Ethics and digital collections: a selective overview of evolving complexities

David Mindel
Library Department, University of Wisconsin-La Crosse, La Crosse, Wisconsin, USA

Abstract
Purpose – Digital collections are becoming more commonplace at libraries, archives and museums around the world, creating potential for improved accessibility to information that may otherwise remain hidden and further support for intellectual exploration. As a result of the growing potential for digital collections to inform and influence, the conversation surrounding ethics and digital collections needs to be continually examined and adapted as technologies evolve, user expectations change and digital information plays an increasing role in our everyday lives. In this context, this paper presents an overview of multifaceted ethical realities that impact the how, why and what digital information is created, accessed and preserved.

Design/methodology/approach – Written from the perspective of a digital collections librarian, this paper relies on existing research in presenting ethical considerations and complements that research with professional observations in providing subsequent reflections on addressing challenges in the age of digital information.

Findings – There are and should be considerations given to not only what information is contained in a given collection, but also how that information is selected, accessed and consumed by the public. The conclusions offered are designed to provoke reflection on the evolving and interconnected nature of information and ethics in the context of digital collections.

Originality/value – Information ethics is multifaceted, with one of those facets relating directly to digital collections. This paper demonstrates that digital collections are more complex than simply a collection of digitized documents and photographs. As the field of information management continually evolves and adapts, so, too, do the ethical realizations identified in this paper, all of which go beyond the (virtual) walls of a library, archive or museum, and carry the potential to have a long-term impact concerning information and its integrity, equity and access.

Keywords Digital collections, Ethics, Image manipulation, Digitization, Born-digital, Digital divide, Digital literacy, Digital infrastructure, Information equity

Paper type Viewpoint

Introduction
Ethics in information management is arguably more important now than ever, specifically in regard to the continued rise of information in digital format. Of course, digital information is not a recent introduction to the world of information management. Information that can be effectively used and consumed by the public has been around since the mid-1990s, although technology leading up to that time had been under development since the 1960s (Campbell-Kelly and Garcia-Swartz, 2013, pp. 18–19). It is the rate of use and consumption of this information that by necessity creates a major facet in information management that focuses on the general ethics surrounding the production, dissemination, use and preservation of information in digital formats. These areas are of particular relevance within the subset of digital collections as they relate to the general field of information management.

As the demand for digital resources, specifically digital collections, continues to increase (Mills, 2015, p. 162), heritage, academic, and governmental institutions attempt to keep pace with expectations for the “extensive online access to...unique...primary sources contained in institutional collections” (Bantin and Agne, 2010, p. 244). This motivation, paired with an improved financial feasibility for digital collections infrastructure, including equipment (Terras, 2015, p. 69), has great potential for preserving information, disseminating that
information, and creating a more informed public now and in the future. However, such growth and potential should maintain adherence to central pillars guiding responsible digital information management. This paper brings together obstacles that require an appreciation of the ethics intertwined with building and maintaining digital collections and provides responses to these obstacles for those managing digital collections.

This paper discusses various elements that relate directly to the work of libraries, archives, museums and other cultural heritage institutions to make available and preserve digital information. As such, the term digital collection(s) is used to express this concept. Although there are many sources for library-related definitions, the term digital collection(s) is used synonymously with the term digital library, or “a library in which a significant proportion of the resources are available in machine-readable format...accessible by means of computers,” as defined by Reitz (2004). This paper also considers electronic resources to differ from materials that are part of a digital collection, in that, from a cultural heritage and open access perspective, the author believes content within a digital collection should be free for the public to access and use, independent of authentication, licensing fees and subscriptions. Although open access journals create some ambiguity to this definition, in this paper, items such as e-journals, e-books and general databases that are not open and free are excluded from what is considered to be typical digital collection content. Additionally, institutional repositories may be defined as databases of “scholarly materials in digital format created by [an] institution” and are often open and free (Reitz, 2004). As such, institutional repository is being included as a subset of the more general digital collection(s) term. To further narrow and clarify the scope of what is considered under the term digital collection for the purpose of this paper, attention is given to digitized archival material that comes from an archives or special collections library as well as discussion related to born-digital information.

Concerning the integrity, equity and accuracy of a digital collection, or more specifically, of the information contained within, this paper suggests that these three concepts are central forces driving digital collection ethics. This paper understands integrity as the content or information represented by a digital image being “the same as at the time of its acquisition” (Korus, 2017, p. 2), whereas equity in digital collections parallels the larger movement of diversity, equity and inclusion in libraries and their collections. For this discussion, equity is more specifically referring to cultural equity aimed at providing representation, preservation and dissemination of information related to, and produced by, historically underrepresented and oppressed populations. For example, in 1985, folklorist Alan Lomax expanded upon his 1972 “Appeal for Cultural Equity” (Lomax, 1972), in which he identified the cultural consequences of “[a] mismanaged, over-centralized electronic communication system...imposing a few standardized, mass-produced, and cheapened cultures everywhere” (Lomax, 1985, p. 40). Lomax went on to suggest that:

All cultures need their fair share of the air-time. When...[people] hear or view their own traditions in the big media, projected with the authority generally reserved for the output of large urban centers, and when they hear their traditions taught to their own children, something magical occurs. They see that their expressive style is as good as that of others, and, if they have equal communicational facilities, they will continue it. (Lomax, 1985, p. 43)

Although Lomax was specifically referring to the “preservation, research, and dissemination of the world’s traditional music and dance” (The Association for Cultural Equity, 2021), his sentiments resonate in the context of equity in the broader sense of digital collections, some of which do contain those elements targeted by Lomax. Finally, this paper understands accuracy to relate to the digital rendering providing a true-to-life (as possible) representation of analog materials in a 2-D environment as well as the display of those renderings and born-digital materials.
Background
As the field of digital humanities broadens and the discipline of library and information science evolves, and as global interconnectivity expands and digital isolation at times seems unattainable, the multifaceted concept of information management requires that it be informed by ethical considerations if the outcome is to produce a world that is a reflection of the quality of the information that it consumes.

