CASE STUDY

Managing Electronic Records in Manuscript Collections: A Case Study from the Beinecke Rare Book and Manuscript Library

Michael Forstrom

Abstract

This paper reports on the management of electronic records in manuscript collections at the Beinecke Rare Book and Manuscript Library at Yale University. The paper offers a case study exploring the InterPARES 1 Authenticity Task Force "Requirements for Assessing and Maintaining the Authenticity of Electronic Records" and archival description as models for assessing and maintaining the authenticity of copies of electronic records in manuscript collections. The paper focuses on rules in *Describing Archives: A Content Standard (DACS)* applied to description in finding aids.

Introduction

The Beinecke Rare Book and Manuscript Library is Yale University's principal repository for literary papers and early manuscripts and rare books in the fields of literature, theology, and history. In addition to its general collection of rare books and manuscripts, the library houses the Yale Collection of American Literature, the Yale Collection of German Literature, the Yale Collection of

© Michael Forstrom.

This is a revised version of a paper presented at the 2008 SAA Conference in San Francisco, California, in session 203, "Getting Your Hands Dirty (and Liking It): Case Studies in Archiving Digital Manuscripts." The paper also grew out of the Beinecke Library's implementation of *DACS* as a content standard for finding aids. I would like to thank Heather MacNeil and Yale colleagues Tom Hyry, Kevin Glick, Jennifer Meehan, and Mike Rush for comments on earlier drafts of the paper.

Western Americana, and the Early Modern and Osborn Collections. The Yale Collection of American Literature (YCAL) is well known for the strength of its archival holdings in early twentieth-century literary papers, but the Beinecke Library, and YCAL in particular, continues to acquire the archives of modern and contemporary writers and, increasingly, these archives include digital components.

Approximately fifty-five manuscript collections¹ now held by the library contain computer media.² Much of this media could be characterized as fugitive³ in the sense that there has been no significant precustodial intervention, the digital content has not been appraised prior to acquisition, and the media is part of a collection consisting chiefly of paper-based materials.⁴ While the library has implemented policies and practices for managing electronic records on media of this sort, both changes in technology, personal computer use, and professional standards and practice, and acquisitions of mixed or hybrid archives are forcing us to re-evaluate and revise policy and practice.

There is a growing body of professional literature, standards, and resources relating to the management of personal electronic records.⁵ Early contributions to the literature (from the mid- to late 1990s) address the value of personal electronic records, computer use habits, types of digital records and environments, and general strategies for managing electronic records.⁶ One gap in the early literature, from the perspective of processing archivists, is the absence of case studies. The one exception, marking a transition in 2000, is

¹ Many of these collections fall into what Lucie Pacquet identifies as the "first generation of users among the individuals who donate electronic records. The first generation are 50 years or older who began, late in their careers, to use the first personal computer." Lucie Pacquet, "Appraisal, Acquisition and Control of Personal Electronic Records: From Myth to Reality," *Archives and Manuscripts* 28, no. 2 (2000): 73.

² Media include common formats (5½, 3½, and zip disks; CDs and DVDs), contain chiefly word processing and text files, and can be found in personal papers and the records of literary presses and journals.

³ I am making a distinction between fugitive digital content in manuscript collections and mixed or hybrid archives. See *Workbook of Digital Private Papers*, 2005–2007, available at http://www.paradigm.ac.uk/workbook/appraisal/appraisal-issues.html, accessed 3 March 2009.

⁴ Adrian Cunningham argues for precustodial intervention in a series of articles: "The Archival Management of Personal Records in Electronic Form: Some Suggestions," *Archives and Manuscripts* 22 (May 1994): 94–105; "Waiting for the Ghost Train: Strategies for Managing Electronic Personal Records Before It Is Too Late," *Archival Issues* 24 (1999): 55–64; and "Digital Curation/Digital Archiving: A View from the National Archives of Australia," *American Archivist* 71 (Fall/Winter 2008): 530–43.

⁵ The Society of American Archivists now maintains an online bibliography to best practice guidelines and literature. See SAA Technology Best Practices Task Force, "Managing Electronic Records and Assets: A Working Bibliography. 6. Electronic 'Manuscript Collections'," available at http://www.archivists.org/saagroups/bptf/mera-emss.asp, accessed 22 June 2009.

⁶ In this paper I use the term *record* broadly to encompass documents and other digital objects that can find their way into an individual's archive. For discussion on concepts of *record*, see the excellent pair of recent articles by Geoffrey Yeo, "Concepts of Record (1): Evidence, Information, and Persistent Representation," *American Archivist* 70 (Fall/Winter 2007): 315–43 and "Concepts of Record (2): Prototype and Boundary Objects," *American Archivist* 71 (Spring/Summer 2008): 118–43.

Lucie Pacquet's article, "Appraisal, Acquisition and Control of Personal Electronic Records: From Myth to Reality," which summarizes practice at the National Archives of Canada from the early to late 1990s.⁷

Since Pacquet's article, in addition to an increase in the literature, several standards have been published and adopted and numerous research projects have been undertaken.⁸ The Paradigm *Workbook of Digital Private Papers*, endorsed by the Society of American Archivists' Technology Best Practices Task Force, covers archival functions, as well as metadata, repository, and legal issues, and provides links to tools and references.⁹

Two recent case studies describe how American and Australian colleagues are managing electronic records at the Harry Ransom Center (HRC) at the University of Texas at Austin¹⁰ and at the National Library of Australia¹¹ respectively. Stollar Peter's discussion of the Michael Joyce Papers at HRC, for example, addresses arrangement of digital records within hybrid collections, preservation strategies and use of DSpace, use of open-source tools and other utilities, and the challenges posed by electronic literature.¹²

This study seeks to add to the growing body of literature by addressing authenticity requirements for electronic records in manuscript collections. ¹³ The purpose of the study is to explore InterPARES requirements and archival description as

⁷ Pacquet, "Appraisal, Acquisition and Control of Personal Electronic Records," 71–91.

⁸ Relevant standards include *ISAD(G)*, *DACS*, *RADS*, *AGRkMS*, ISO 14721 (or OAIS), and *PREMIS*. Projects focusing on personal electronic records include: Paradigm, futureArch, and Digital Lives. For discussion of digital preservation projects, and an annotated bibliography, see InterPARES 3, "General Study: Past and Present Digital Preservation Project Devoted to the Long-Term Preservation of Digital Records and Digital Information," version 1.4 (January 2009), available at http://www.interpares.org/display_file.cfm?doc=ip3_gs01_digital_preservation_projects_v1-4p.pdf, accessed 27 February 2009.

