secure storage of digital archival material. Each project has either purchased or built a digital storage and preservation system. The storage and preservation system has been created to deal with the challenges of digital preservation that precede all other digital archival

problems, including the intellectual management of archival material. As one archivist involved in the study commented, people have to look for solutions to preservation management problems first, and all other problems come after and are, therefore, secondary.

The impact of digital preservation on the organization of digital material is significant. In the Wellcome case, the digital curator worked very hard to turn the multi-page files into single-page files to reduce them to a manageable level so that the library could assume long-term preservation management responsibilities. In the LAC TDR case, digital preservation metadata, along with other discovery and administrative metadata, are identified and linked at the item level, including the file classification code that maintains the internal relationships of records in the system. The PeDALS project uses a metadata extractor to automatically capture the metadata required for digital preservation, such as file format and file size. Some metadata used for preservation purposes are embedded in digital items, which are stored in the system as flat files, along with structural metadata used to declare hierarchical relationships in the representation system.

All three institutions studied use digital preservation systems for managing digital library materials – Tessella SDB, LAC TDR, and LOCKSS. In these systems, digital archival material is retained as itemized objects to meet preservation needs. There is apparent consensus in all three cases that digital preservation management cannot take place at the aggregate level – it must occur at the item level. As a result, item-level information management, while exceptional in traditional archival practice, may have become the norm in digital archives. Whereas archivists normally trace the original order of paper records down to the file level, they are forced to take an item-centric approach in dealing with digital and electronic records because digital objects require item-level control. For preservation purposes, digital material has to be stored discretely with its exclusive metadata; consequently, the archival control of original order has been pushed down to the item level in response to this item-centric methodology in digital preservation.

The archival control of item-level metadata may transform what archivists do and how they do their work. Item-level metadata is crucial for automating the digital archival transfer processing. The practice of automatic mapping of electronic records and associated metadata into digital archival systems is seen in two of the three cases studied. In all three digital systems, creator-supplied descriptive metadata at the file and item levels are retained as part of digital archival description. Archivists/digital curators interviewed in this study have all articulated the necessity of keeping digital material identifiable and traceable at the item level, so as to perform basic migration and preservation functions. Item-level metadata also hold great potential to provide more granular access to archival material, and enable archivists to manage access to material for confidentiality or privacy protection purposes. Because the metadata that

accompany an electronic record are an integrated part of the record itself, metadata may also be used to document the authenticity and integrity of the record.

Key Finding 6

There are two levels of representation in digital archives. Higher-level description (provenance and series levels) is supplied manually by archivists. Lower-level metadata (item and file levels) generated by records creators can be automatically mapped to the digital preservation system and linked to the archival representation system.

This study shows a pattern of a two-level representation system in digital archives. The higher-level description (series and up) is manually created by archivists, and the lower-level representation (file and item levels) is mainly based on the original file and item-level descriptive metadata supplied by records creators and automatically mapped into digital archival systems. While the higher-level description can be compliant with archival description standards, the lower-level representation is mostly dependent on how records are organized and what metadata are generated in the record creation and recordkeeping system. The top-level description is hierarchical and provides provenance and series-level context for digital material, whereas the lower-level metadata are mostly flat and provide for granular searching of record content.

The two levels of representation are handled differently, more for reasons of practicality than for theoretical assumptions. As shown in this study, aggregate information is generally not included in the automatic process of item-level metadata mapping. Although such information can be inferred from transferred material or gathered from background research, including working with donors or offices of origin, aggregate information is mainly the product of archivists' manual work. On the other hand, because of the extensive amount of digital material being generated every day and the huge amount of work that would be needed to process such material, it does not seem possible to describe digital archival material down to the item level if the work is not done automatically. The practice arising from the automation of archival processing is challenging traditional notions. As shown in the LAC TDR and PeDALS cases, instead of working with individual records, archivists may work with computer-programmed business rules to transfer and process item-level metadata.

Automatic mapping of file- and item-level metadata will lead to automatic description at the lower level in digital archives. As a result, less descriptive work will be done by archivists processing digital material at the lower level because they will be able to depend more on creator-supplied metadata than archivist-contributed description to provide access to digital archives. The

increasing volume of digital archival material will prevent archivists from undertaking manual processing while, at the same time, the availability of item-level metadata will provide more access to archival material than is possible in other archival contexts. The two-level representation system creates an interesting discovery scenario in digital archives. Users may rely on the higher-level description for contextual information and the lower-level representation for content search.