Digital information ethics is by no means a small or recent concept. Rafael Capurro identifies Norbert Wiener’s, 1950 work and Joseph Weizenbaum’s 1976 work, as foundational in establishing the combined concept of ethics and computer technology (Capurro, 2009; Weizenbaum, 1976; Wiener, 1950). Capurro also discusses issues related to how society and the environment are impacted by digital technologies, including “privacy, information overload, Internet addiction, digital divide, surveillance and robotics particularly from an intercultural perspective” (Capurro, 2009). More specifically, the niche areas of digitization, digital collections and cultural heritage follow similar thought and give rise to questions regarding the ethics of selection, privacy, copyright, preservation and access, all of which are discussed in Cokie Anderson’s book, Ethical Decision Making for Digital Libraries (Anderson, 2006). These themes revolve around the core idea that “[l]ibrarians have long been the guardians of knowledge, striving to collect, protect, and preserve the intellectual output of civilization” (Anderson, 2006, p. 81). This guardianship requires adaptation to meet not only existing challenges, but challenges that have yet to be faced. Peterson Brink, Mary Ellen Ducey and Elizabeth Lorang debated the inclusion of the historical controversial newspaper, Awgwan, in the University of Nebraska–Lincoln’s digital collections, demonstrating the need for ethics to evolve when contemplating the balance between unpleasant historical events or publications and values related to information access (Brink et al., 2016, p. 22). In 2004, Johannes Britz and Peter Lor discussed “communal and individual information rights” for digitized cultural heritage materials in Africa, and how people of those represented cultures need protections and assurances for privacy and access (Britz and Lor, 2004, p. 219). Over a decade later, Zinaida Manžuch presented ideas related to the defining of cultural heritage from a Western perspective, both in relation to digitization selection and metadata creation, while also discussing elements of heritage objects as “sacred or secret” and balancing access with respect for worldviews and traditions (Manžuch, 2017, p. 5). Alternatively, Anthony Cocciolo proposes email as an emerging form of cultural heritage in that email has “the potential to capture the emergence of thoughts and decisions that illustrate how and why things developed the way they did” (Cocciolo, 2016, p. 89). Such a proposal, of course, demands that institutions develop broader practices related to born-digital institutional archiving, which includes the access to, and selection and preservation of, “born-digital documentation created by an institution with historic and legal value” (Cocciolo, 2014, pp. 238–239). This emphasis on born-digital material and its preservation and access relies on assumptions of infrastructure and capacity, related to both an institution performing those tasks and the general public benefiting from those efforts.

As early as 1996 Dawn Mercedes wrote of the host of problems that comes with the widespread use of digital media, including the potential for misuse (Mercedes, 1996, p. 44). Dino Brugioni and Mia Fineman provide two separate works discussing the history of photographic manipulation, both focusing on pre-digital photography (Brugioni, 1999; Fineman, 2012). The significance of Brugioni’s and Fineman’s works is that they demonstrate the long history of photographic manipulation, well before the days of digital cameras and Photoshop. Today, digital cameras are often combined with mobile phones, and techniques for digital photographic manipulation can easily be obtained on YouTube or other Internet forums.

Technology, however seemingly widespread, is not guaranteed for many communities around the world. From Marshall McLuhan’s mention of an “Age of Information and
Communication” in 1964 (McLuhan, 1964, p. 248) to Anthony Normore and Antonia Issa Lahera’s 2019 Crossing the bridge of the digital divide: a walk with global leaders (Normore and Issa Lahera, 2019), a plethora of evolving literature explores issues related to the global access and use of information in today’s digital world, and how technology has brought a new complexity to the world’s digital divides.

Alternatively, in societies that are equipped with technological infrastructure, the evolving nature of the broader digital data has brought about its own complexities surrounding ethics. Lambér Royakkers, Jelte Timmer, Linda Kool and Rinie van Est explore the topic of data ethics, building on the idea that “digitization penetrates every aspect of our lives” (Royakkers et al., 2018, p. 127). This concept of digitization, however, goes far beyond the idea of converting historical materials into digital files and refers more generally to the presence of “digital” throughout daily life. This includes:

[T]echnology [nestling] itself in us (for example, through brain implants), between us (through social media like Facebook), knows more and more about us (via big data and techniques such as emotion recognition), and is continually learning to behave more like us (robots and software exhibit intelligent behavior and can mimic emotions) (Royakkers et al., 2018, p. 127).

Part of digital daily life no doubt revolves around the search for information using digital technologies. The reality is, as pointed out by Safiya Umoja Noble, that in contrast to the early Internet’s reliance on “human-labor curation,” today’s Internet employs “complex algorithms of machines to make selections and prioritize results for users” (Noble, 2018, p. 25). Adding to the complexity is that this system of searching is a cyclical process that both “informs and is informed in part by users” (Noble, 2018, p. 25), further catering to dominant nations, cultures and beliefs. Of course to say that Google, for instance, is the first information system to magnify societal misrepresentations would be an injustice to traditional library cataloging systems, which have long demonstrated “how women, Black people, Asian Americans, Jewish people, or the Roma, as ‘the other,’ have all suffered from the insults of misrepresentation and derision in the Library of Congress Subject Headings (LCSH) or through the Dewey Decimal System” (Noble, 2018, p. 24). Together with the reach and influence of the Internet and big data, misrepresentations such as these “maintain and exacerbate social relationships” and thereby serve to “[maintain] racial and gender subjugation” (Noble, 2018, p. 31). Noble goes on to state that “[a]ccess to high-quality information, from journalism to research, is essential to a healthy and viable democracy” (Noble, 2018, p. 153). This suggests that society needs both information access and quality information, and simply the presence of one or the other serves as a major obstacle in creating an informed and information literate society.