⁹ Personal Archives Accessible in Digital Media (Paradigm) workbook, the result of a collaboration between staff at the Bodleian Library at the University of Oxford and the John Rylands University Library at the University of Manchester. Cited as *Workbook of Digital Private Papers*, 2005–2007, available at http://www.paradigm.ac.uk/workbook, accessed 6 February 2009.

Ocatherine Stollar Peters, "When Not All Papers Are Paper: A Case Study in Digital Archivy," Provenance 24 (2006): 23–35. Also Catherine Stollar and Thomas Kiehne, "Guarding the Guards: Archiving the Electronic Records of Hypertext Author Michael Joyce," New Skills for the Digital Era: Proceedings of a Colloquium Sponsored by the National Archives and Records Administration, Society of American Archivists and Arizona State Library, Archives and Public Records (31 May-2 June 2006), available at http://rpm.lib.az.us/NewSkills/CaseStudies/4_Stollar_Kiehne.pdf, accessed 6 February 2009.

¹¹ Douglas Elford et al., "Media Matters: Developing Processes for Preserving Digital Objects on Physical Carriers at the National Library of Australia," World Library and Information Congress: 74th IFLA General Conference and Council, 10–14 August 2008, available at http://www.ifla.org/IV/ifla74/papers/084-Webb-en.pdf, accessed 12 February 2009.

¹² See Electronic Literature Organization, http://eliterature.org/. For discussion on electronic literature, see, for example, N. Katherine Hayle's "Electronic Literature: What Is It?," version1.0 (2 January 2007), available at http://eliterature.org/pad/elp.html, accessed 11 February 2009.

¹³ A recent Swedish study tested the InterPARES requirements on government records. See Kenneth Hänström, "Autenticitet i en Digital Värld: Långsiktsbevarande av Allmänna Handlingar [Authenticity in a Digital World: Long-Term Preservation of Public Records]," Human IT 9, no. 1 (2007): 67–109, available at http://etjanst.hb.se/bhs/ith//1-9/kh.pdf, accessed 5 March 2009.

models for assessing and maintaining the authenticity of copies of electronic records in manuscript collections. The study consists of two parts, an introduction to the InterPARES 1 Authenticity Task Force "Requirements for Assessing and Maintaining the Authenticity of Electronic Records" and an analysis of the George Whitmore Papers, focusing on preservation and description, testing the requirements. The study focuses on rules in the American standard, *Describing Archives: A Content Standard (DACS)*, applied to description in finding aids.

Authenticity, InterPARES, and Electronic Records in Manuscript Collections

The framework for assessing and maintaining the authenticity of electronic records has been the focus of a series of InterPARES and InterPARES-related projects dating back to 1994. The first project, known as the Preservation of the Integrity of Electronic Records (also UBC Project) ran from 1994 to 1997; InterPARES 1 ran from 1999 to 2001; InterPARES 2 ran from 2002 to 2006; and InterPARES 3 started in 2007 and ends in 2012. The general purpose of the first two projects was to test the concepts and methods of diplomatics on modern, electronic records. Early work was conceptual and ambitious: reviewing literature, defining terms, explaining methodology, and presenting research findings, recommendations, and models. While InterPARES 1 focused on electronic records in institutional settings, starting with InterPARES 2, work focused on case studies of electronic records created in different environments, including or encompassing records in an author's archive, and InterPARES 3 proposes to "translate theory and methods . . . into concrete action plans."

InterPARES defines an *authentic record* as "a record that is what it purports to be and is free from tampering or corruption." To assess and maintain the authenticity of a record, a record preserver must be able to establish its identity and demonstrate its integrity. ¹⁸ InterPARES defines the

¹⁴ InterPARES 1, Authenticity Task Force, "Requirements for Assessing and Maintaining the Authenticity of Electronic Records" (March 2002): 1–11, available at http://www.interpares.org/display_file.cfm? doc=ip1_authenticity_requirements.pdf, accessed 31 March 2009.

¹⁵ InterPARES stands for International Research on Permanent, Authentic Records in Electronic Systems. For the website, see http://www.interpares.org/. For summaries of early InterPARES research on the topic, see Heather MacNeil, "Providing Grounds for Trust II: The Findings of the Authenticity Task Force of InterPARES," *Archivaria* 54 (2002): 24–58, or Yvette Hackett, "The Search for Authenticity in Electronic Records," *The Moving Image* 3, no. 2 (2003): 100–107.

¹⁶ Diplomatics emerged in the late seventeenth century as a "technique for determining the authenticity of records issued by sovereign authorities," and it has been used "to help determine a record's authenticity for legal purposes and to assess medieval records as historical sources." Hackett, "Search for Authenticity," 101.

¹⁷ InterPARES 3 website, available at http://www.interpares.org/ip3/ip3_index.cfm, accessed 7 April 2008.

¹⁸ InterPARES 1, "Requirements for Assessing and Maintaining Authenticity," 1.

identity of a record as the "attributes of a record that uniquely characterize it and distinguish it from other records," and the *integrity of a record* as the quality of being complete and unaltered in any essential respects. A record is understood to be complete and uncorrupted in any essential respects "if the message that it is meant to communicate in order to achieve its purpose is unaltered." ¹⁹

InterPARES 1 established formal criteria and recommendations for assessing and maintaining the authenticity of electronic records: one, the benchmark requirements supporting the presumption of authenticity; two, verification; and three, the baseline requirements supporting the production of authentic copies.²⁰ The benchmark requirements relate to the activities of the author or record creator and are meant to serve as the basis for assessing the author's records. These requirements cover

- record attributes, including identity and integrity (A.1);
- access privileges (A.2);
- protective procedures (A.3–A.4);
- establishment of documentary forms (A.5);
- authentication (A.6);
- identification of the authoritative record (A.7); and
- the removal and transfer of relevant documentation (A.8).

Specific identity requirements include the record's author(s), name of action, date(s), and expression of archival bond. Integrity requirements include the name of the author, annotations, and technical changes to the record.

