Article

Balancing Care and Authenticity in Digital Collections: A Radical Empathy Approach to Working with Disk Images

Monique Lassere and Jess M. Whyte

ABSTRACT

Radical empathy frameworks ask us to carefully consider the presence of sensitive information within digital archives; those who created, are captured by, and are affected by a record (or the absence of that record); and the consequences of retaining or discarding that information. However automated digital archiving workflows often discourage contextual and empathetic decision-making in order to efficiently handle the scale and volume of digital content.

This paper explores the default practice of "take and keep it all" during the acquisition of media-bound digital content and its ethical implications on labor and privacy within the context of radical empathy. Accessioning and processing practices which promote retention of complete disk images and encourage the creation of access copies with redacted sensitive data increase the risk of harm for creators, subjects, and those affected by a record. The decision to discard must be deliberate and, often, must be enacted manually, outside of the automated workflow.

The motivation for the disk imaging model is that the researcher, archivist, curator, or librarian can always return to the original disk image in order to defer additional labor, demonstrate authenticity, allow for emulation or access, or to generate new access copies. However, this practice poses ethical privacy concerns and does not demonstrate care for creators and subjects. We recognize that the resources necessary to review disk images and make contextual decisions that balance both privacy and

Lassere, Monique and Jess M. Whyte. "Balancing Care and Authenticity in Digital Collections: A Radical Empathy Approach to Working with Disk Images," in "Radical Empathy in Archival Practice," eds. Elvia Arroyo-Ramirez, Jasmine Jones, Shannon O'Neill, and Holly Smith. Special issue, *Journal of Critical Library and Information Studies* 3.

ISSN: 2572-1364

authenticity are significant due to the manual nature of this work: this places strain and further labor on staff and practitioners using current digital archival and preservation tools. However, we believe that there is an outstanding need to develop tools which aid in efficient and explicit redaction, but also allow for much-needed contextual and empathetic decision-making that aligns with developing concerns and evolution in archival practice. In this paper we propose that more resources, particularly staff time, are required to make these decisions and if those resources are not available, then the institution should consider itself incapable of ethically stewarding the content and protecting those affected.

INTRODUCTION

This paper explores the implications of the common practice to "take and keep it all" in the acquisition of digital archives and its labor and privacy implications.

Radical empathy frameworks in archiving ask us to carefully consider the presence of sensitive information; those who created, are captured by, and are affected by a record (or its absence); and the consequences of retaining or discarding that information for those affected. It is our position that available, recommended digital archiving workflows and tools, often in order to handle the volume of digital content or out of fear of violating preservation practices, discourage contextual and empathetic decision-making in favor of default decisions influenced by the primary audiences for whom these tools are developed, policy enforcement, and surveillance forces.

This paper begins by defining the scope of inquiry and the technical terms employed. We will then demonstrate that the creation and retention of disk images and associated metadata is often the default practice and discuss the reasons for this, outlining the historical relationship between archival work and digital forensics. We will provide a definition of radical empathy in this context, and demonstrate how its application might benefit the development of workflow tools. Finally, our outline of currently available digital curation tools will, in the context of radical empathy, evaluate their potential for allowing archivists to make contextually informed, empathetic, and ethical decisions.

DEFINING SCOPE AND TERMS

Michelle Caswell and Marika Cifor call upon archivists and scholars to further consider how radical empathy and feminist ethics of care can help us rethink the roles or decision-making of collectors and describers and consider the perspectives and vulnerabilities of others, and awareness of power relationships in our work.[1] This paper is an answer to that call, focusing specifically on the acquisition and retention of previously media-bound digital content, particularly in archival personal collections.

We will begin by defining terminology specific to digital forensics and digital preservation, before delving into workflow mechanics and the application of radical empathy within those structures. First, what do we mean by media-bound digital content? This is data encoded, or "saved," on a media carrier, typically a disk such as a hard drive, USB key, CD, DVD, or floppy. As contemporary content creation moves away from single-media storage like local hard drives to redundant, distributed storage (e.g., a

[1] Michelle Caswell and Marika Cifor, "From Human Rights to Feminist Ethics: Radical Empathy in the Archives," *Archivaria* 81 (Spring 2016): 42.

personal Google Drive), the presence of media-bound digital content in a personal collection will become less prevalent. But, as the influx of new archival acquisitions can often follow a few decades behind that of creators', its presence in a new donation is still likely. Archivists and collection stewards may also inherit a backlog of media-bound materials in existing collections.

Before beginning to work with media-bound digital content, many digital archivists and librarians will first create what is called a disk image. By this, we do not mean a photograph of the media itself, nor do we mean a logical copy of the files on that disk. We defer here to Matthew Kirschenbaum's succinct definition,

> "a literal representation of every *bit* of information on some original instance of source media...it is not simply a copy of all of the files that were once on that original diskette; rather the disk image...preserves all of the *information* that was recorded on the disk in its original storage geometry."[2]

Kirschenbaum describes how a disk image lets us treat a carrier as its own artifact or entity, instead of a carrier for individual objects or files. With a disk image, we represent the content not as files, but as individual bytes encoded into a whole. That disk image can then be examined for content unorganized into files (e.g., deleted content or content not in the filesystem), metadata relating to the carrier (e.g., volumetric properties, file structure, etc.) and the organization of files or bytes on that media carrier. It can also be mounted and explored as a logical disk and one can extract individual files from it. Disk images are often referred to as a "bit-level copy" or a "raw image." It may, depending on the media and context, also be referred to as an "ISO" (typically for CD-ROMs), a "DMG," a "dd," or a "block-level" image. A disk image is often stored as a single file or a collection of files in a single directory whereas a logical copy of a disk would be a traditional file system and directory structure containing copied, individual files (see Diagram 1). An analogy that clearly demonstrates the difference between disk images and logical directories is recording the contents of an entire cassette tape as one entity versus individual songs.

Because disk images are representations of the physical media, they may contain content that has been deleted from the file system but are still present on the disk itself. This content may be unwanted by the collecting repository or is unethical to keep due to a donor or creator being unaware that this content was preserved.[3] For example, a media

---

[2] Matthew Kirschenbaum, *Mechanisms: New Media and the Forensic Imagination* (Cambridge, MA: MIT Press, 2008), 115.

[3] Kam Woods, Cal Lee, and Simson Garfinkel, "Extending Digital Repository Architectures to Support Disk Image Preservation and Access," in *Proceedings of the 11th Annual International*

carrier might contain four files, three are relevant to the collection, but the fourth belongs to a third-party, say a family member of the donor. The fourth file is not relevant, contains private information, and, if it were a piece of paper, would likely be destroyed or returned. Choosing to keep a full disk image of that media carrier would retain that third file.