**Digital collections and content selection**

By the 2000s, collection development policies began being applied more commonly to digital collections (Xie and Matusiak, 2016, p. 37), but materials selected for digitization frequently revolve around the accessibility of special collections and archival materials (Mills, 2015, p. 161). Therefore, early digital collections may have existed under separate umbrellas for collection development from general library items. Those items selected for digitization, as part of a library’s existing collection, need not be purchased, thereby excluding them from traditional collection development budgets. Even so, digital collections, being less bound by traditional collection development budgets but still under the guidance of an umbrella collection development policy, require that some institutions’ collection policies “adapt and reflect the changes in the environment” (Xie and Matusiak, 2016, p. 40) as digital collections become more widespread. Other digital resources, such as e-journals, may comprise elements of a digital library that are governed by a collection development policy that factors in costs...
associated with subscriptions and licensing of e-materials. Typical costs associated with building digital collections, although potentially high, are often dependent upon infrastructure and staffing considerations. By the nature of the items being digitized, digital collections are able to mirror a library’s physical collection with the potential to increase access through its digital dissemination. Digitization therefore provides an additional gateway for accessing library materials, and even can attract donations, with donor expectations of digital accessibility (Bantin and Agne, 2010, p. 246).

By extension, the cliché, “history is written by the victors,” has truth that is applicable to digital collection development and its impact on the historical narrative. Defined by the United Nations Educational, Scientific and Cultural Organization (UNESCO), culture heritage is “the legacy of physical artefacts and intangible attributes of a group or society that are inherited from past generations, maintained in the present and bestowed for the benefit of future generations” (UNESCO, 2017). Using this definition, it is easy to understand how a people’s legacy that is not maintained in the present and bestowed for future generations will face the great likelihood that it may cease to exist for those future generations. Similarly, born-digital publications, some of which possess intangible attributes of a group or society, legacies yet to be realized, or are part of a collective knowledge base, are all susceptible to being lost to history if not captured, maintained and passed on. The significant difference between the analog-to-digital materials and born-digital materials is that the former already has a physical presence in a library (independent of being or not being digitized), whereas the latter, if not captured, may cease to exist over time. So the question arises as to whether a library, as a repository, is ethically obligated to ensure born-digital content is captured, preserved and disseminated.

Of course, responses to this question may focus more on the basic plausibility of taking action than acting upon the obligation itself. Certainly, many working in libraries and archives are familiar with tight budgets and the seemingly endless need to produce more with less. The theoretical prospect that the average library will be ethically obligated to capture, preserve and disseminate all born-digital materials that pertain to the institution and its community is not only daunting, but simply unrealistic. Therefore, the issue shifts from upholding an impractical ideal to implementing a practical standard that allows heritage institutions to strike a balance between working within their means and not sacrificing the ethical responsibility to preserve some “information heritage,” despite the inability to preserve it all.

From an ethics’ perspective, the selection of library information is identified as one of four key areas of significant influence or control performed by those in the library profession, in addition to information organization, preservation and dissemination (“Code of Ethics of the American Library Association”, 2008). Of course, digital collections are supported through all four areas, with not one being more important than another. However, it is the selection of materials to include in a digital collection that serves as a threshold that information must meet in order to be organized, preserved and disseminated. Preservation and dissemination may drive the selection, but it is the act of selection itself that imposes standards on information and those evaluating that information to ensure that it is of “genuine merit and...utility” (Cotton des Houssayes, 1906, p. 43). At the same time, e-resources provide the potential for collection development to be more directly patron-driven, such that, for example, e-book records loaded into a catalog are purchased when users access the resource. Although not used or endorsed by all libraries as a selection method (Johnson, 2013, p. 48), this concept introduces the potential to shift collection building from a solely educated and informed decision from within the field of information science to a process that links library users more directly to the collection development. This diversity or evolution of approach to collection development, however, generally lacks applicability to digital collections.
This is not to suggest that patron-driven digital collection development does not exist, because it does, occurring, for example, as researchers request certain resources to be digitized and made accessible online (Novara, 2010, p. 166). However, this patron-driven approach to building a digital collection is limited in comparison to typical e-resource patron-driven collection development, such as the e-book example above. What this reinforces is the significance of collection building for digital collections residing with a librarian or equivalent staff member who is familiar with the physical and electronic collections, as well as campus curricula for educational institutions, in order to make collection decisions. As is true with traditional collection development, the decisions made by information professionals do not guarantee that a collection will be free from bias during the selection process. Areas of specialty and interest among librarians should be cultivated and embraced, but caution should be exercised that they not “overshadow their collection development responsibility” (Morrisey, 2008, p. 167). At the same time, digital collections may require time to develop themes or focuses through student and faculty requests, digitization for preservation and the identification of content that reinforces campus and library strengths. Since physical materials have previously undergone the selection process, and patron-driven collection development is limited to specific requests, digital collection development can pose difficulties in producing a balanced collection that reflects both historically and culturally significant materials, as well as materials that enable students, faculty and community to advance their own knowledge. A digital collection that is deficient of a balanced representation of a library’s materials deprives those using that collection of accessing information that is independent of “private interests” and “personal convictions,” and deprives users of assurance that censorship is combated and privacy rights are upheld (Morrisey, 2008, pp. 165, 167).