The authenticity of electronic records may also be verified by examining the records themselves and other "reliable information available from other sources about the records and the various contexts in which they have been created and maintained." Methods of verification include steps common to archival practice: comparison of the records with entries on a register; comparison of the records with copies stored elsewhere; analysis of the records' content; and "testimony of a trusted third party." For example, records may be compared to entries on preliminary, appraisal, or transfer lists; drafts of writings in electronic form may be compared to paper drafts or published texts; and donors or executors of an estate may be able to provide trusted testimony.

The baseline requirements supporting the production of authentic copies relate to the activities of the archivist or record preserver and include

¹⁹ InterPARES 1, "Requirements for Assessing and Maintaining Authenticity," 1-2.

²⁰ InterPARES 1, "Requirements for Assessing and Maintaining Authenticity," 5–10.

²¹ InterPARES 1, "Requirements for Assessing and Maintaining Authenticity," 3.

²² InterPARES 1, "Requirements for Assessing and Maintaining Authenticity," 3.

- controls over the record (B.1),
- documentation of the reproduction process (B.2), and
- archival description (B.3).²³

These requirements are meant "to guarantee the records' identity and integrity." 24

One of the questions this paper asks is whether the InterPARES 1 requirements can serve as a standard for assessing and maintaining authenticity of electronic records in manuscript collections. ²⁵ An individual author's electronic records are rarely, or at least less often, subject to the juridical-administrative requirements governing institutional records. Further, to the extent that the records reflect the work of an individual outside an institutional or networked environment, there may be few or no transactional attributes linked to the records. ²⁶ An author working on a home computer, for example, may not implement any access, protective, authentication, or other measures (A.2–A.4 and A.6–A.8). ²⁷ For these and other reasons, benchmark requirements A.2–A.4 and A.6–A.8 can seem irrelevant or impractical. ²⁸

The options would seem to be

1. to adopt the existing InterPARES requirements as a standard for assessing and maintaining electronic records in manuscript collections, acquiring

²³ The OAIS reference model and preservation metadata standards address the framework for record control and documentation requirements. See Consultative Committee for Space Data Systems, Reference Model for an Open Archival Information System (OAIS) (January 2002), available at http://public.ccsds.org/publications/archive/650x0b1.pdf; and PREMIS Editorial Committee, PREMIS Data Dictionary for Preservation Metadata, version 2.0 (March 2008): 1, available at http://www.loc.gov/standards/premis/v2/premis-2-0.pdf, both accessed 10 March 2009. The PREMIS Data Dictionary "defines preservation metadata that: Supports the viability, renderability, understandability, authenticity, and identity of digital objects in a preservation context. . ."

²⁴ InterPARES 1, "Requirements for Assessing and Maintaining the Authenticity," 7.

²⁵ One of the stated objectives of InterPARES 3 is "to establish when such theory and methods, concepts and principles apply across jurisdictions, regardless of legal/administrative, social and cultural environment; and, in the situation where they do not apply, to identify why, and determine the measures that are required to ensure the preservation of digital records." See "Intellectual Framework," version 2.0 (September 2008): 1–2, available at http://www.interpares.org/display_file.cfm?doc=ip3_intellectual_framework.pdf, accessed 27 February 2009.

²⁶ Once in archival custody, records in manuscript repositories are governed by contracts with authors and donors, library policies, and federal legislation. Section 108 in Title 17 of *United States Code* covers reproduction and distribution of copies. See *Copyright Law of the United States* (October 2007), available at http://www.copyright.gov/title17/circ92.pdf, accessed 25 February 2009.

²⁷ According to an InterPARES 2 report, "the case studies showed that record creation in the digital environment is almost never guided by considerations of preservation over the long term. As a result, the reliability, accuracy, and authenticity of digital records can either not be established in the first place or not be demonstrated." See InterPARES 2, "A Framework of Principles for the Development of Policies, Strategies and Standards for the Long-Term Preservation of Digital Records," version 1.2 (March 2008): 2, available at http://www.interpares.org/display_file.cfm?doc=ip2(pub)policy_framework_document.pdf, accessed 27 February 2009.

²⁸ For an early test of the Preserve Electronic Records (now Chain of Preservation) model, based on an InterPARES case study, see InterPARES 1, Preservation Task Force, "Appendix 7: Walkthrough Applying the 'Preserve Electronic Records' Model," version 5.1 (8 October 2001), available at http://www.interpares.org/book/interpares_book_p_app07.pdf, accessed 27 February 2009.

- as much supporting documentation as possible, even if it is incomplete and either weakens the case for authenticity or makes it insufficient; or
- 2. to modify the benchmark requirements and, for example, adopt the record attributes (A.1) and documentary form (A.5)²⁹ requirements, but make other requirements (A.2–A.4 and A.6–A.8) optional; or
- 3. to drop the benchmark requirements; and
- 4. to adopt other criteria.

Beinecke Rare Book and Manuscript Library Case Study: The George Whitmore Papers

The Beinecke Library usually acquires computer media toward the end of an author's career and life or through transfers of no-longer-used media, 30 what the Paradigm project identifies as "traditional" and "transfer via retired media" approaches respectively. 31 In all cases, disks and other media are pulled and logged during accessioning and forwarded to me as the archivist in the Manuscript Unit responsible for managing digital records in collections. 32

We have pursued different preservation strategies with born-digital material in collections over the past decade. As a rule, we follow what Cedars and CAMiLEON project researchers dubbed the "migration on request" approach.³³ This approach requires preservation of the original data object and a means of interpreting or rendering the object when needed. We preserve the original data stream by refreshing³⁴ files into the Yale University

We should try to determine documentary forms when possible. When no determination can be made, the record acquired by the library or archive becomes, by default, the documentary form. InterPARES 1, Authenticity Task Force, "Appendix 2: Template for Analysis," available at http://www.interpares.org/display_file.cfm?doc=ip1_template_for_analysis.pdf, accessed 26 February 2009. This would seem to be especially important with unique and complex works of electronic literature and art.

³⁰ With personal or literary papers, the author is often either dead before the library acquires the archive or there is no communication between the author and archivist before the archive is transferred to the library. Susan Davis makes this point in her recent survey. Susan E. Davis, "Electronic Records Planning in 'Collecting' Repositories," *American Archivist* 71 (Spring/Summer 2008): 169.

³¹ Paradigm identifies several approaches to collection development: regular snapshot accessions, post-custodial approaches, combining snapshot and postcustodial approaches, traditional, transfer via retired media, and self-archiving.