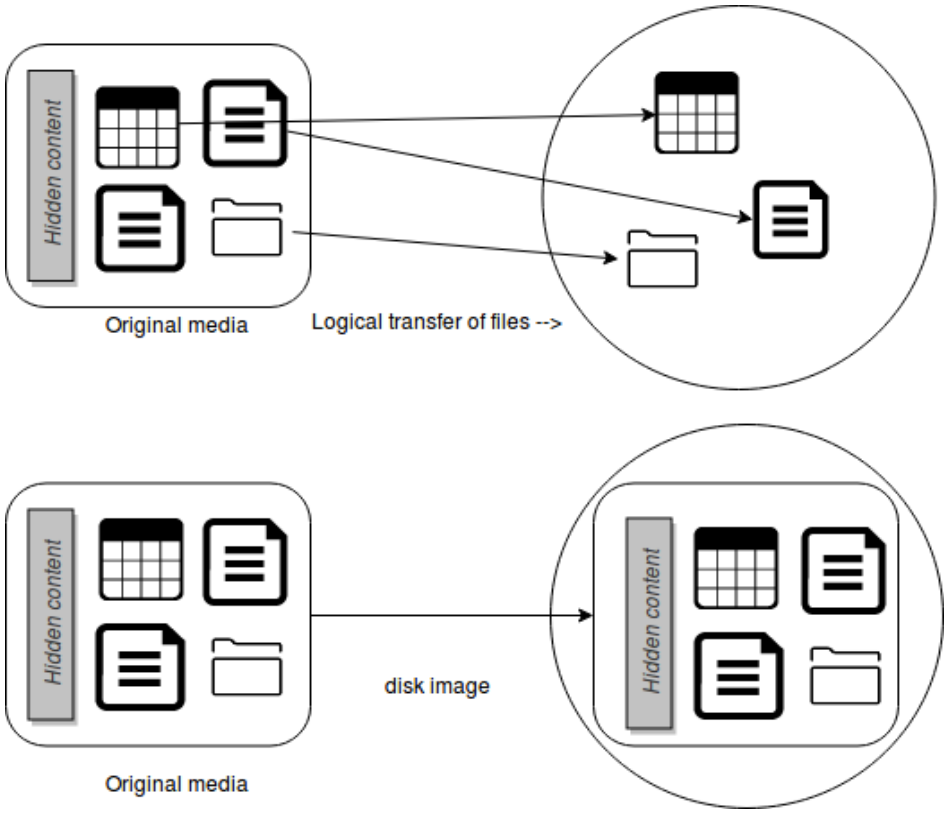

Figure 1. Illustration of a logical file transfer compared to a disk image. Special thanks to Steve Marks for providing the whiteboard inspiration for this illustration.

ACM/IEEE Joint Conference on Digital Libraries (New York: Association of Computing Machinery, June 2011), 64.

BACKGROUND ON THE CAPTURE AND RETENTION OF DISK IMAGES IN DIGITAL PRESERVATION WORK

In their 2013 article examining current management of sensitive information in born-digital collections, Ben Goldman and Timothy D. Pyatt noted that disk imaging and forensic acquisition methods were increasingly a part of best practice workflows in digital archives.[4] They address the privacy challenges arising from this approach:

- Media was often unexamined at acquisition and donors may no longer be available for consultation if questionable content is surfaced;
- The backlog and volume of incoming content is sometimes not reviewable with available resources;
- Technical solutions for scanning for sensitive personal information are not sufficient;
- Seemingly innocuous hidden file metadata can pose privacy issues;
- The range of hidden or deleted content available in disk images is extensive; and
- Despite the coalescing of practices around the disk imaging process as part of the digital archives workflow, both archivists and donors are not always aware of the types of content that can be surfaced with forensic methods.[5]

Given the privacy issues raised by this method, as identified by Goldman and Pyatt, why is disk imaging the preferred method, and why are disk images retained, at times indefinitely? We believe there are four predominant reasons the capture and retention of disk images and/or their associated metadata became a recommended practice in this field. First, traditional paper-based conservation principles discourage any action that may reduce a record's value as evidence, or any process of repair that is irreversible.[6] Secondly, there is an existing relationship between diplomatics, archival science, and digital forensics.[7] Thirdly, disk images allow us to defer labor. Finally, the clearest interpretation of a digital object may require additional data outside of its own boundaries. In practical terms, a full disk image is sometimes required in order to fully

---

[4] Ben Goldman and Timothy D. Pyatt, "Security Without Obscurity: Managing Personally Identifiable Information in Born-Digital Archives," *Library and Archival Security* 26, no. 1-2 (2013): 43.

[5] Goldman and Pyatt, "Security," 43-44.

[6] Roger Ellis, *The Principles of Archive Repair: A Paper Read by Roger Ellis* (London, UK: London School of Printing and Graphic Arts, 1951).

[7] Luciana Duranti, "Diplomatics: New Uses for an Old Science, Part I," *Archivaria* 28 (Summer 1989): 7-27; Luciana Duranti, "From Digital Diplomatics to Digital Records Forensics," *Archivaria* 68, (Fall 2009): 39-66.

render a digital object in such a way that it is seen not in violation of the traditional conservation principles listed above.

In 1951, Roger Ellis, under the guidance and mentorship of Hilary Jenkinson[8] at the UK National Archives, formulated a set of five principles to guide conservators and archivists. These principles were succinctly summarized by David Baynes Cope, an expert in document examination and preservation, as the following: first, no process of repair should remove, diminish, or obscure the document's value as evidence; second, no process of repair should be used which would in any way damage or weaken the materials of which the document is made; third, processes of repair should not interfere in any way with any subsequent treatment that the document may require; fourth, the process of repair should not diminish in any way the aesthetic appearance of non-archival material; and lastly, no process of repair should be irreversible.[9]

These principles are not, nor have ever been, the final word on conservation and preservation. Nevertheless, their influence is felt.[10] Jenkinson's influence on the conception of archival materials as evidence is particularly relevant to disk imaging, as are his conclusions that no grounds exist for records destruction and that records must remain as received.[11]

The importance of the evidentiary nature and capacity of records can further be seen in the development and theory of diplomatics, advanced in the 1989 article, "Diplomatics: New Uses for an Old Science, Part I" by Luciana Duranti, professor of archives and the director of the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) project. Diplomatics can be loosely defined as the study of documents and was born out of a need to standardize how documents were authenticated during the late Medieval period. In the series of articles produced by InterPARES, Duranti notes a feedback loop between records management (the archivist's original duty) and diplomatics, outlining the inescapable "historical-administrative-legal-archival" nature of both disciplines and the inherent link between the two.[12] At the root of each is a need to authenticate records across historical, administrative, and legal mandates, and to prove their integrity. According to Duranti, when records are digital, integrity is linked directly to the chain-of-custody and the object's 0's and 1's:

---

[8] David Baynes-Cope, "Thoughts on Ethics in Archival Conservation," *Restaurator: International Journal for the Preservation of Library and Archival Material* 9, no. 3 (1988): 137.