Developing a digital collection still requires that it reflects and serves its stakeholders’ needs and wants while balancing general access to, and preservation of, the materials (Mills, 2015, p. 161). At the same time, the “choice to present certain information while leaving others behind is a decision easily impacted by politics, cultural preferences, and social values” (Ma, 2020, p. 275).

As with any library collection, a digital collection’s content has the ability to promote, support, and guide intellectual exploration by the materials included in the collection, and, through assessment, enables “institutions to identify items or collections of value to patrons, and consequently, determine user priorities” (Mills, 2015, p. 163). Such influence requires particular consideration that, at a minimum, echoes the purpose and implementation of traditional collection development policies to ensure the impact of a digital collection that supports and reinforces critical thinking with little or no bearing on the direction that thinking will take.

**Discovery, access and ethics**

Although the images that make up a digital collection are no doubt a key source of information, the metadata, in particular the descriptive metadata, enables discovery while both supplementing and contextualizing the content of the images. Despite a picture being worth a thousand words, without its metadata, the potential to inform is significantly hindered. For these reasons, metadata demands attention as a significant area, one that is explicitly tied to language. As such, and given common ambiguities of one language, let alone the potential for information to be lost or misconstrued in translation, metadata presents ethical challenges in ensuring not only accurate, equitable and current language is being used, but also that that language comprises a description that also meets those criteria.

The significance of language is made apparent when investigating library classification systems, a non-narrative pillar of metadata that enables access and understanding. However,
as Jens-Erik Mai describes worldviews in the context of library classification, “people come to the world from different perspectives, view the world differently, and understand the world differently” (Mai, 2013, p. 243), and therefore there is no universal understanding of the world. This serves as a significant realization in the realm of library classification because it creates a ripple effect on the historically Western-dominant hold on classifications (Whaanga et al., 2015, p. 527). It forces the recognition that, among other subjects, race, religion, gender or sexual orientation are not universally compartmentalized nor are they necessarily able to be compartmentalized and that a greater consideration of what and to whom it means to self-identify within any or all of those areas cannot be assumed. What is more is that societal norms, trends and taboos, and language itself change over time. Mai insists that the flawed concepts of “an all-knowing authority independent of time and place” and “cultural relativistic views” need to be removed and instead replaced with a framework based on Aristotelian logic from which one can be “objectively right or wrong to call something good, because that is a matter of fact within a specific tradition, purpose, or function” (Mai, 2013, p. 250). Following this line of thought equips libraries to make classification decisions “within the standards, norms, and authority of the practice” thereby offering the potential to “regain the authority [libraries] have sacrificed in the name of presumed neutrality” (Mai, 2013, p. 251).

This area of ethical consideration also extends into an equally complex landscape when examining indigenous knowledge. Metadata have the power to facilitate discoverability and, as a result, access, in the sense that the information to which the metadata pertain can be found. The next level of access for digital collections is the visual or auditory level of interaction (noting that this presents additional layers of accessibility obstacles for communities of differently-abled as it relates to sight and hearing, making both discoverability and access rely solely on the quality and granularity of the metadata). Manzuch presents a thorough investigation of the ethical challenges involved specifically with digitizing cultural heritage, and brings to light memory institution biases, including how cultural heritage is interpreted and selected for digitization; the inadequacy of applying “metadata schemas that the ‘Western’ world commonly uses to describe and arrange indigenous heritage;” and how ethically problematic it is to create and provide digital access to items that are sacred, secret, or of particular significance to a community (Manzuch, 2017, pp. 4–5). In the context of the last point, access is no longer concerned about simply making digital objects available online, but instead the concern is shifted to the ethics surrounding whether or how an object should be made available online, if at all.

Discovery and access are more complex than simply describing an object, assigning subject headings and publishing the item online. In emphasizing cultural equity, assumptions about the desire for dissemination need to be questioned. The quality of metadata needs to be a focal point, in the sense that subject headings and descriptions must be informed by the origins of the object in the context of cultural heritage. Although seemingly contrary to mass digitization and Google Books projects, within cultural heritage, the limiting of access now presents the potential for ensuring access for the future.

The shifting digital divide
In 1964, Marshall McLuhan wrote of an “Age of Information and Communication” founded on the “electric media[s]” ability to instantly and constantly enable interactivity of humans and events (McLuhan, 1964, p. 248). Nearly 30 years later, in 1995, the United States Department of Commerce’s National Telecommunications and Information Administration (NTIA) produced the publication, “Falling Through the Net: A Survey of the ‘Have Nots’ in Rural and Urban America,” which largely focused on the reach of telephone services to the United States public (National Telecommunications and Information Administration, 1995). In it, distinctions based on location, race, age, economic status and education were found to play
roles in determining a population’s connectivity to information using both telephone and the Internet (National Telecommunications and Information Administration, 1995). Importantly, the survey recognizes the relevance access to information plays in equipping a society with the information infrastructure necessary to enable individuals to achieve “economic and social well-being” through the access, accumulation, and assimilation of information (National Telecommunications and Information Administration, 1995). Six years later, in 2001, Pippa Norris wrote of the emerging information society and varied perspectives of optimism and pessimism regarding its impact on the future of the digital divide and associated inequalities (Norris, 2001, p. 26). Norris identifies three sub-categories of division:

The global divide refers to the divergence of Internet access between industrialized and developing societies. The social divide concerns the gap between information rich and poor in each nation. And finally within the online community, the democratic divide signifies the difference between those who do, and do not, use the panoply of digital resources to engage, mobilize, and participate in public life. (Norris, 2001, p. 4)