³² A separation sheet is inserted at the location from which a disk(s) or media is removed. The sheet contains contextual information, such as description off a disk label or container.

³³ Phil Mellor et al., "Migration on Request, a Practical Technique for Preservation," The Sixth European Conference on Research and Advanced Technology for Digital Libraries (2002), available at http://www.springerlink.com/content/752vmvw0g0w40dj2/fulltext.pdf (subscription only), accessed 11 February 2009.

³⁴ Refreshment is defined by InterPARES as "the process of copying the digital content from one medium to another." InterPARES 2 Glossary, available at http://www.interpares.org/ip2/display_file.cfm? doc=ip2_book_glossary.pdf, accessed 24 February 2009. Refreshing files into the Rescue Repository preserves the directory structure.

Library Rescue Repository,³⁵ and we use commercial file viewing software (Quick View Plus) to view or render files.³⁶ The software supports common word processing file formats, but there are risks to this variation on the model. The software does not support all file formats, so we cannot render all files, and it is proprietary, so we have no guarantee of its continued availability and viability.³⁷

Hardware and software issues stemming from formats that have come our way and from changes implemented by our systems office have forced us to pursue other preservation strategies, including technology preservation and migration. The first case, with regard to technology preservation, we have maintained a $5\frac{1}{4}$ -inch disk drive to copy files off $5\frac{1}{4}$ -inch disks formatted on DOS-based operating systems. In the second case, with regard to migration, our $5\frac{1}{4}$ -inch disk drive cannot read disks formatted on machines running CP/M operating systems, and the $3\frac{1}{2}$ -inch disk drive in our workstation cannot read old $3\frac{1}{2}$ -inch disks formatted on Macintosh computers. In the past, we have outsourced these disks to have files migrated to readable and standard formats.

³⁵ Starting in 1999, files were refreshed onto a Beinecke server(s), and since 2006, into the Yale University Library Rescue Repository. For information on the Rescue Repository, see http://www.library.yale.edu/%7Elso/databaseadmin/RRweb/AboutRescueRepository.html.

In 2005, PREMIS identified steps for verifying fixity, integrity, and authenticity of digital objects. PREMIS recommends a fixity check program for verifying fixity and format identification and validation for verifying integrity. Technical criteria on authenticity include bit-level preservation, documentation on the history of the object, and use of digital signatures. *Data Dictionary*, 200–201.

Based on PREMIS recommendations, the Rescue Repository runs fixity checks and the file identification and validation tool JHOVE and generates documentation on the history of the object; digital signatures have not been adopted. For repository documentation, see Audrey Novak, "Fixity Checks: Checksums, Message Digest and Digital Signatures" (Yale University Library, Digital Preservation Committee, November 2006), available at http://www.library.yale.edu/iac/DPC/AN_DPC_FixityChecksFinall1.pdf, accessed 24 February 2009. Also Roy Lechich, "File Format Identification and Validation Tools" (Yale University Library, Integrated Library and Technology Services, February 2007), available at http://www.library.yale.edu/iac/DPC/FileIDandValidate.pdf, accessed 24 February 2009

For discussion on authenticity, electronic records, and digital signatures, see Filip Boudrez, "Digital Signatures and Electronic Records" (2005), available at http://www.expertisecentrum david.be/docs/digitalsignatures.pdf, accessed 30 March 2009. For discussions on authenticity as it relates to the OAIS model and to digital objects respectively, see Mahnaz Ghaznavi et al., "InterPARES 2 Written Comments to the Review Committee of ISO 14721:2003 Space Data and Information Transfer Systems—Open Archival Information System—Reference Model (OAIS)." October 2006, unpublished, personal copy from author; and Seamus Ross and Margaret Hedstrom, "Preservation Research and Sustainable Digital Libraries," *International Journal on Digital Libraries* 5, no. 4 (2005): 317–24.

³⁶ Ownership of Quick View Plus has changed since 2002, but the product has been maintained and updated (from version 7 to 10).

³⁷ CAMILEON project advocates recommend development and use of open-source interpreting or rendering tools. Mellor, "Migration on Request," 11.

³⁸ For discussion on migration strategies, see Paul Wheatley, "Migration: A CAMiLEON Discussion Paper," *Ariadne* 29 (September 2001), available at http://www.ariadne.ac.uk/issue29/camileon/, accessed 11 February 2009.

³⁹ CP/M stands for Control Program for Microcomputers. It was in use from the late 1970s to mid-1980s.

These preservation efforts have been incremental and ad hoc. The principal migration-on-request strategy, which seems to have served us well to this point, has also been based in part on a wait-and-see approach. Other preservation strategies⁴⁰ have been adopted as needed, and, over time, the idea of a suite of preservation options has emerged as a general strategy, implemented through a suite of digital preservation tools.⁴¹

Since 2007 we have processed two literary manuscript collections containing born-digital materials: the papers of George Whitmore and the papers of James Welch. Whitmore (1945–1989), a member of the Violet Quill circle of gay male writers, is probably best known for his book, *Someone Was Here*, on the AIDS epidemic. Whitmore's papers, acquired in 1996, several years after his death in 1989, contain writings and research files, a small amount of professional correspondence, and miscellaneous personal papers, including journals and photographs. The papers also include fifty-eight 5½-inch disks containing

⁴⁰ For a list of preservation strategies, see Appendix C of the InterPARES Preserver Guidelines: Preserving Digital Records: Guidelines for Organizations, available at http://www.interpares.org/ display_file.cfm?doc=ip2(pub)preserver_guidelines_booklet.pdf, accessed 10 February 2009. For a survey of trends in "preservation of digital content, with an emphasis on electronic records," see Michele V. Cloonan and Shelby Sanett, "The Preservation of Digital Content," Portal: Libraries and the Academy 5, no. 2 (2005): 213-37. For discussion on the "Performance Model" leading to the strategy (normalization) adopted by the National Archives of Australia, see Heslop et al., "An Approach to the Preservation of Digital Records" (December 2002), available at http://www.naa.gov.au/Images/An-approach-Green-Paper_tcm2-888.pdf, accessed 27 February 2009. For documentation on the XENA software application developed by the NAA, see http://www.naa.gov.au/records-management/secure-and-store/e-preservation/at-NAA/software.aspx#section2. Other strategies include the Multivalent Document Model. See Thomas A. Phelps and P. B. Watry, "A No-Compromises Architecture for Digital Document Preservation," Proceedings of the 9th European Conference on Research and Advanced Technology for Digital Libraries September 2005), available at http://multivalent.sourceforge.net/Research/ Live.pdf, accessed 12 February 2009.