[9] David Baynes-Cope, "Principles and Ethics in Archival Repair and Archival Conservation. Part 1: The Principles of Archival Repair and of Archival Conservation," *Journal of The Society of Archivists* 15 (1994): 18.

[10] Richard Stapleton, "Jenkinson and Schellenberg: A Comparison," *Archivaria* 17 (1983): 83

[11] Hilary Jenkinson, *A Manual of Archive Administration Including the Problems of War Archives and Archive Making (*Oxford, UK: The Clarendon Press, 1922) 19-84; Stapleton, "Jenkinson and Schellenberg," 81.

[12] Duranti, "Diplomatics: New Uses," 10-11.

> The integrity of a record is linked to its ability to convey the message it was intended to communicate when generated. Thus, it does not matter if the ink is fading, the medium (i.e., the material support) is falling apart, or the bit-stream is not the same as in the first manifestation of the record, as long as the content is readable and is the same as it was originally intended, the medium does not have missing parts, or the manifestation we see on the computer screen is the same as it was the first time the record was saved.[13]

While diplomatics holds that no instantiation of a digital object constitutes an original, digital forensics and archival science rely on chain of custody and the existence of an original record to prove authenticity and integrity.[14] Disk images play a crucial part in ensuring that the structure and nature of a digital object remains unmodified throughout its handling, regardless of the overarching intention for that object and the content it represents. The InterPARES Project defined the trusted custodian (or recordkeeper) as one who can attest to, ensure, and demonstrate the authenticity and availability of records, usually through the fulfillment of baseline requirements for a trusted recordkeeping system.[15] The role of the trusted recordkeeper extends to the software and computing environments digital records inhabit through the archival and forensic process. The disk image then can become one means by which the recordkeeper secures the archival bond between records.

In 2001, the British Library, began processing digital personal papers on physical media (which it referred to as "eMANUSCRIPTS"),[16] through its Digital Manuscripts Project. The project identified three key requirements: 1) to capture full contextual disk images for authentication purposes; 2) to retain "exact copies" of files; and 3) to meet confidentiality requirements sensitive to depositors.[17] They were also early adopters of forensic methods. The careful protection and consideration of this content, its evidentiary value, and the heralding of disk images for authentication purposes resonates with Ellis's

---

[13] Duranti, "From Digital Forensics," 53.

[14] Duranti, 58.

[15] Heather MacNeil et al., "Part 1: Establishing and Maintaining Trust in Electronic Records – Authenticity Task Force Report," *The Long-Term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* (2001), 21, http://www.interpares.org/book/interpares_book_d_part1.pdf, 21.

[16] A. Summers and J. Leighton John, "The W. D. Hamilton Archive at the British Library," *Ethology, Ecology & Evolution* 13: 373-384; 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, London* (September 2008): 48-55.

[17] Leighton John, "Adapting Existing Technology."

and Duranti's influence. Yet concerns about privacy implications remained. In his 2008 outline of the eMANUSCRIPTS work, Jeremy Leighton John stresses repeatedly the "essential need to involve potential depositors in the capture process."[18] Leighton John adds that, while recovery of deleted drafts or overwritten works might be of scholarly interest, retrieval of this content "must involve the originators and accord with their wishes." He also remarks that the originator may have only intended to donate a specific file, not a full forensic image of their drive. However, the forensic imaging processes and recommended practices for "capture" outlined in Leighton John's work did not yet include a method for discarding content entailing sensitive information. It was recommended that this content "be identified by the curator and bookmarked," but it remained unclear whether the original disk image would be discarded, or related metadata redacted after unwanted content had been identified. From early days, the field was concerned about the surfacing of sensitive information by forensic methods, but without the technical mechanisms or labor capacity to address it.

In the 2010 Council for Libraries and Information Resources (CLIR) report "Digital Forensics and Born-Digital Content in Cultural Heritage Collections" Kirschenbaum, Richard Ovenden and Gabriela Redwine highlight shared practices between disparate fields.[19]The report puts into practice some of Duranti and Elizabeth Diamond's theoretical work—which highlights the convergence of the distinct fields of forensic science, archival studies and diplomatics—by examining varying institutional practices regarding the management of digital records. Their practices, which constitute a general digital forensics workflow, hinge on the capture and retention of disk images, although other strategies such as emulation, migration, or using legacy media for rendering and access are put forth as viable, but less secure options.[20] The authors of the CLIR report echo previous assertions made about the authenticity and integrity that disk images afford (the capture of complete information, and the capability to understand relationships between data) and further state a bold reality: disk images also act as a container to retain complete data sources until a later time during which the archivist can deal with analysis and redaction.[21] By 2010, it was well-established within the field that disk images provide a more complete original presentation and documentation of the original digital object, while serving the practical function as a safeguard against time, labor shortages, and a lack of knowledge related to media-bound digital content. The disk image allows those of

[18] Leighton John, 50.

[19] Matthew G. Kirschenbaum, Matthew G., Richard Ovenden, and Gabriela Redwine, "Digital Forensics and Born-Digital Content in Cultural Heritage Collections," *Council on Library and Information Resources (CLIR) Reports* 149 (2010), 2.

[20] Kirschenbaum et al., 15, 21-22.

[21] Kirschenbaum et al., 26.

us without adequate resources to "take and keep it all" until a later, undefined date that may or may not arise.

The CLIR authors go on to discuss the nuances of managing personally identifiable information (PII) and sensitive data in digital form, noting the differences in security measures needed to ensure the safekeeping of sensitive information among paper-based and digital archives, the overall lack of specificity in current guidelines and recommendations, and the complications cloud-based personal information pose to privacy.[22] Inherent in their perspective is the known fact that risks and concerns regarding the identification and redaction of sensitive information have increased at a rate unequal to the rate of developments in digital forensics theory and practice, thus necessitating a range of workflows, policies, and procedures to mitigate associated risks. The risks related to mismanagement or disregard of sensitive information are both local and global, holding the potential to endanger record creators and expose record subjects' identities and behaviors, thus affecting the collecting institution's ability to generate and maintain trust.[23] We must therefore plan that, even with the best identification and redaction tool, we are likely to miss sensitive information. The authors emphasize the need to educate and engage donors, record creators, and researchers across the archival and forensic life cycle; working with data creators; and recruiting and allocating training and staff.[24]

By 2011, according to Kam Woods, Christopher A. Lee and Simson Garfinkel in their argument for greater institutional repository support for and increased access to disk images, the capturing and retention of disk images was "widely used today by practitioners of digital forensics [because] disk images can serve as baselines for comparison for digital preservation activities, as they provide fail-safe mechanisms when curatorial actions make unexpected changes to data; enable access to potentially valuable data that resides below the file system level; and provide options for future analysis."[25] Woods, Lee and Garfinkel also discuss the multifaceted ethical concerns related to disk images, though they advocate for diligent curation, care, and access restrictions, rather than deletion. Their justification is that the researcher, archivist, curator, or librarian can always return to the original disk image in order to demonstrate authenticity, allow for emulation or access, or to generate new access copies, and that this capacity and service to the user might outweigh other relationships with creators, subjects, or communities. It is then left to curators of digital materials to be "responsible for understanding and implementing plans to handle 'hidden' data."[26]

---

[22] Kirschenbaum et al., 48-49, 51-52, 54.
[23] Kirschenbaum et al., 51.
[24] Kirschenbaum et al., 50-64.
[25] Woods, Lee, and Garfinkel, "Extending Digital," 57.
[26] Woods, Lee, and Garfinkel, 64.

