ARCHEION
2025, T. CXXVI, s. 10–29
ISSN 0066-6041
e-ISSN 2658-1264

DOI 10.4467/26581264ARC.25.008.21260

Getting to Digital

Anne J. Gilliland

University of California (USA)
gilliland@gseis.ucla.edu, ORCID 0000-0002-4897-7780

ABSTRACT
It has been over 40 years since archives began to automate their processes and digitize content. During that period, archival practices and ideas have experienced unprecedented development, archives have established substantial digital presences and records creators have shifted almost completely to digital production. Today archives are being looked to by Artificial Intelligence developers and scholars alike as one of the few remaining sources of big data, both for training Large Language Models and for surfacing new knowledge about the past and humanity. And yet archives have yet to reach "full" digital capacity and fluency, and not simply because of resource limitations. One might ask whether they ever will – or even should? This lecture looks back on the history of digital developments in archives, examining the motivations and challenges arising at different points. It contemplates why and how the goalposts have repeatedly moved, and how the questions that have arisen in archival decision-making have changed as a consequence and how they may continue to do so as we look to the future nature, roles and practices of archives.

KEYWORDS
archival concepts, archival theory, Artificial Intelligence, digital archives, electronic records management

Droga do cyfryzacji

STRESZCZENIE
Od ponad 40 lat archiwa automatyzują swoje procesy i digitalizują treści. W tym czasie praktyki i koncepcje archiwistyczne przeżyły bezprecedensowy rozwój, archiwa zbudowały znaczącą obecność w sieci, a dokumentacja tworzona jest niemal wyłącznie w formie cyfrowej. Obecnie archiwa są postrzegane przez twórców sztucznej inteligencji i naukowców jako jedno z niewielu pozostałych źródeł dużych zbiorów danych, zarówno do szkolenia dużych modeli językowych, jak i do odkrywania nowej wiedzy o przeszłości i ludzkości. Niemniej jednak archiwa nie osiągnęły jeszcze pełni cyfrowych możliwości i funkcjonalności, i to nie tylko z powodu ograniczonych zasobów. Można zadać pytanie, czy kiedykolwiek to osiągną – a nawet czy powinny? Wykład ten stanowi przegląd historii cyfryzacji archiwów, zawierający analizę motywacji i wyzwań pojawiających się na różnych jej etapach. Rozważa, dlaczego i jak ewoluowały jej cele, w jaki sposób zmieniały się pytania pojawiające się w procesie podejmowania decyzji oraz jak mogą one dalej ewoluować z perspektywy przyszłej natury archiwów, ich roli i praktyk archiwalnych.

SŁOWA KLUCZOWE
koncepcje archiwalne, teoria archiwalna, sztuczna inteligencja, archiwa cyfrowe, zarządzanie dokumentacją elektroniczną

"Twentieth-century archival science is in peril. Modern bureaucracy has rapidly undergone qualitative and quantitative changes. Gone will be the manuscripts along with their meaning when machine prints replace them. Much of the semi-official business will be absorbed by the phone. Extreme growth of state machinery and the ease in copying written text – typewriter or mimeograph machine – cause a flood of paper by enabling the multiplication rather than the duplication of an endless number of prints, each stored in a different registry needlessly increasing its volume; hence resulting in veritable »geological layers« of paper that will flood the archives with its shapeless mass"[1].

Introduction[2]

We are now living in an era when increasing numbers of scholars, bureaucrats and members of the wider general public have no memory of a time before the World Wide Web and digital almost everything. It is a time when expectations and demands for digital content are at an all-time high. Researchers are desirous to apply computational tools as well as innovative intellectual frameworks to investigate the past and surface hidden narratives in unprecedented and unanticipated ways, while Artificial Intelligence developers hunger to get access to previously inaccessible stores of what is variously termed "legacy data" or "big

---

1 K. Konarski, *On the Issues of Modern Polish Archival Science,* translated and edited by Bartosz Nowożycki, "American Archivist" 2017, vol. 80, no. 1, p. 226.