⁴¹ For discussion on digital forensics and forensic workstations, see Jeremy Leighton John, "Adapting Existing Technologies for Digitally Archiving Personal Lives: Digital Forensics, Ancestral Computing, and Evolutionary Perspectives and Tools," *iPRES 2008: The Fifth International Conference on Preservation of Digital Objects* (29–30 September 2008), available at http://www.bl.uk/ipres2008/presentations_day1/09_John.pdf, accessed 30 March 2009. Also "Digital Manuscripts: Capture and Context," *Email Curation: Practical Approaches for Long-Term Preservation and Access* (24–25 April 2006), available at http://www.dcc.ac.uk/events/ec-2006/EC_Digital_Manuscripts_Jeremy_John.pdf, accessed 1 April 2009. See also the National Library of Australia's Prometheus Digital Preservation Workbench, available at http://prometheus-digi.sourceforge.net/. Douglas et al. (2008) describe a "decentralized" processing model, using portable jukeboxes.

⁴² Guide to George Whitmore Papers (YCAL MSS 274), available at http://hdl.handle.net/10079/fa/beinecke.whitmore. For the guide to the James Welch Papers (YCAL MSS 248), see http://hdl.handle.net/10079/fa/beinecke.welch, accessed 25 August 2009.

⁴³ The Whitmore papers were acquired from Michael Canter, executor of Whitmore's estate; Canter's authority as executor is documented on Whitmore's death certificate, which is in the collection.

⁴⁴ The fifty-eight disks in Whitmore and twenty-seven disks in Welch are in fact large quantities for personal papers at Beinecke. Most manuscript collections contain significantly fewer disks.

drafts of writings, research and interview notes, journal entries, correspondence, and other documents. $^{\rm 45}$

The disks, which contain Wordstar⁴⁶ files for CP/M, were outsourced to a data recovery company so that files could be migrated to Wordstar for DOS and ASCII formats.⁴⁷ The converted files were later refreshed into the Yale University Library Rescue Repository, and the physical media were rehoused with the collection (as Restricted Fragile).⁴⁸ Various security measures and access restrictions are in place throughout the workflow. For example, I am currently the only staff member in the Manuscript Unit with rights to the Rescue Repository and repository software.

The "Apparatus of Authenticity" in the George Whitmore Papers

In her 2005 American Archivist article, "Picking Our Text: Archival Description, Authenticity, and the Archivist as Editor," Heather MacNeil recommends that we explore archival description as an aggregate attestation for authenticating born-digital material in collections and that we make our description more transparent. ⁴⁹ MacNeil cites Michelle Light and Tom Hyry's argument

⁴⁵ When dealing with fugitive media, the number of disks and files in a collection may, along with other processing considerations, determine whether or not to conduct item-level analysis. At the Beinecke Library, most files on these formats include drafts of writings or materials relating to writing projects and correspondence (in word processing formats). When possible, we respect original order or context in arrangement. When original order or context cannot be established, in general, small numbers of disks and files lend themselves to item-level analysis and arrangement by content, and larger numbers of disks and files, and disks with mixed files (e.g., writings, correspondence, software files, etc.), lend themselves to arrangement by format. These approaches with fugitive media can also be combined.

When dealing with digital records accessioned through scheduled snapshots and postcustodial approaches, respect original order or context in arrangement, as recommended in the "Paradigm Exemplars for Arrangement," *Workbook of Digital Private Papers*, available at http://www.paradigm.ac.uk/workbook/cataloguing/ead-exemplars.html, accessed 2 March 2009. The electronic records in the Whitmore papers are arranged by content and by format. Records arranged by content are described variously at the file or item level through Series I–V and include folder notes directing patrons to Access Services for reference/use copies. Series VI, Electronic Files, is arranged in disk number order. In the James Welch Papers, drafts of writings in electronic form are arranged by content.

- 46 Wordstar was a word processing application originally written for CP/M (in 1978) and later ported to DOS.
- ⁴⁷ Files were migrated from an unidentified version of Wordstar for CP/M to Wordstar 4.0 for DOS. The collection includes Kaypro program disks for WordStar version 3.3 (1983) and unidentified, possibly earlier, versions.
- ⁴⁸ The Manuscript Unit's processing manual defines *Restricted Fragile* as follows: Items that would be damaged through handling or whose media or material composition would likely incur loss of content over time may be restricted from regular public access and a reference surrogate supplied in its place.
- ⁴⁹ Heather MacNeil, "Picking Our Text: Archival Description, Authenticity, and the Archivist as Editor," *American Archivist* 68 (Fall/Winter 2005): 264–78. MacNeil identifies archival description as an "apparatus of authenticity," 271. The InterPARES 1 Authenticity Task Force discusses the idea that description should serve as a collective attestation in "Requirements for Assessing and Maintaining the Authenticity of Electronic Records," 8, 11.

for a colophon as one means for achieving greater descriptive transparency and use of existing standards as another.⁵⁰ I will limit my discussion in this section to use of existing standards, reviewing the documentation relating to the electronic records in the Whitmore papers and the Whitmore description. The questions I am asking here are whether the InterPARES baseline requirements are being met in practice and whether *DACS* supports the InterPARES baseline requirement (B.3) for description.⁵¹

Several forms of documentation commonly exist on the provenance and archival custody of born-digital objects in a collection. The library's curatorial files typically include correspondence and associated documentation leading up to and following an acquisition. The curatorial files for Whitmore include the following: a professional appraisal of Whitmore's papers, dated 25 January 1996, noting the existence of a "box of disks" associated with *Someone Was Here*; a pickup note, dated 25 March 1996, presumably documenting the physical transfer of the collection; and an acquisition authorization, signed and dated 6 May 1996, documenting curatorial review of the collection.⁵²

Next, the existence and number of disks are documented in the accession record for the collection in the library's acquisitions database (see Figure 1). The presence of disks is usually captured at the time the record is created; otherwise, the record is updated when disks are discovered during rehousing. For patron and staff discovery, we have found it useful to note disks in the title field, notes fields, and the material type field.⁵³ The title field provides a key entry point, the notes fields provide the number of disks and, in more recent records, the physical location of disks in the collection, and the material type ("computer file" and "cf") provides access by format.⁵⁴

Third, because the Whitmore disks were outsourced, we have a loan form with the data recovery company, signed and dated, documenting the names of

MacNeil proposes a model for archival description, re-envisioning the finding aid "as part of a complex network of hyperlinked and interactive documentation relating to the history, appraisal, preservation, use, and interpretation of a body of records over time." MacNeil, "Picking Our Text," 275–76. See also Michelle Light and Tom Hyry, "Colophons and Annotations: New Directions for the Finding Aid," American Archivist 65 (2002): 216–30.