Our final motivation for obtaining and retaining a full disk image is the occasional interconnectedness of data across a full volume. By this, we mean content that may first appear to be contained within one easily extractable logical file, but is dependent on bits inscribed elsewhere on the media. Examples include pointer or resource files, proprietary software or databases that are environment-specific, and linked content in documents. John Durno, a digital archaeologist and head of library systems at the University of Victoria, writes that in some file systems, such as HFS+, "data in the resource fork sometimes extends beyond system-specific metadata including, for example, the embedded images in a word processing document."[27]Proprietary software or databases may also be dependent on specific environment configurations, or files or keys held elsewhere on the system. Finally, linked data within a single volume, but spanning multiple files poses similar challenges. For these reasons, a practitioner might be hesitant to let go of a full system disk image without first doing extensive testing.

RADICAL EMPATHY AND THE RETENTION OF DISK IMAGES

The reluctance to discard a disk image for evidentiary purposes or to enable a future ability to render content, along with the insistence that future custodians will appropriately handle this data, does not take into consideration relationships to and with creators, subjects, users, or communities. It takes advantage of "the different levels of power and privilege between creators, subjects, users and communities" and as a result can "reveal whose needs are served by different archival decisions"[28] because of differing access to technical capacity, thus employing knowledge as a weapon.

However, there is now a dilemma created by friction between the assumed wants of the user or community, the donor, the creator, and the subjects. Keeping a disk image may allow for greater access in the future, preserve more contextual information, or better demonstrate authenticity, but as a collection decision it may lack care for the privacy and safety of the creators or subjects of records. There is a tension here and our aim in this paper is to foreground that tension. On one hand, archivists have a mandate to preserve an authentic object, but on the other, archivists have a responsibility not to endanger those captured within that object.

Empathy as consideration for subjects and the navigation of the ensuing ethical dilemmas and questions are not new concepts for digital collections. As early as 1978, Alice Robbin, now a professor of library and information science and career expert on digital inequality and privacy, wrote about the difficulties of balancing an individual's privacy with researchers' access, and the challenges inherent to the collection of data

---

[27] John Durno, "Digital Archaeology and/or Forensics: Working with Floppy Disks from the 1980s," *The Code4Lib Journal*, no. 34 (October 2016).

[28] Caswell and Cifor, "From Human Rights," 195.

sets, rapidly evolving technology, and machine-readable formats.[29] In her work, Robbin laid out a set of ethical principles to guide data archivists. While her principles focus primarily on actions *after* selection, she nods to selection when she writes, "archives will have to reevaluate their acceptance of certain data if facilities and competence to protect the information are not available. The probability of increased costs, perhaps heavy, to an archive as a result of implementing these procedures cannot be ignored."[30] Here, Robbin argues the decision to collect is not just about the sensitivity of the content, but also our ability to both understand and protect it. Nine years later, Heather MacNeil, now a professor at the University of Toronto, published her thesis on disclosure of personal information, which later became the 1992 book, *Without Consent: The Ethics of Disclosing Personal Information in Public Archives*. MacNeil's work, like Robbin's, explored the topic of reconciling subject-researcher conflicts alongside the challenges raised by the rapidly increasing volume of digital records, the sometimes-obscured nature of their content, and their security.[31] MacNeil, like Robbin, focuses more on access than selection or criteria for destruction though she does not shy away from the topic, describing the ongoing privacy discussion as a professional obligation. In her work, she makes a series of recommendations describing how access requests might be reviewed and collection policies developed, and notes that ethical standards almost always give higher priority to the subject over the researcher or future access request.[32] She also encourages the ongoing questioning of technological imperatives that seemingly require the collection of so much personal information. MacNeil argued, over 25 years ago, that just because we are technically able to collect and store this information, it does not mean we should.[33]

In both Goldman and Pyatt's work on the privacy risks associated with disk imaging as well as Kirschenbaum, Ovenden, and Redwine's CLIR report discussed earlier, both sets of authors make recommendations to mitigate those risks. These include developing policies with more specific language about digital retention, engaging donors and creators at acquisition, encouraging donor self-appraisal and selection, developing preservation and management strategies, engaging IT staff in security and information risk management, developing institution-specific access methods, and engaging researchers on their responsibilities for privacy protection.[34]

---

[29] Alice Robbin, "Ethical Standards and Data Archives," *New Directions for Program Evaluation* 4 (1978): 7-18.

[30] Robbin, 12.

[31] Heather MacNeil, *Without Consent: The Ethics of Disclosing Personal Information in Public Archives* (Metuchen, NJ: Society of American Archivists and The Scarecrow Press, 1992).

[32] MacNeil, *Without Consent*, 185-201; MacNeil, 185.

[33] MacNeil, 199.

[34] Goldman and Pyatt, "Security," 47; Goldman and Pyatt, 46-51; Kirschenbaum et al., "Digital Forensics," 50-64.

While these recommendations are thoughtful and needed, we have two criticisms. The first is the absence of specific considerations for subjects and affected communities. The second is that, for Goldman and Pyatt, the work is rights-based, concerning itself with "content archivists are required to protect under most laws and regulations."[35] While regulations and policies may provide useful benchmarks, they are not sufficient; it may be legal to release someone's data, but this doesn't preclude it from causing harm. Archivists lately have increasingly pointed out and become more sensitive to the complex, prejudiced relationship between record creators and stewards and the subjects or communities they describe. Jarrett Drake illustrated the harmful influence of law enforcement on the creation of records meant to document a 2015 New Orleans police shooting, and Chelcie Juliet Rowell and Taryn Cooksey explored the implications of a digital publisher developing a collection of Ku Klux Klan newspapers in 2017.[36] We must not forget that multiple axes of recordkeeping—from its litigious nature to the terrorizing history that records may record, narrate, and demonstrate—are often at work against the interests of users, subjects, and adjacent or marginalized communities represented within the record.[37]

As authors, we were moved to consider whether we should recommend against keeping disk images at all, as though a universal definitive that did not consider context was required. We discarded the temptation of a universal judgement, recognizing that in radical empathy, absolutes are not required. According to Caswell and Cifor, "these relationships may conflict in dramatic ways if the needs of the creator are vastly different from those of their subjects, for instance, but using radical empathy can guide an archivist to the right questions to ask when processing a collection, ensuring the result is as fair and inclusive as possible."[38] In other words, a decision needs to be made, a dilemma must be resolved, and it is likely that decision can't be easily mapped on to a decision tree or automated workflow.