2 This article presents the transcript of a lecture delivered by Professor Anne J. Gilliland on October 16, 2025, at Maria Curie-Skłodowska University in Lublin, as part of the international lecture series "Konarski Lectures", organized by the Polish State Archives. Gilliland is an internationally renowned archivist, scholar, and professor specializing in archival studies. She currently serves as Director of the Center for Information as Evidence at the University of California, Los Angeles (UCLA) School of Education & Information Studies. Professor Gilliland is widely regarded as one of the most influential voices in contemporary global archival discourse. For years, she has shaped critical conversations about the nature of archives and their societal roles. Her work – from the seminal *Conceptualizing 21st Century Archives* to her recent studies on refugee archives and contested heritage – continues to redefine the boundaries of archival scholarship. She is a co-originator of concepts such as "affective archives" and the driving force behind an international research initiative that explores the rights of individuals and communities to their own histories. A Fellow of the Society of American Archivists, she has received numerous accolades for her groundbreaking contributions to the field.

data of the past" (i.e., archives) to train Large Language Models and potentially support entirely new applications and explorations and expositions of the past[3].

Many archives still remain far from being at the "full" digital capacity needed to satisfy such visions and demands, however. Moreover, the pace of change in record-creating technologies as well as the increasing elusiveness of evidence of creative activity that Kazimierz Konarski noted almost 100 years ago[4], have only increased exponentially, leaving the archival field in a constant state of trying to keep up with the latest technologies and public demands while potentially losing an existential battle to more agile and less evidentially concerned information fields. We may have survived the 20th century, but to riff off Konarski, is 21st century archival science in the new era of AI in peril because it has not yet fully "got to digital?".

If so, it would be in spite of the fact that many professional bodies, national and international, have worked unceasingly since the 1980s to develop standards and other infrastructure for archival information exchange, digitization, and digital preservation; as well as the considerable, and in archival terms, extraordinary amounts of government and other agency funds have been sunk into research and development of regulations and methods for working with electronic or born-digital records[5]. It would also be in spite of the fact that during this period the scope and conceptual sophistication of the archival field has undergone unprecedented growth as have professional education, and archival engagement with whole new audiences and community spaces – a growth that is clearly reflected in the extensive body of recent archival scholarship regarding the conceptual, practical and ethical concerns of the field. A growth that I would argue in no small part can be attributed to the generative effects of having to respond to how digital technology has been taken up across society for government, business,

---

3 Consider, for example, the ambitious vision of Time Machine Europe, which is seeking to digitize and collate archives across Europe and combine this content with digital infrastructure and tools that will allow it to surface and present the richness of European history in completely new ways including Augmented/Virtual Reality (AR/VR) applications that, in the developers' words "may lead to simulations of hypothetical spatiotemporal 4D reconstructions of European economic, social, cultural and geographical evolution across times". See: Time Machine, About us, https://www.timemachine.eu/about-us/ [access: 5.10.2025].

4 K. Konarski, *On the Issues of Modern Polish Archival Science*…

5 For example, all of the work of members of the DLM-Forum since 1991; almost 3 decades of InterPARES-related research projects, and vast investments in digital library developments to which archives have been contributors during the same period.

social engagement, creativity and so many other purposes, purposes that bridge time and space, and peoples and communities in whole new ways.

So why then do gaps remain between the present reality and the promise and aspirations for archives in terms of being fully digital? Setting aside the obvious limitations of available financial and technical resources and expertise, to what extent have shifting priorities and the complexities of responsibly stewarding the archival record had an impact on getting to digital? Are wider aspirations to full digitality actually shared by archivists, or might this *status quo* perhaps also reflect a certain professional queasiness about abdicating the unique evidentiary role of archivists, ceding control of archives, or ethical concerns about how archives should or might be used if they were indeed fully digital?

Canadian archival theorist, Terry Cook famously argued in 2013 that:

> "archival paradigms over the past 150 years have gone through four phases: from juridical legacy to cultural memory to societal engagement to community archiving [...] as the broader intellectual currents have changed from pre-modern to modern to postmodern to contemporary"[6].

Cook notes that these shared mindsets emerged separately but in a cumulative fashion across and within the archival profession to comprise the archival ecosystem in which we find ourselves today. I would argue, however, that while broader intellectual currents have certainly influenced the changing and diverse ways in which archivists today frame their work, technology and associated "information", "digital" and now "AI" paradigms, have also been integral factors that have led archivists to reconsider and recalibrate their most fundamental concepts and practices.