⁵¹ InterPARES baseline requirement (B.3) for description states that "the archival description of the fonds containing the electronic records includes—in addition to information about the records' juridical-administrative, provenancial, procedural, and documentary contexts—information about changes the electronic records of the creator have undergone since they were first created." InterPARES 1 (March 2002): 8.

⁵² The appraisal highlights drafts and research materials associated with Whitmore's writings, and the "box of disks" associated with Someone Was Here accounts for only a portion of the disks in the collection.

⁵³ Description in the acquisition database does not follow *DACS*. The title, for example, is essentially a brief scope and content note.

⁵⁴ In principle, one should retrieve all collections with computer disks by a keyword search on the title field or a material type search for computer files.

Accession Number

19960506-k

Date Entered

7 May 1996

Call Number

YCAL MSS 274

Author

Whitmore, George, 1945-1989

Title

Papers: ms., notes, printed material, audio tapes and computer disks by or relating to George Whitmore. Includes The Confessions of Danny Slocum, typescript and setting copy; The Caseworker (play), typescript; and poetry, including Tricking, ms., and printed version

Place

V.p.:

Date

[ca. 1960-1989]

Physical Description

9 boxes (2 boxes added 2/06)

Notes

Restricted Fragile: Boxes 7-8, 58 disks, use copies available; consult Public Services. Formerly Uncat.ZA MS.484

Department

Yale Collection of American Literature

Material Type(s)

archives and manuscripts (ms) audio-visual material (av) computer file (cf)

FIGURE 1. Accession record for George Whitmore papers.

the institutions and staff responsible and the dates the disks went out and returned, and summarizing the request.⁵⁵

Fourth, when files are refreshed into the Yale University Library Rescue Repository, all events are recorded, creating an audit trail. Events are documented in a session log, a JHOVE log, and in a database. ⁵⁶ The metadata is based loosely on PREMIS elements and includes, for example, the identity of the individual user, the identity of the collection (contributing unit or library), session dates and times, software, checksums, and file verification and validation results.

Fifth, in addition to the metadata captured automatically in the Rescue Repository, the unit maintains metadata on preservation events preceding

⁵⁵ I transported the Whitmore disks to and from the data recovery company, toured the facility and resources, discussed the job with the project manager, etc.

⁵⁶ JHOVE stands for JSTOR/Harvard Object Validation Environment. For the JHOVE website, see http://hul.harvard.edu/jhove/, accessed 22 June 2009.

refreshment, including a unique identifier, the file title, eventType, eventDate, eventOutcome, and agentIdentifier. And finally, when files are refreshed into the Rescue Repository, brief descriptive records (based on Dublin Core) are created for individual files, including a unique identifier, title, contributor, collection, call number, and, when present, disk number.⁵⁷

This documentation, from the paper slips in the curatorial files to distributed preservation metadata, supports the baseline requirements on record controls and documentation (B.1–B.2) and informs the final description (B.3).⁵⁸ The identity of individual records is addressed in a preliminary way through our initial preservation efforts, reproduction into the Yale University Library Rescue Repository, and then again through analysis and arrangement in processing. The research notes and drafts in electronic form for Whitmore's book on the AIDS epidemic, for example, are identified and contextualized within the writing series in the collection.⁵⁹

The integrity of electronic records in personal papers, and on fugitive media in particular, is a challenge to assess and guarantee when we know little or nothing about the manner in which the records have been created and maintained. Questions about the integrity of the electronic records in Whitmore's papers arise from the records' creation and maintenance and later reproduction. In the first case, the estate's executor confirmed that he had stored the archive between Whitmore's death in 1989 and transfer of the archive to the library in 1996, and that Whitmore's "early Kaypro" computer was discarded in 1990 or 1991, but he was not able to provide additional information on the creation and maintenance of Whitmore's electronic records. ⁶⁰ In the later reproduction, the dates of the files on the $5\frac{1}{4}$ -inch disks were lost in the migration, and the migration dates prevent us from confirming that the original dates predate Whitmore's death. ⁶¹ Lack of information on the records' creation and maintenance and lack of the original (disk) file dates and certain metadata on the reproduction process weaken the documentary support.

⁵⁷ The Dublin Core Contributor element is used in place of the Creator element at this point in the workflow. The collection title (e.g., George Whitmore Papers), call number, and disk number are all treated as Relation elements.

⁵⁸ Metadata produced at the Beinecke Library meets minimum requirements and is not sufficiently integrated and automated.

⁵⁹ In Series IV, Writings on AIDS, described at the file level, and in Series VI, Electronic Files.

⁶⁰ Email communication with Michael Canter, 15-17 April 2009.

⁶¹ The files on the disks would have a creation or modification (possibly even a transmission) date. See InterPARES 1 Authenticity Task Force, *Benchmark Requirement A.1.a.iii.*, 5. Files were migrated from 28 July to 3 August 2005. The collection was processed in 2007. Analysis of files with Quick View Plus yielded dates between 1984 and 1988, predating Whitmore's death in 1989, for some files. All other files are listed as "undated" in the finding aid.

MacNeil's recommendation that we make our description more transparent also raises questions. First, what are the rules addressing electronic records? Second, are the existing rules and examples sufficient? And third, what level of transparency is required or desirable?

Information about the record's provenancial, procedural, and documentary contexts is addressed in *DACS* Title (2.3), Name (2.6), Administrative/Biographical History (2.7), Scope and Content (3.1), System of Arrangement (3.2), and Acquisition (5.1–5.2) elements.⁶² Other *DACS* elements likely to be used in describing electronic records in personal papers include Extent (2.5.5–2.5.6), Information on Access (4.2.7 and 4.3.6), Related Materials (6.1–6.2), and Notes (7.1.4).⁶³ The peculiarities of media in collections, arrangement, and local means of access and descriptive practice will most likely dictate use of the Extent and Information on Access elements.