Eszter Hargittai’s investigation goes beyond simply access to the Internet and focuses on one’s ability to effectively use the Internet and “demonstrate[s] that a second-level digital divide exists relative to specific abilities to effectively use the medium” (Hargittai, 2002). Shanyang Zhao and David Elesh go further, arguing that the second digital divide highlights the parallel between inequality in the offline world and inequality in the online world, in such a way as “saying the sky is free for everybody to use, but bring your own airplane” (Zhao and Elesh, 2007, p. 184). This is demonstrated by various social barriers, such as individuals in lower socio-economic classes, some of whom may not be literate (World Development Report 2016, 2016, p. 8); although they may have access to social networks, their active participation is hindered (Zhao and Elesh, 2007, p. 184). Building on this and other research into first- and second-level digital divides, a third-level divide emerged:

The third-level digital divide concerns disparities in the returns from Internet use within populations of users who exhibit broadly similar usage profiles and enjoy relatively autonomous and unfettered access to ICTs [Information and Communications Technology] and the Internet infrastructure. Third-level divides, therefore, relate to gaps in individuals’ capacity to translate their Internet access and use into favorable offline outcomes. (van Deursen and Helsper, 2015, p. 30)

This evolution of the digital divide does not imply that the various levels are superseded by the next level. Rather, as the digital information landscape continues to increase in complexity, ways by which to break down and measure how the information is dispersed, accessed and used will become similarly complex. In 2016, the World Bank estimated that of the households in the bottom 40% of their countries’ income distribution, 21% did not have access to a mobile phone, while 71% lacked Internet access (World Development Report 2016, 2016, p. 104). Although the potential access to worldwide information is becoming more sophisticated, the digital divide in its original form is still a reality for some populations.

Digital collections represent one of many subsets of information available on the Internet. Given the fact that digital collections themselves are limited by institutional capacity, potentially introducing yet another level to the digital divide, they are not the primary source of information for the majority of the world. By expanding the definition of digital collections to include similarly grouped sources of digital information located in an online, central location and freely available, the term begins to encompass more than just heritage and academic institutions, including also governments, non-governmental organizations and even private businesses that are choosing to disclose information to the public. All of these information producers have the potential to disseminate information that is culturally, economically and intellectually relevant to the well-being of a country’s population if made available, accessed, understood and used. However, given the many complex barriers that
prevent any or all of those components from being achieved, information producers must embrace the benefits and understand the limitations of digital collections, as their over-reliance on the production of solely digital information has the potential to disenfranchise those who lack a centralized or personal digital infrastructure, as well as those who do have access to those infrastructures but lack digital literacy. Simply put, “the perception of universal access made possible only by means of digital technologies is an illusion, because a set of complex power, financial, infrastructural, literacy, and other factors precondition availability of digital content” (Manzuch, 2017, p. 11).

For example, government documents created and disseminated in digital format in countries that lack the widespread infrastructure to support Internet accessibility or the bandwidth necessary to provide reliable and consistent access to that information do not produce a more informed public, despite an online presence. Similarly, countries in which there is a developed infrastructure to support the distribution of government information do not necessarily provide all citizens equal access to this information if there is a gap in the public’s “user digital capability” (Doong and Ho, 2012, p. 519). From an ethical perspective, the dissemination of information is only as good as the information’s accessibility and subsequent preservation. Access and preservation hold the power to create an informed society, and lacking either presents immediate and long-term implications for the understanding and guidance of a governmental body that is designed to be accountable to the people.

An additional question arises regarding the means of access to digital information, including via both personal computers and mobile phones. Both present great potential for the retrieval of digital information. However, there are also drawbacks, such as a weak infrastructure (West, 2015), accessibility obstacles of mobile phone interfaces versus personal computers (Bartikowski et al., 2018, p. 373) and the cost of devices not including connection fees (West, 2015, p. 2). Too much reliance on one or the other may result in the exclusion of certain parts of a population, despite both technologies being significant conduits for information. Furthermore, “smartphones are less appropriate than regular computers to engage in economic value-creating activities” (Bartikowski et al., 2018, p. 373), adding yet another potential area of divide, this one related to the quality of experience. Certainly, the burden falls on a government to understand the reach of its own country’s digital infrastructure, that infrastructure’s capacity to provide information to the public, and the limitations of the same infrastructure to reach everyone. An intentional use of select means to disseminate information enables a governmental body to influence understanding through the indirect limiting of transparency. There is now division within the digital divide that necessarily requires a balance of how information is distributed to ensure equal access and future preservation.

It is clear that basic access to the Internet is not the sole solution to bridging the variety of evolving digital divides. Creating an informed public requires technological infrastructure, digital literacy, affordability and acknowledgement of, and sensitivity to, social, cultural and economic barriers. As cultural heritage institutions continue work capturing, preserving and making information available, digital collections complement that work, but do not replace it.

The underlying challenge of digital information vulnerability
Digital information is made up of the same “1s” and “0s” regardless of file format, meaning digital file modifications may occur on documents, spreadsheets, datasets, images and the like. As a result, additional burdens fall to information managers to identify original information, verify that that information is free from manipulation and modification, and then preserve and make it available in ways that maintain its intended form. The particular challenge facing information workers goes beyond these burdens, however, extending into
the more ambiguous act of exercising one’s judgment regarding the integrity of a digital file, whether that file be born-digital or a result of in-house digitization.