Arthur David Baynes-Cope arrives at a similar conclusion in his 1988 work, *Thoughts on Ethics in Archival Conservation*. Baynes-Cope argues that without choice and judgement on the part of the conservator, there can be no ethics, and that strict obedience to a workflow or a rigid set of best practices is self-defeating.[39] Baynes-Cope is

---

[35] Goldman and Pyatt, "Security," 38.

[36] Jarrett M. Drake, "Insurgent Citizens: The Manufacture of Police Records in Post-Katrina New Orleans and Its Implications for Human Rights," *Archival Science* 14, no. 3 (October 1, 2014): 376-378; Chelcie Juliet Rowell and Taryn Cooksey, "Archive of Hate: Ethics of Care in the Preservation of Ugly Histories," *Lady Science* (blog), January 19, 2019, https://www.ladyscience.com/blog/archive-of-hate-ethics-of-care-in-the-preservation-of-ugly-histories.

[37] Drake, "Insurgent Citizens," 366.

[38] Caswell and Cifor, "From Human Rights," 195.

[39] Baynes-Cope, "Thoughts on Ethics," 140.

basing his argument on the logic that repair or conservation actions cannot always be congruent with the conservation principles as laid out by Ellis and Jenkinson. However, the sentiment is still applicable: his idea that preservationists "should have the mental and moral provision to make moral, ethical judgements, and beyond this what can only be called the gift of doing what is right in the circumstances and doing it beautifully"[40] can easily be applied to digital forensics and preservation. In other words, while workflows and axioms can guide us and hasten work, room must always be left for choice.

Those choices, the thoughtful and empathetic decisions that Caswell and Cifor encourage as part of radical empathy in archives, do not lend themselves well to automation. They are slow decisions, ones that require space and time, and they are human decisions. They are subjective and contextual, and the questions change. They often don't have a default setting. We recognize this can be unsettling for those who prefer a more rigid and rules-based order.

SUCCESS CRITERIA FOR DISK IMAGE ANALYSIS AND REDACTION TOOLS WITHIN THE LENS OF RADICAL EMPATHY

Computer-aided analysis and redaction tools can support, though may not replace, the human decisions required when working with disk images. There are a range of open-source and proprietary tools to assist a librarian or archivist tasked with making empathetic decisions about these collections. Comparison of these tools is not our goal due to their disparity in functionality, audience, and purpose. Instead, we aim to explain each tool's function, and how it fulfills certain aspects of our success criteria for radical empathy. These criteria are:

- Open-source and thus transparent software, ideally developed or informed by the cultural heritage community;
- Clear documentation of the tool and its functionalities, capabilities, and limitations;
- Allows for transparent and thorough deletion of content as a direct result of analysis;
- Clear and useful reporting mechanisms;
- Does not retain any unwanted content, including in log files, past-versions of content, or metadata, or makes that retention transparent and editable; and
- Low technical barriers to entry.

[40] Baynes-Cope, "Thoughts on Ethics," 144.

The following are a sampling of popular tools and environments available to archivists to help them identify and make decisions about potentially sensitive information in disk images, and an analysis of how they meet these criteria.

Bulk_extractor

Bulk_extractor, developed by Simson Garfinkel,[41] is a forensics tool that can scan a disk image and look for particular types of content: strings that match a certain format or type, and metadata associated with certain image files and compressed files. It is very effective at identifying, for example, an email address, URL, or any specified string. It is not, however, as effective at helping an archivist evaluate a video or image, nor is it meant to act as a substitute for thoughtful and considerate evaluation of content. For example, according to bulk_extractor, the contents of a file might be benign due to absence of credit card numbers and Social Security Numbers (SSNs), but that doesn't mean those contents would not be harmful to its creator or others if kept. Bulk_extractor is designed to be an identification tool, not a redaction or deletion tool. Its purpose is to help find information, not remove it.

Archivematica

Archivematica is a free, open-source digital preservation system that automates the process of preparing digital objects for ingest into a repository and/or storage, and provides access to the archived material.[42] Archivematica uses bulk_extractor as its core application for identifying, analyzing, and reviewing personally identifiable information, such as SSNs, email addresses, and credit card numbers, but includes options for deletion of found content. Once an object is examined by bulk_extractor it must be sent to Backlog in the application, where the results can be viewed per file in the Appraisal tab. The bulk_extractor reports are divided into two categories, "PII" and "Credit card numbers". The interface also provides a preview pane for visual evaluation of certain file formats (e.g., .jpg files or .doc files). Archivematica's process transparency and high-level of user control (both behind the scenes in its configuration, and from the user-end) comes at a complexity cost. Archivematica will tell you what it is doing, but finding and deciphering that information requires time, knowledge of where logs are stored, knowledge of where discarded materials awaiting deletion are stored, and ultimately knowledge of the

---

[41] Simson Garfinkel, "Digital Media Triage with Bulk Data Analysis and Bulk_Extractor," *Computers and Security* 32 (2013): 56-72.

[42] Digital Curation Centre, "Archivematica," *External Resources*, http://www.dcc.ac.uk/resources/external/archivematica; Artefactual, "Archivematica Homepage," https://www.archivematica.org/en/.

infrastructure and system upon which it is installed. For example, a user may extract desired files from a disk image and delete the original disk image along with unwanted extracted files (this is part of the backlog process for storage and requires a dual-user authorization process). It is unclear though if all records of these files (e.g., filenames, file-level metadata) are wiped. If appraisal and report generation occur after a METS.xml document is generated on the initial transfer, but *before* a file is ultimately rejected, the residue of those files and their existence is difficult to remove.

BitCurator

The BitCurator Environment is an Ubuntu-derived Linux stack of free and open-source digital forensics tools and software libraries. The environment is designed to help collecting institutions triage, acquire, describe, identify, and analyze born-digital materials, incorporating software and practices adopted from the digital forensics community with a low barrier to entry. The environment groups tools and scripts under common steps in the digital forensics workflow: Forensic Disk Imaging, Forensic Processing and Identification of Potentially Sensitive Information, Data Triage, and Metadata Export. A number of tools allow users to view, browse, and analyze disk images (BitCurator Disk Image Access Tool, BitCurator Reporting Tool, fiwalk, bulk_extractor), identify and prioritize important information in or about disk images (ssdeep, sdhash, ClamAV, DFXML tools, FSlint), as well as additional scripts and tools that further support or enhance the described workflow steps (BitCurator Mounter, GHex, a hex editor/viewer, custom BitCurator tools and Nautilus scripts).