For example, at the Beinecke Library, we record the quantity of material at the collection or fonds level in terms of linear feet and number of boxes in accordance with *DACS* 2.5.7. In a hybrid archive, however, I would recommend the material type be highlighted in accordance with 2.5.6 or 2.5.8. The extent of fugitive media is appropriately described, depending on arrangement, at series, subseries, or file levels. In the Whitmore papers, the extent is recorded in the finding aid at the series level.⁶⁴

Access restrictions on original media and objects, as well as technical requirements for access, are noted in the Access Restrict element <accessrestrict> in accordance with DACS 4.2.7.65 The statement in the Whitmore finding aid reads as follows:

⁶² InterPARES 1 Authenticity Task Force, Benchmark Requirement A.1.a.iii, 2. InterPARES defines the provenancial context as "the creating body, its mandate, structure, and functions," procedural context as "the business process in the course of which the record is created," and documentary context as "the archival fonds to which a record belongs."

⁶³ At the Beinecke Library we have mostly used the Existence and Location of Originals element in accordance with 6.1.4 to refer to originals located in another repository. When the library holds copies of electronic records on portable storage media, if it can be established that the records on an individual's hard drive(s) have been destroyed, it may be appropriate to note this in a Scope and Content (3.1) or Custodial History (5.1.3) element. The exclusion (6.2.1) to the Existence and Location of Copies element means that the existence of copies is noted in the Access Restrict element.

⁶⁴ Extent may be recorded in terms of bytes or number of files. Though *DACS* does not offer an example of digital extent recorded in terms of bytes, 2.5.5 seems to allow for it. See also *RADS* 9.5B2, *ISAD(G)* 3.1.5, and the Paradigm fonds-level description recommendations, available at http://www.paradigm.ac.uk/workbook/cataloguing/ead-fonds.html, accessed 2 March 2009.

⁶⁵ Technical requirements for patron access can be noted in the Physical Characteristics/Technical Requirements element <phystech> in accordance with DACS 4.3.6. As this element is rarely used in Yale University Library finding aids, this information has been noted at least provisionally in the Access Restrict element <accessrestrict>.

Original computer disks are restricted. Copies of electronic files are available through Access Services. 66 System requirements include a Beinecke Library laptop with CD/DVD drive and file viewing software.

The custodial history is addressed in Acquisition elements 5.1 and 5.2.67 DACS 5.1 "provides information on changes of ownership or custody of the material being described, from the time it left the possession of the creator until it was acquired by the repository, that is significant for its authenticity, integrity, and interpretation," 68 but then excludes (5.1.1) information about the donor or source to the Immediate Source of Acquisition Element (5.2). In the Whitmore finding aid, conforming to DACS guidelines, information on the custodial history and acquisition is split between <custodhist> and <acquiro> elements.69 (See Figure 2.)

DACS 7.1.4 recommends that we give details of any migrations or logical reformatting since transfer to archival custody in a notes element. This notes element is optional in DACS, but we adopted it at the Beinecke Library in accordance with the InterPARES baseline requirement B.3. In our finding aids, this goes into the processing note processinfo. Following current DACS guidelines, the Whitmore example might look as follows:

Electronic files migrated by National Data Conversion from the original word-processing software (WordStar for CP/M) to WordStar 4.0 for DOS and to ASCII to maintain readability of data. Technical specifications are filed with the media in *Restricted Fragile*.

An informal survey of description in finding aids across institutions in the United States, Canada, and the United Kingdom reveals varied use of these terms, though preference may be emerging for the term *digital*. See, for example, InterPARES 2, "A Framework of Principles for the Development of Policies, Strategies and Standards."

Given variations in language in the description standards and in institutional practice, guidance could be provided by the International Council on Archives (ICA) or SAA.

⁶⁶ The heading "Electronic Files" was adopted by the Manuscript Unit in 1999 for all levels of description in finding aids. The rationale for use of these terms is documented in the department's processing manual as follows: "Electronic (*RAD2* 1.1C4) is preferred to Computer (*AACR2* 1.1C1) as a broader and ostensibly more accurate term, one, for example, that can encompass files created on contemporary portable devices (such as digital cameras, cell phones, PDAs, etc.) not commonly identified as computers. Electronic is preferred to Digital as a broader term, encompassing both analog and digital formats. File(s) is preferred to record, document, copy, and other terms."

⁶⁷ In an article currently under review, "Trusting Description: Authenticity, Accountability and Archival Description Standards" (2009), MacNeil explores the link between authenticity and description through examination of the *International Standard for Archival Description* or *ISAD(G)* and proposes we expand the definition of provenance to better capture the history of records.

⁶⁸ Describing Archives: A Content Standard (Chicago: Society of American Archivists, 2004), 60.

⁶⁹ DACS exclusions (5.1.1. and 5.2.1) do not prohibit use of the custodial history element alongside the immediate source of acquisition when, in cases like Whitmore, the custody overlaps with the source of acquisition. Rather, they simply separate information. In the EAD Tag Library version 2002, however, <custodhist> and <acquinfo> guidelines indicate that the <custodhist> element is reserved for "information about the chain of ownership... before [emphasis is mine] the material reached the immediate source of acquisition," 88.

Administrative Information

Provenance

The George Whitmore papers were purchased from Michael Canter on the Ezra Pound Archive Fund in 1996.

The Whitmore archive was stored by Canter between Whitmore's death in 1989 and transfer of the archive to the library in 1996. According to Canter, Whitmore's "early Kaypro" computer was discarded in 1990 or 1991.

Information about Access

This collection is open for research.

Original computer disks are restricted. Copies of electronic files are available through Access Services. System requirements include a Beinecke Library laptop with CD/DVD drive and file viewing software. Surrogates of all audio recordings are available within the collection.

Restricted Fragile in boxes 20 and 21 may be consulted only with permission of the appropriate curator.

Ownership & Copyright

The George Whitmore Papers is the physical property of the Beinecke Rare Book and Manuscript Library, Yale University. Literary rights, including copyright, belong to the authors or their legal heirs and assigns. For further information, consult the appropriate curator.

Cite As

George Whitmore Papers. Yale Collection of American Literature, Beinecke Rare Book and Manuscript Library.