Conclusion

In 1994, Heather MacNeil described the shift in our understanding of the meaning of original order as the principle of original order has adjusted to the reality of electronic records:

> In paper-based record systems, where records are physically ordered in labelled files, usually in accordance with a classification scheme, the physical and contextual aspects of the records are intimately connected: original order has tended, for that reason, to be associated with physical arrangement. That association is no longer valid for most electronic records. In their case no such arrangement and indexing system exists. … What has changed here is not the principle: preserving original order has always meant preserving the records' documentary relations. What has changed is its application: for electronic records, applying the principle will increasingly mean ensuring the preservation of data directories, which is where those relations will be described.[52]

The findings of this study support MacNeil's statement that the principle of original order continues to be valid but the expression of documentary relationships of archival records has undergone transformation in the digital environment. Records transferred to digital archives can be originated by creators utilizing file directories, classification coding systems, and metadata schemas. While file directories can be structured in a variety of ways, the documentary relations of records are mainly controlled by personal recordkeeping behaviours and computer file organization systems. In electronic records management software environments, file classification codes can be designed and implemented to formalize documentary relationships among records, primarily based on business functions and activities and associated with records retention and disposition plans. In metadata-centric records creation and management environments, metadata schemas play an essential role in structuring how records are related and expressed. Unlike file directories and records classification coding systems, relationships among metadata-enriched records can be expressed in multiple ways, depending on business needs embedded in the design and implementation of metadata schemas.

52 MacNeil, "Archival Theory and Practice: Between Two Paradigms," 10.

This study shows that the organizational structures of records – whether they are file directories, classification systems, or metadata schemas – are what archivists rely on to identify, transfer, preserve, process, and make available for use records created in the digital environment. Original file structures, classification codes, and records metadata are inherited and preserved by archivists to protect context and facilitate access. In addition, they are used to keep track of individual digital objects for their technical features in order to achieve long-term preservation. Because records preserved in digital form need to be discretely identified and stored for digital preservation purposes, special attention has been paid to the identification and representation of records at the item level in digital archives. Not only are electronic files or digital objects individually named or assigned various metadata; relationships among records established in their creation and use can also be associated and expressed at the item level so that digital records can be linked to their appropriate groupings and broader context.

The item-centric approach in digital processing and preservation may pose some challenges for traditional archival hierarchical representation. While it makes sense to expect a more granular representation and retrieval system for digital records owing to pre-existing item-level metadata, the reality may proceed in an unexpected direction, such as the segregation of higher-level from lower-level representation found in this study. Digital archival objects, along with digital library and museum objects, are most likely stored in preservation systems as flat files with associated file- and item-level metadata to achieve an optimal preservation goal. When it comes to linking them out to representation systems for information discovery purposes, file- and item-level metadata may not be fully utilized, partially because of limitations in traditional archival representation. As shown in this study, some archives may routinely exclude lower-level description in their representation systems and, as a general practice, they describe archival materials only at the collection and series levels. Those trying to make use of item-level metadata may find it difficult to integrate multi-faceted metadata into an archival hierarchical representation system, which, as a norm, requires an all-encompassing descriptive title for an archival item, not a standard set of metadata elements.

In digital archival representation, the challenge for archivists is not only to represent digital archival items by their original metadata to maintain their metadata-empowered retrieval capacity, but also to preserve the context in which records are created and used. The preservation and representation of archival context makes it necessary for archivists to gather and evaluate information to generate higher-level description for the digital records identified and transferred to their collections. At the same time, as one of the cases in this study implies, in order to embrace the full representation of original order in a hybrid archival environment, it is important to explore the representation

of a broader archival context that takes into account both digital and analog records generated from the same origin and business purpose.

With a focus on preserving the context of record creation and use, the concept of original order adopted in the study goes beyond the narrow sense of sequencing of records traditionally identified in paper files. It is interpreted as a broad concept that identifies, characterizes, and represents the structures and relationships of records in an archival system. As Adrian Cunningham points out, in the digital world archivists may stick to "the real principle underlying the old principle of 'maintaining original order'" by "linking records to their business and social context and ensuring that such linkages persist over time."[53] This study presents examples to demonstrate a variety of structures, relationships, and linkages of records created digitally and preserved as such in digital archives.

53 Adrian Cunningham, "The Postcustodial Archive," in *The Future of Archives and Recordkeeping: A Reader*, ed. Jennie Hill (London, 2011), 185.