As an open-ended and modular environment, BitCurator provides a lot of space for contextual decision-making. Users can mount disk images and explore them manually and visually, and they can run a variety of reporting or analysis tools (such as bulk_extractor). That said, the goals of BitCurator (like those of forensics) are collection, analysis, identification, and documentation, not deletion.

Bulk Reviewer

Bulk Reviewer, an open-source tool, aids in the identification and review of sensitive information in disk images or logical files.[43] As an answer to Goldman and Pyatt's concerns about the prevalence of forensic methods, Tessa Walsh, began work on Bulk Reviewer in 2018 to help librarians and archivists better and more quickly review born-digital content. In a departure from law enforcement's traditional keyword or phrase searching tools, and recognizing what is required for empathetic collections decisions, Walsh envisioned a tool

[43] Tessa Walsh, "BulkReviewer-README," *Github*, Retrieved March 26, 2019 from https://github.com/bulk-reviewer/bulk-reviewer.

that assists the human review process, allows for annotation and discussion, and helps overcome technical obstacles to redacting content that is both visibly performed and sometimes hidden in places like file slack or headers.

Bulk Reviewer is still a work-in-progress, but the second production release (0.2.0) is available. Its functionality includes scanning for user-supplied regular expressions, keywords, names, and structured strings. Its evaluation of a disk image is primarily text-based (e.g., a photo's EXIF metadata can be evaluated, but not the photo itself). As with Archivematica, most of this functionality is provided by bulk_extractor, but user control and reporting are significantly expanded. BulkReviewer does not provide an option to redact within files (a difficult and technically unstable tactic), but it does provide decision support through exports that can be classified as "cleared" (found not to contain defined sensitive information) or "private" (files found to contain defined sensitive information). The reporting mechanisms are abundant and clear, and the active project is installed as a desktop application with installer scripts provided and extensive documentation.

Forensic ToolKit (FTK)

Forensic ToolKit (FTK) is a proprietary computer forensics software made by AccessData designed to "help law enforcement officials, corporate security, and IT professionals access and evaluate the evidentiary value of files, folders, and computers," therefore its features correspond with stages in the e-discovery process: identification, preservation, collection, processing, review, and production.[44] Functionality includes creation of disk images (e.g., of hard drives, floppy, CD/DVD, portable media such as USB drives, and/or live data from any common electronic source); identification, analysis, and redaction of deleted files and objects (data carving), contact information, and other data specified by the user; filter searching to quickly locate specific item types and/or exclude specific data from review and reporting; as well as decryption.[45] The filter feature allows users to search for keywords and specified data (e.g. email addresses, phone numbers) at a high level of granularity: predefined filters can target specific mobile and communication data such as email and addresses or signatures within them, phone history, calendar information, etc.

FTK presents high financial and technical barriers, as well as obstacles to understanding and terminology. First, the software costs a few thousand dollars without a support contract. Second, because the tool relies on distributed processing and allows for multiple installation methods, installation requires moderate technical understanding regarding database management. Language and metadata further pose an obstacle.

---

[44] AccessData, "Forensic ToolKit (FTK) Version 6.4 User Guide," *AccessData,* February 2018, https://ad-pdf.s3.amazonaws.com/ftk/6.4.x/FTK_UG.pdf.

[45] AccessData, "Forensic ToolKit (FTK)."

Language used in the FTK manual and user interface are drawn directly from law enforcement, automatically creating a barrier of knowledge and understanding related to terminology for users of the tool outside of those audiences. While features can be useful to those outside of the designated user group, use of the tool comes at an additional labor cost. Workflow steps and metadata fields must be mapped between FTK's terminology and those more prominently known in the archival and digital preservation communities (e.g., "evidence" and "case" vs. collection or accession, "custodian" vs. subjects, creator, or communities). Further, because recovery of deleted data is a core feature and the software's data model maps to e-discovery practices, the process of deletion and redaction are convoluted and lengthy, at times requiring the user to detach multiple records from each other in order to delete "evidence." Disk images and their contents are described and processed as evidence to be documented and retained rather than destroyed.

DISCUSSION

We do not yet have a tool that will do it all, nor should we expect to. Computer-aided analysis and deletion can support and assist the archivist's work. However, the identification of potentially harmful information and decisions on its inclusion need to consider the relationships to objects' creators, subjects, users, or communities. This requires labor, time, and empathy.

From our own collections, we pulled a floppy disk for preservation treatment that was collected over 20 years ago, and contained content that raised many of these questions. The disk's origin was as part of a submission to a magazine, whose editor was the actual donor. Contained within the deleted content were personal letters detailing experiences battling cancer. These letters did not contain credit card numbers or SINs/SSNs. They did not violate any laws. To keep them however, even buried in the invisible recesses of a disk image, would be wrong. These letters were not meant to be collected by the repository and they were not intended for a public audience. We made the decision to pull the one file (the author's original submission) that was part of the collection's focus, and delete the disk image. Internal discussion is still ongoing about whether the original media carrier should be disposed. Once an object makes its way into special collections, it can be difficult to remove. Discovering the letters and then making these decisions were all actions that required human labor and consideration.

Recommendations and Future Work

The following recommendations follow a feminist ethics of care, and they create space for the consideration of creators, subjects, users, communities, and other cultural

heritage professionals in our work beyond what is possible in a traditional, legalistic, rights-based framework. Our hope is to contribute to this subset of our field, helping ourselves and others make empathetic, thoughtful, and informed decisions about digital materials. To this end, beyond our success criteria for analysis and redaction tools, we include several recommendations for future work, both within our profession and field of scholarship. We value the recommendations put forth by Goldman, Pyatt, Kirschenbaum, Ovenden, and Redwine, which include more specific policy language, donor engagement at the time of acquisition, self-appraisal, improving management strategies, engaging IT staff, developing institution-specific access methods, collecting stories and use cases, defining requirements for the development of new tools, developing regional collaboration networks, and educating researchers on their responsibilities.[46] In considering radical empathy, we would also like to expand the list of recommendations to *add* the following.

Allocate Staff, Training, and Time Needed to Review Born-Digital Content

Appropriate provision of staff training and allocation of time is required to fully review all content, before the decision to keep it. When possible, additional staff to review born-digital content should be a condition of acquisition, particularly for collections with high-risk or sensitive information. As demonstrated, current technical solutions to scan for sensitive information are aids, not replacements. These tools are also incapable of making nuanced decisions about retention, destruction, or redaction.