From the born-digital perspective, an example can be seen in the influx of digital images created with mobile phones, shifting the “control, distribution, and knowledge” relative to traditional photography (Gómez Cruz and Meyer, 2012, p. 212), or the production of digital images occurring in a controlled environment, such as a digital reformating lab or publication office. This also introduces the issue of original purpose and the creator’s intended use. For instance, political communication has shifted from using only the more traditional forms of communication, such as speeches, printed press and posters, to now also embracing digital media as one of the main tools for political communication (Maarek, 2014, p. 14), thereby creating a new potential for digital misinformation related to political marketing. This example is not to suggest that historical political communication and marketing were free from deception and misinformation, which they were not (Brugioni, 1999, p. 127). But, in the context of a similar purpose and with unknown creators and creation environments, the speeds at which information spreads today has the potential to quickly reach large populations. Lyndon Kennedy and Shih-Fu Chang hypothesize that Internet users are more often than not presented with images that have already undergone a derivative process and are manipulated in a way so to shape a user’s interpretation (Kennedy and Chang, 2008, p. 1). Not only does this emphasize the significance of capturing born-digital items as close as possible to the original source, but also demonstrates how widespread the use and availability of digital tools to manipulate digital information have become.

From the perspective of in-house digitization, elements of manipulation may be introduced internally, rather than externally. As Fineman suggests:

…one might argue that there is no such thing as an absolutely unmanipulated photograph. The process of making a photograph—of translating the constantly changing, full-color, three-dimensional world into a flat, static, bounded image—involves dozens of conscious and unconscious decisions, from the moment one picks up the camera through the distribution of the final image. (Fineman, 2012, p. 6)

More specifically, the digitization process at its most fundamental state creates a digital rendering of a physical scene based on the encoding of light and its computer-based interpretation into recognizable digital images, often undergoing many “corrections and adaptations” along the way (Rudolf, 2006, p. 200). In today’s world, this complex process seems as easy as pressing a button on a camera or a phone, but the capturing process still involves multiple steps using both software and hardware, during which variations in image reproduction may be introduced. The important takeaway from this is that even if an original image undergoes no post-processing during digitization, there are many other variables at play throughout the capture process itself. Even at the stage of post-processing, there is little guidance as to “how much is too much, or if any at all is too much” (Mindel, 2016, p. 186). The ethical element at play here is related to the issue of intention and transparency: What is the desired effect that drives performing post-processing, and is that process documented for users to see and understand?

Digital image manipulation with particular emphasis on digital collections may occur for many reasons, some of which may be acceptable, such as restorative manipulation and compensatory enhancements, or for public appeal, aesthetics, and academic pursuits (Mindel, 2016, pp. 182–184). Alternatively, image manipulation performed for the sake of propaganda and misinformation, or to create “outright forgeries” (Hofer and Swan, 2005, p. 292), not only adds to the ethical concerns, in general, but specifically places a burden on those working to preserve cultural heritage to adapt and use new software, incorporate additional workflows and exercise even more attention to detail and scrutiny. Although image manipulation has been around for well over a century, these burdens are of even greater importance now, given
the widespread availability of digital creation and manipulation tools and the skills needed for their operation.

Digital content is vulnerable to manipulation both inside and outside of collection institutions. Manipulation can be beneficial or deceptive, depending on the intention and transparency of the manipulation performed. The “provenance of both analog and digital materials, as well as documentation about their storage environment, what has been done to them, and by whom, are the key aspects of establishing and maintaining trust” (Kirschenbaum et al., 2010, p. 27). Enhancing an image to yield greater information may be an acceptable form of manipulation, particularly with transparency regarding the steps taken to enhance. Information that is intentionally altered for the purpose of misrepresentation or deception clearly is not ethical image manipulation. Whether it be digital information in the form of images, audio, video, datasets or documents, alterations to an original with the intent to deceive and mislead is a reality of a digitally connected world. However, “necessity is the mother of invention,” and this threat continues to drive the field to develop and implement tools to counter deceptive information. In a field that is far from black and white, information management relies on current and future information managers to develop an awareness of both the potential and the deficiencies of digital information as it informs the public now and in the future, and the ethical implications presented by manipulation of any kind to digital images and how those implications challenge a collective trust in visual information.

An even more complex future

Similar to digitization operations, there are clear capacity issues related to the capture, dissemination and preservation of born-digital objects. A notion “that all materials have been, or ultimately will be, converted to a digital format” (Mills, 2015, p. 162) is most likely out of touch with the realities of capacity related to equipment, personnel, skills, storage and general funding. Similarly, a typical institution will not be able to capture every born-digital item produced within its walls, particularly given the funding scarcity in small and mid-sized institutions (Kay Rinehart et al., 2014, p. 29). Furthermore, libraries often serve as information centers for communities and regions, and, as such, the responsibility to capture, preserve and disseminate publications with local connections is often an expectation. With this responsibility comes the challenge of locating sources producing relevant information, monitoring those sources over time, and continually searching for new sources, which may appear and disappear without warning. Born-digital information may include reports, publications, emails, scholarship, audio and video, or, as Richard Pearse-Moses on behalf of the Society of American Archivists defines it, “information created in electronic format” (Pearce-Moses, 2005). In these cases, although the content differs from traditional digital collections that may focus more exclusively on cultural heritage and analog-to-digital conversion, given the nature in which it is stored and accessed, this paper still treats born-digital content as part of a digital collection. As the medium of information evolves and adapts, so, too, must the institutions in charge of capturing, preserving and disseminating that information.