Processing Notes

In addition to the paper files that document Whitmore's major works in fiction and non-fiction, there are fifty-eight computer disks and seven cassettes with audio recordings. In most cases, the electronic documents have paper printouts in the collection, but readers who would like to consult the electronic files may do so through the Beinecke Library's Access Services staff. The finding aid has been annotated to include these files in their appropriate places within the collection's intellectual arrangement. Surrogates of all audio recordings have been created for patron use, and are filed at the appropriate place within the collection.

The electronic files were migrated by National Data Conversion from the original word-processing software (WordStar for CP/M) to WordStar 4.0 for DOS and to ASCII to maintain readability of data. WordStar 4.0 for DOS and ASCII files were refreshed into the Yale University Library Rescue Repository. Technical specifications are filed with the media in *Restricted Fragile*.

In this finding aid, the term "typescript" is used to identify papers that Whitmore produced with either a typewriter or a computer and printer.

Alternative Formats

Audio Recordings.

FIGURE 2. Administrative information for George Whitmore Papers.

This argument is preliminary, but the "apparatus of authenticity" may work better as a means of aggregate attestation if we expand recommendations on preservation actions in *DACS* 7.1.4 to include refreshment or ingest into a

⁷⁰ DACS 7.1.4 reads: "If the materials being described are in electronic form, give details of any migration or logical reformatting since its transfer to archival custody..." DACS (2004), 78. A revised rule might read as follows: "If the materials being described are in electronic form, give details of any preservation actions following transfer to archival custody, including ingest into a digital repository and migration or logical reformatting. Indicate the location of any relevant documentation."

digital repository.⁷⁰ If we are more transparent about the reproduction process, an expanded processing note might look as follows:

Electronic files migrated by National Data Conversion from the original word processing software (WordStar for CP/M) to WordStar 4.0 for DOS and to ASCII to maintain readability of data. WordStar 4.0 for DOS and ASCII files refreshed into Yale University Library Rescue Repository.⁷¹ Technical specifications are filed with the media in *Restricted Fragile*.

In my experience at the Beinecke Library since 2002, most born-digital objects in a manuscript collection are successfully preserved through the same preservation strategy, with only the odd exception or problem. In such cases, when limited in number, we might go one step further and document preservation exceptions and failures. An expanded note might look as follows:

Electronic files migrated by National Data Conversion from the original word processing software (WordStar for CP/M) to WordStar 4.0 for DOS and to ASCII to maintain readability of data. WordStar 4.0 for DOS and ASCII files refreshed into Yale University Library Rescue Repository. Disk#34 unreadable; no files copied. Technical specifications are filed with the media in *Restricted Fragile*.

In addition to the recommendations on preservation events for electronic records in archival custody, *DACS* 7.1.4 tells us to "Indicate the location of any relevant documentation." Provisionally, we are filing printouts of metadata associated with electronic records, as well as a summary of preservation events, with the media in *Restricted Fragile*.

Conclusion

The literature examining changes in personal computer use over the past several years shows that the digital components of individuals' archives are growing and becoming increasingly complex.⁷² Fugitive media will no doubt

⁷¹ The Yale University Library is currently developing a Fedora-based digital repository to replace the Rescue Repository. *Ingest* is defined as the "Process of adding objects to a preservation repository's storage system." In the context of OAIS, ingest includes services and functions that accept Submission Information Packages (SIP) from producers and transforms them into one or more Archival Information Packages (AIP) for long-term retention. *Data Dictionary*, 212.

⁷² There is a "second wave" of literature addressing contemporary computer use habits, types of digital records and environments, and preservation strategies. See, for example, Neil Beagrie, "Plenty of Room at the Bottom? Personal Digital Libraries and Collections," *D-Lib Magazine* 11 (June 2005), available at http://www.dlib.org/dlib/june05/beagrie/06beagrie.html, accessed 16 February 2009; Catherine C. Marshall, "Rethinking Personal Digital Archiving: Part 1," *D-Lib Magazine* 14 (March/April 2008), available at http://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html, accessed 8 April 2009.

continue to come in with collections through traditional and transfer-viaretired media approaches, but to better understand the contexts in which authors are now working, and to assess and maintain their records, staff members in collecting repositories need to be proactive. Staff members at collecting repositories have made some recent efforts to work with authors earlier in the records' life cycle: repository employees are creating and sharing digital preservation guidelines with authors, conducting records surveys, developing new donor and legal agreements and transfer tools, and arranging snapshot accessions of authors' digital archives.⁷³ At the Beinecke Library, for example, we have adapted the InterPARES *Creator Guidelines* and the Paradigm records survey for use with literary authors, we have tested the records survey with an author in midcareer, and we are developing new transfer tools for a snapshot accession.⁷⁴

Though this case study is intended as a preliminary examination of the authenticity requirements for electronic records in manuscript collections, I hope that it demonstrates the value of the InterPARES requirements and the capacity of *DACS* to meet most of our needs. Future work on collections containing electronic records should allow us to further test and refine these requirements.⁷⁵ The number of manuscript collections containing electronic records acquired by the Beinecke Library has been increasing over the past several years, and, as the size and complexity of the digital component of archives increases, it will help to apply standards consistently in the management of these records. Moving forward, I recommend that repositories adopt the InterPARES 1 Authenticity Task Force requirements as a model for assessing and maintaining the authenticity of electronic records in manuscript collections, and, in addition to possible revisions to *DACS* 7.1.4, I believe we need to be transparent in our descriptions of the history and management of electronic records in our collections.⁷⁶

⁷³ In the United States, for example, the Beinecke Rare Book & Manuscript Library at Yale; the Harry Ransom Center at UT Austin; the Manuscript, Archive, and Rare Book Library (MARBL) at Emory University; and the Dodd Research Center at the University of Connecticut have implemented or are in the process of implementing these or similar practices.

⁷⁴ Beinecke Rare Book & Manuscript Library, "Authors' Guidelines for Digital Preservation," available at http://www.library.yale.edu/%7Enkuhl/AuthorsGuidelines.pdf, accessed 22 June 2009. See also InterPARES 2, *Creator Guidelines: Making and Maintaining Digital Materials: Guidelines for Individuals*, available at http://www.interpares.org/ip2/display_file.cfm?doc=ip2(pub)creator_guidelines_booklet.pdf, accessed 6 March 2009.

⁷⁵ It would be great to provide access to copies of electronic records through the finding aid—delivery of individual records also requires some form of authentication.

⁷⁶With fugitive media, the burden falls on archivists to try to verify authenticity.