Collect Less

As Alice Robbin remarked in 1978, the increased costs of digital collections are both probable and heavy. If those costs cannot be covered, it is important to consider that the archive may be unable to accept the materials. As we have shown, it is nearly impossible to know with precision what exists on born-digital media-bound carriers without rendering that content. Therefore, more often than not the appraisal and review process for born-digital material occurs *after* the archive acquires these materials and *while* the media is processed for preservation and access. Unless the acquisition process can be amended to better appraise born-digital materials, archivists and curators must explicitly limit acquisitions of born-digital materials in accordance with the archive's ability to ethically steward these materials.

[46] Goldman and Pyatt, "Security," 46-51; Kirschenbaum et al., "Digital Forensics," 62-64.

Consider the Subjects and Affected Communities of Born-Digital Materials, and Their Presence in Those Materials

Goldman and Pyatt, as well as Kirschenbaum, Ovenden, and Redwine, include consideration of and engagement with *donors* in their recommendations. The authors of the 2010 CLIR report also emphasize the critical need to engage *researchers* in discussions around their responsibilities to protect sensitive information. However, none explicitly mention the *subjects*, or adjacent and marginalized communities represented in the archival materials. In a feminist approach, the steward also cares for and considers subjects and communities when making archival decisions.[47]

Define Acceptable Risk for Retention of Sensitive Data and PII

Goldman and Pyatt called on the digital preservation community to begin "collectively attempting to define what constitutes acceptable risk when it comes to born-digital materials" to mitigate and manage risk.[48] As we have demonstrated in this paper, the barriers for thoroughly and thoughtfully reviewing sensitive information are numerous. Therefore we, as community members and staff at individual collecting institutions, must begin an iterative staged approach for tackling the issues that sensitive data and PII present. We can begin by creating classes of the sensitive data that exists within and across our collections, then prioritizing those risks, in an initial attempt to manage risk. By first defining what sensitive data exists in or across collections, what risks these data pose, and determining the priority of data or collections to review and redact, we can better plan future resourcing and staff labor for both the short and the long term. This work can also inform further decisions to acquire or accept future digital materials. As Robbins stated, we must consider not only the sensitivity of the materials but our ability to understand and manage that data as part of our decision to collect.

Generate Data and Documentation on Processing Time and Staff Required to Review Sensitive Information

Building on Kirschenbaum, Ovenden, and Redwine's recommendation to gather use cases, we believe it is critical for staff to document their identification, reviewal and redaction activities in order to generate appropriate allocation of resources and labor; as well as identify challenges and risks posed to staff, collecting institutions, record creators, users, subjects, and other third parties. Even though predictions or understanding of the process may fluctuate among collections and materials, it is more prudent to gather use

[47] Caswell and Cifor, "From Human Rights," 36.
[48] Goldman and Pyatt, "Security," 52.

cases, lessons learned, and data in order to give insight into the process of managing sensitive personal information.

Use Tools Developed and Informed by Libraries, Archives, and Museums

Archivists, curators, and other stewards of born-digital cultural heritage should develop or aim to use tools developed and informed by their own communities and subjects. When we use tools primarily designed for a different "type" of custodian, we create more labor that must be performed for outside of the tool in order for the digital objects to be properly stewarded in a way that aligns with archival and preservation best practices. We also risk fundamentally altering our perspective. Take, for example, a Twitter discussion in early 2019 on the use of equipment and software from vendors primarily serving legal enforcement. The discussion began with Eddy Colloton, a Project Conservator at the Hirshhorn Museum, noting, "Lots of GLAM folks (including me) using tools from vendors that primarily service law enforcement. Thinking about forensic bridges and disk imaging software especially. Kinda fucked up right?"[49] The subsequent conversation drew heavily on Elvia Arroyo-Ramirez, Kelly Bolding, Faith Charlton, and Allison Hughes' work at Princeton as described in, *"Tell Us about Your Digital Archives Workstation": A Survey and Case Study*. They write, "the company that manufactures FRED machines, Digital Intelligence, has a primary customer base not of cultural heritage institutions but law enforcement agencies. The sobering fact that our purchase would indirectly help support the tools of the criminal justice system, and by extension, the prison industrial complex, caused some unease."[50] From this, Elizabeth England, an archivist, stated a desire for functionality in the open-source disk imaging tool, Guymager, to edit or use different metadata fields as opposed to the preset fields, "evidence number" and "examiner."[51] Within 24 hours, Euan Cochrane, a digital preservationist, and Guy Voncken, Guymager's developer, had created a solution to change the fields.[52] This interaction perfectly highlights the need for archivists, curators, and other stewards of born-digital materials to have access and input into the development of software and tools utilized in their work.

---

[49] Eddy Colloton, Twitter Post, March 21, 2019, 3:36pm, https://twitter.com/EddyColloton/status/1108814751281360896.

[50] Elvia Arroyo-Ramirez, Kelly Bolding, Faith Charlton, and Allison Hughes, "Tell Us About Your Digital Achives Workstation: A Survey and Case Study," *Journal of Contemporary Archival Studies* 5 (2018): 9.

[51] Elizabeth England, Twitter Post, March 21, 2019, 4:00pm, https://twitter.com/elizabeengland/status/1108820635650793473.

[52] Guymager Ticket #13 as submitted by Euan Cochrane, "Add Ability to Define Metadata Field Names Through Settings," SourceForge, March 25, 2019, https://sourceforge.net/p/guymager/feature-requests/13/.

CONCLUSION

Given the number and range of hurdles surrounding the management of sensitive data, we must maintain a critical eye regarding the theory and practices our digital archival work depends on, particularly with respect to personal archives. Archival science, traditional recordkeeping, and digital forensics and preservation best practices hold that the disk image is the gold standard preservation format for born-digital materials. The disk image engenders trust, authenticity, and integrity but also potentially discloses a multitude of hidden and sensitive information to the recordkeeper, who must decide on the proper course for review and redaction (or not). The disk image also serves as a deferral method, a way to hold content in stasis until some future date when a more thorough review of its contents can be completed. For some, this date might be triggered by an access request. For others, it may be when more resources, like staff time, become available. In some cases, it may not come at all.

Radical empathy lends a feminist intersectional lens to digital forensics practices, which originate in law enforcement. Rather than focusing on the gathering and retention of evidence, it asks us to center human experience in archives, and consider all users, creators, subjects, and communities that the existing records affect. Here, radical empathy points to a lack of care and ethical accountability on the part of the archivist in managing sensitive personal information. Building upon recent work, case studies, and understanding of sensitive information in born-digital archival material, we recognize a need to hold ourselves accountable to creators, users, subjects, and marginalized communities; define the risks data and our workflows pose towards those groups; and either allocate resources, training, and human labor to the management of sensitive data or collect what we can appropriately manage.