Active participation in the collection development of born-digital content requires that institutions find a balance between the limitations of local infrastructure and the capture of high-value sources of information in a landscape of the ever-increasing production of information. The ethical complexities surrounding the selection of materials becomes complicated by the very nature of the digital medium and its susceptibility to being lost forever. Ross Harvey and Martha Mahard highlight this in their discussion of appraising born-digital materials:

The selection or appraisal decision traditionally came after enough time had passed to allow an object’s significance to be determined. With vulnerable digital objects that do not last unless they
receive attention from an early point in their existence, we do not have the luxury of time in which to develop an understanding of their significance. Selection changes from being a process carried out after time has passed to an ongoing activity starting as close as possible in time to the creation of a digital object. (Harvey and Mahard, 2013, p. 6)

Intellectual and cultural values are universal and overarching in concept, but differ in their determination and assignment from institution to institution. Subject matter, information source and uniqueness guide the selection and appraisal processes (Consultive Committee for Space Data Systems, 2012). Going beyond these elements, however, Jinfang Niu proposes a selection method for born-digital content that incorporates statistical sampling, risk analysis, and an appraisal process that uses mission alignment, value, cost, and feasibility as its criteria (Niu, 2014, p. 68). Additionally, Niu provides an overview of the intersection of traditional archival selection and appraisal approaches with that of born-digital selection and appraisal, including an outline of value in multiple contexts, such as administrative, fiscal, legal, evidential, informational, scientific, historical, and research (Niu, 2014, p. 72). Each category of value comes with its own set of sub-criteria that are influenced by professional and personal lenses through which appraisers evaluate. In the context of ethics, more significant than specific guidelines is an acknowledgement of both the subjectivity of selection and appraisal as well as the pace and volume at which digital content is created and potentially lost. The importance of ensuring that those in the cultural heritage fields acquire the skills necessary to allow their professional lenses to provide guidance in the face of limited resources, finite time and an endless supply of new digital information is central to preventing a future that reflects an incomplete past.

In addition to the selection, itself, is the challenge of handling simply the quantity of digital information being produced, complicating the “verification, provenance, and contextualization work that archivists have traditionally undertaken” (D’Ignazio and Klein, 2020, p. 171). In fact, stepping back even further, “the people who work in libraries and archives and in related preservation fields to ensure that knowledge of the present remains accessible for generations to come” (D’Ignazio and Klein, 2020, p. 199), are really part of the broader field of data work. In this context, libraries and archives are one part of “a larger ecology of knowledge, one that must be both sustainable and socially just” (D’Ignazio and Klein, 2020, p. 200). These concepts of sustainability and social justice, however, go beyond digital content as a product and its subsequent consumption. Instead, the ethical focus here is on those producing and managing the content and data in a world where “power is not equally distributed,” such that “those who wield power are disproportionately straight, white, able-bodied, cisgender men from the Global North” (D’Ignazio and Klein, 2020, p. 8). From this perspective, Catherine D’Ignazio and Lauren Klein identify “data feminism” as a term and concept, both founded on:

...a belief in and commitment to co-liberation: the idea that oppressive systems of power harm all of us, that they undermine the quality and validity of our work, and that they hinder us from creating true and lasting social impact with data science. (D’Ignazio and Klein, 2020, p. 9)

The ethical realization that those “behind the scenes” are facing ethical challenges internal to the fields in which they work presents a layer of complexity at the foundation of how digital information is created, used, preserved and managed. This should raise concerns about implicit biases and their potential to impact nearly every aspect of the data lifecycle.

Similarly, the data lifecycle is also subject to external influences, some of which have been mentioned previously in reference to its consumption, namely infrastructure and information literacy. Not explored in this paper is how search engines, or more specifically, their underlying algorithms, can determine what information is consumed by the masses. Safiya
Umoja Noble writes of these underlying influences in her book, *Algorithms of Oppression: How Search Engines Reinforce Racism*. In it, Noble describes that:

Part of the challenge of understanding algorithmic oppression is to understand that mathematical formulations to drive automated decisions are made by human beings. While we often think of terms such as “big data” and “algorithms” as being benign, neutral, or objective, they are anything but. The people that make these decisions hold all types of values, many of which openly promote racism, sexism, and false notions of meritocracy, which is well documented in studies of Silicon Valley and other tech corridors. (Noble, 2018, p. 1)

As a result, everyday technology that relies on computer code, including the Internet and artificial intelligence, are susceptible to discrimination being embedded within the code itself (Noble, 2018, p. 1). Whether intentional or unintentional, through this mode of perpetuating discriminatory ideologies, minority groups influenced by majority culture “are often subject to the whims of the majority and other commercial influences such as advertising when trying to affect the kinds of results that search engines offer about them and their identities” (Noble, 2018, p. 16). At the same time, as Internet companies grow, they possess greater influence over the information to which they provide access. In the context of digital collections, for example, Google digitized over 10 million books to make available online, and aside from issues related to the potential for violations of antitrust law, control in the form of digital content ownership makes terms of access vulnerable to shareholder interests (Noble, 2018, p. 157). Such vast and growing control over information should give pause as it has the potential to negate ethically informed curatorial decisions made by information professionals in libraries, archives and museums.

**Further conversation**

Although institutions may be faced with various capacity issues related to equipment and online infrastructures, an additional area of infrastructure is directly tied to skills capacity. What this means is that new digital collection programs require both the human and non-human elements in order to be successful. As mentioned, the landscape of information is continually changing, and therefore the fields connected to information creation, management, dissemination and preservation must adapt to ensure that these four components of a digital collection are fulfilled. Education in the academic realm is one major resource through which the information fields can equip future professionals with the skills necessary to overcome obstacles and prepare for future challenges. Ensuring that library, archives, and museum studies programs incorporate and keep pace with the ever-changing landscape of digital collections is the first line of defense for the field as a whole to collectively combat threats against information creation, access, dissemination and preservation. Secondly, institutions that are able to provide professional development for those involved in managing information have the potential to prevent information mismanagement by avoiding the reassignment of staff who do not have the requisite skills for new positions that demand current knowledge and skills related to digital information management. Combining both strategies to emphasize education and training at both the pre-professional and professional levels is one step to ensure topics discussed in this paper are acknowledged and addressed.

Similarly, as the complexity of the world of digital information will only continue to increase, it becomes essential that the tools being developed to counter “bad information” evolve and diversify. The discipline of digital forensics will no doubt continue to expand, and so, too, will those threats. Seeking new educational opportunities and working collaboratively with information technology professionals can lead to the development and increased strength of the ability to create a networked effort that is better able to carry out obligations related to the sound preservation of digital information.
Through education and collaborative efforts, additional tools need to be developed and made freely available. Specifically, open source content management systems are vital to expand institutional capacity as it relates to providing access to, and preservation of, digital content. Of the numerous barriers that exist for individuals to find and use digital content, there are also barriers that institutions face in providing that content and safeguarding it for future generations. Having access and the ability to use open source software that enables access and preservation would remove one major barrier to the creation of a more informed and connected world. However, even with general improvements in the accessibility of verified digital information, without an adequate digital literacy those efforts may have little effect. As this paper describes, the digital divide is complex, with many divisions and sub-divisions existing from the levels of community and region to countries and continents. Improving digital literacy at all levels will help to support the responsible access and use of digital information. Understanding how to improve digital literacy and the means that are the most effective is an area within this topic that deserves attention now and moving forward.

Following the logic that it takes more than just simple access to “good” information to create an informed public, the influence that search engines have in the retrieval of information should also be considered a relevant area for exploration. Once content is created or captured, made available online, and preserved on servers elsewhere, the producing institution’s role in how that information is reached and consumed is largely out of the institution’s control. This further reinforces the importance of creating a digital-literate public in order to navigate and understand how information access may be acted upon by third parties, and understand that the search for information is as much a skill as the ability to evaluate that information.

Alongside education, information literacy and open access tools, efforts focused on developing policies and procedures for institutions are essential to respond to the ethical considerations presented in this paper. Policies and procedures are reflections of everyday practices, and as such, must adapt to changes in the field of digital information. While there are some nationally and internationally accepted policies (Janowski and Hector, 2016) or statements (Garcia-Febo et al., 2012) (American Library Association, 2017), there needs to be internal organizational structures in place to provide ethics oversight. Emphasizing internal oversight will allow for greater flexibility and the ability to more quickly adapt and adopt policy that reflects larger ethical concepts as they evolve. UNESCO highlights broad knowledge society policy from 11 countries, concerning topics that revolve around information including, research, development, innovation, economy, skills, education and government (Janowski and Hector, 2016). Annual reporting and internal audits allow institutions to reflect on digital collection equity, cultural representation, historical context and accessibility. Furthermore, engaging with communities and stakeholders should be established and reinforced as policy for institutions working more exclusively with cultural heritage materials.

Conclusion
This paper approaches the topics of ethics and information management from the perspective of a digital collections librarian. Although interdisciplinary in nature, this topic is particularly relevant and important for those in the field of information science, and specifically for those who manage digital collections. At the same time, each section is not intended to serve as a surrogate for the extensive literature that is available on each topic. Rather, it is designed to identify the ethical common thread among the areas discussed and demonstrate the interconnectedness of all areas of information management, with specific attention given to digital collections. Finding balance in building a digital collection, ensuring that that collection’s digital files are properly maintained and meet a level of digital integrity,
providing transparency regarding a file’s history and understanding the limitations of the
digital information contained within a collection are key areas that deserve the attention of
both those working in the field of information science and the public at large. The choices
made in the acquisition, creation, management, dissemination and preservation of
information generate a ripple effect in how and where that information is consumed,
understood and used. At the same time, the public has a responsibility to hold those
accountable who withhold information, misrepresent information, discard information and
generally limit access to information, as complacency only encourages information
mismanagement.

Information is a powerful tool that enables one to critically understand the world, make
decisions and judgments about the world, and therefore be able to constructively contribute
to society. Ethical management of information provides a pathway to achieve those abilities,
and it runs counter to information “gatekeeping.” Digital formats have expanded how
information is managed, requiring new approaches and considerations to ensure that this
format maintains accessibility and is preserved for future access. The concepts discussed in
this paper have specific relevance to professionals working with digital collections, whether
those be at cultural heritage institutions, governmental bodies, academic institutions, or other
organizations that collect, manage, disseminate and preserve information. A sensitivity to,
and an awareness of, both the potential that digital information has to inform a large
population and information’s vulnerabilities, including manipulation, misrepresentation and
lack of access or preservation, are the first steps to embracing the ethical management of
digital information in today’s digital world.

References
tools/ethics (accessed 20 April 2021).
and the well-being of ethnic minority and majority consumers: a multi-country investigation”,
digitization and access for archives”, The Reading Room: A Journal of Special Collections, Vol. 2
No. 1, pp. 7-25.
Brugioni, D.A. (1999), Photo Fakery: The History and Techniques of Photographic Deception and
Manipulation, 1st ed., Brassey’s, Dulles, Va.
museumedited by Anne Thurston, D”, Records Management Journal, Vol. 24 No. 3,
pp. 238-250.
Cocciolo, A. (2016), “Email as cultural heritage resource: appraisal solutions from an art museum

Consultive Committee for Space Data Systems (2012), Recommended Practice for an Open Archival Information System Reference Model, NASA, Washington, DC.

Cotton des Houssayes, J.-B. (1906), The Duties and Qualifications of a Librarian: A Discourse Pronounced in the General Assembly of the Sorbonne, McClurg, Chicago, p. 1780.


Corresponding author
David Mindel can be contacted at: dmindel@uwlax.edu