BIBLIOGRAPHY

AccessData. "Forensic ToolKit (FTK) Version 6.4 User Guide," February 2018. https://ad-pdf.s3.amazonaws.com/ftk/6.4.x/FTK_UG.pdf.

"Archivematica," External Resources, Digital Curation Centre. Last modified July 24, 2015, http://www.dcc.ac.uk/resources/external/archivematica.

Arroyo-Ramirez, Elvia, Kelly Bolding, Faith Charlton, and Allison Hughes. "Tell Us About Your Digital Archives Workstation: A Survey and Case Study." *Journal of Contemporary Archival Studies* 5, article 16 (2018). https://elischolar.library.yale.edu/jcas/vol5/iss1/16.

Baynes-Cope, David. "Thoughts on Ethics in Archival Conservation," *Restaurator. International Journal for the Preservation of Library and Archival Material* 9, no. 3 (1988): 136-146. https://doi.org/10.1515/rest.1988.9.3.136.

Baynes-Cope, David. "Principles and Ethics in Archival Repair and Archival Conservation, Part 1: The Principles of Archival Repair and of Archival Conservation," *Journal of The Society of Archivists* 15 (1994): 17-26. https://doi.org/10.1080/00379819409511727.

Caswell, Michelle and Marika Cifor. "From Human Rights to Feminist Ethics: Radical Empathy in the Archives," *Archivaria* 81 (Spring 2016): 23-43. https://archivaria.ca/archivar/index.php/archivaria/article/view/13557.

Colloton, Eddy. Twitter Post. March 21, 2019. 3:36pm. https://twitter.com/EddyColloton/status/1108814751281360896.

Drake, Jarrett M. "Insurgent Citizens: The Manufacture of Police Records in Post-Katrina New Orleans and Its Implications for Human Rights." *Archival Science* 14, no. 3 (October 1, 2014): 365–80. https://doi.org/10.1007/s10502-014-9224-2.

Duranti, Luciana. "From Digital Diplomatics to Digital Records Forensics." *Archivaria* 68, (Fall 2009): 39–66. https://archivaria.ca/index.php/archivaria/article/view/13229/14548.

Duranti, Luciana. "Diplomatics: New Uses for an Old Science, Part I." *Archivaria* 28 (Summer 1989): 7–27. https://archivaria.ca/archivar/index.php/archivaria/article/view/11567/12513.

Durno, John, and University of Victoria Libraries. "Digital Archaeology and/or Forensics: Working with Floppy Disks from the 1980s." The Code4Lib Journal, no. 34 (October 2016). https://journal.code4lib.org/articles/11986.

Ellis, Roger. *The Principles of Archive Repair: A Paper Read by Roger Ellis*. London, UK: London School of Printing and Graphic Arts, 1951.

England, Elizabeth. Twitter Post. March 21, 2019, 4:00pm. https://twitter.com/elizabeengland/status/1108820635650793473.

Garfinkel, Simson. "Digital Media Triage with Bulk Data Analysis and bulk_extractor." *Computers and Security* 32 (2013): 56-72. https://simson.net/clips/academic/2013.COSE.bulk_extractor.pdf.

Goldman, Ben and Timothy D. Pyatt. "Security Without Obscurity: Managing Personally Identifiable Information in Born-Digital Archives," *Library and Archival Security* 26, no. 1-2 (2013): 37-55. https://doi.org/10.1080/01960075.2014.913966.

Guymager Ticket #13 as submitted by Euan Cochrane. "Add Ability to Define Metadata Field Names Through Settings." SourceForge. March 25, 2019. https://sourceforge.net/p/guymager/feature-requests/13/.

Jenkinson, Hilary. *A Manual of Archive Administration Including the Problems of War Archives and Archive Making.* Oxford, UK: The Clarendon Press, 1922. https://archive.org/details/manualofarchivea00jenk/page/n5.

John, Jeremy Leighton. "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, London* (September 2008): 48-55. https://phaidra.univie.ac.at/o:294101.

Kirschenbaum, Matthew G. *Mechanisms: New Media and the Forensic Imagination*. Cambridge, MA: The MIT Press, 2008.

Kirschenbaum, Matthew G., Richard Ovenden, and Gabriela Redwine. "Digital Forensics and Born-Digital Content in Cultural Heritage Collections," *Council on Library and Information Resources (CLIR) Reports* 149 (2010). https://www.clir.org/pubs/reports/pub149.

MacNeil, Heather. *Without Consent: The Ethics of Disclosing Personal Information in Public Archives.* Metuchen, NJ: Society of American Archivists and The Scarecrow Press, 1992.

MacNeil, Heather, Chen Wei, Luciana Duranti, Anne Gilliland-Swetland, Maria Guercio, Yvette Hackett, Babak Hamidzadeh, et al. "Part 1: Establishing and Maintaining Trust in Electronic Records – Authenticity Task Force Report." In *The Long-Term Preservation of Authentic Electronic Records*. 2001. http://www.interpares.org/book/interpares_book_d_part1.pdf.

Moravec, Michelle. "Feminist Research Practices and Digital Archives." *Australian Feminist Studies* 32, no. 91-92 (April 3, 2017): 186-201. https://doi.org/10.1080/08164649.2017.1357006.

Robbin, Alice. “Ethical Standards and Data Archives,” *New Directions for Program Evaluation* 4 (1978): 7-18. https://doi.org/10.1002/ev.1222.

Rowell, Chelcie Juliet, and Taryn Cooksey. “Archive of Hate: Ethics of Care in the Preservation of Ugly Histories.” *Lady Science* (blog), January 19, 2019. https://www.ladyscience.com/blog/archive-of-hate-ethics-of-care-in-the-preservation-of-ugly-histories.

Stapleton, Richard. "Jenkinson and Schellenberg: A Comparison," *Archivaria* 17 (1983): 75-85. https://archivaria.ca/index.php/archivaria/article/view/11021/11956.

Summers, A., and John, J. L. “The W. D. Hamilton Archive at the British Library.” *Ethology, Ecology & Evolution* 13 (2001): 373-384.

Walsh, Tessa. “Harvard Library Innovation Lab Fellowship, Part One: What We Talk About When We Talk About PII.” Last modified July 16, 2018. https://www.bitarchivist.net/blog/2018-07-16-lil-part-one.

Walsh, Tessa. “BulkReviewer-README.MD.” *Github*. Retrieved March 26, 2019 from https://github.com/bulk-reviewer/bulk-reviewer.

Woods, Kam, Christopher A. Lee, and Simson Garfinkel. “Extending Digital Repository Architectures to Support Disk Image Preservation and Access.” In *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries - JCDL 11*, 57-66. New York: Association of Computing Machinery, June 2011. https://doi.org/10.1145/1998076.1998088.