This lecture will focus on how, in the past several decades, those concepts and practices have emerged, been reconsidered and recalibrated as the field responded first to the arrival of the "information age" and then to the current immersive "digital age", and most immediately in the incipient "AI age". It will review 3 different strands of archival activity relating to digital implementations that came to prominence more or less chronologically and separately over the past several decades but today are much more imbricated. As it took hold,

---

[6] T. Cook, *Evidence, Memory, Identity and Community: Four Shifting Archival Paradigms*, "Archival Science" 2013, vol. 13, pp. 95–120.

each strand tended to engage different groups within the archival field, to collaborate with different external partners, and to apply different approaches. Each strand also moved through discernible "phases" that were influenced in various ways by the paradigms that were at work within the wider professional and technological environments. I have loosely named these 3 strands "electronic records management", "automating archival description" and "disseminating digital content". They are rather old-fashioned terms today, but they capture the emphases of the time. For each, I will briefly review how they emerged and evolved in order to highlight specific conceptual and practical questions that arose or were revisited during each phase. In conclusion, I will reflect on where these developments have brought us to today. It is a history that is beginning to fade from professional memory, but it is worth remembering because it illuminates some of the reasons why we have seen increasing divergence and debate within the field around what our priority concerns and constituencies should be. It reminds us that in some cases "we have been here before" and it informs us about what we might take into consideration as we move forward. Above all it reassuringly demonstrates that the questions with which archivists must grapple can be very complex and dynamic, and that even if we seem sometimes to fall short of public expectations, we have already come a really long way in thinking them through.

Let me begin first by setting up some context. If we consider, as Cook does, that the era of the modern archival profession began in the later 19th century, then it is interesting to remember that this closely parallels, temporally, the development of computing capacities. The use of analog computing technology for government censuses, as well as scientific and engineering applications dates back as far as the 1880s. Nevertheless, 20th century archival consensus around principles such as provenance, respect des fonds and original order that are central to what Cook has dubbed, the "evidence" paradigm, were derived from the observations and practices of archivists who had deep experience working with the hardcopy textual records that were generated by government and institutional bureaucracies and were not thinking about the by-products of emerging computing technologies as records. The profession coalesced, therefore, around a profoundly physical paradigm that was based on an understanding of the physical characteristics of records and on principles that emphasize the organic relationship of a record or a fond to the activity that generated it, in order to retain as much as possible of the contexts of creation and

administrative use. These principles in turn prioritized the juridical needs and historical evidentiary practices of the time.

Major steps toward the recognition of this as the operating paradigm for an archival profession were the endorsement of the Dutch archivists Samuel Muller, Johan A. Feith and Robert Fruin's *Manual of Descriptive Practices* at the 1910 International Conference of Archivists and Librarians in Brussels and the many subsequent translations of the Dutch Manual into other languages. The 1920s saw the publication of several key texts in different countries further articulating archival practice and an associated body of concepts, including those by Hilary Jenkinson in London in 1922[7], Pierre Fournier in Paris in 1924[8], Eugenio Casanova in Siena in 1928[9], and Kazimierz Konarski here in Poland in 1929[10]. The publication of these texts opened up a fundamental question that goes to the heart of the professional nature of the field and later would drive diverging approaches to the management of electronic records: does theory derive from practice (i.e., it is inductively developed) or does theory dictate the parameters of practice (i.e., through a top-down or deductive approach)?

These publications in turn were followed in the 1930s by a flurry of new journals and archival associations, and in 1948, by the establishment of the International Council on Archives, a body that would play an important role in delineating professional education and standards. In these developments, the arrangement and description of archives tended to be the primary focus. But there was always a certain tension between what might be agreed upon as universal principles and what was considered to be locally-specific for reasons of historical, juridical, political, cultural, or even linguistic differences. This raises a question that has since frequently recurred and elicited different arguments and decisions at different moments especially when contemplating digital developments because of their implications for standardization and shared digital infrastructure: what is the optimal balance that needs to be struck between generally understood archival principles and local needs and practices?

---

7 H. Jenkinson, *A Manual of Archive Administration*, London 1922.

8 P. Fournier, *Conseils pratiques pour le classement et l'inventaire des archives et l'édition des documents historiques écrits*, Paris 1924.

9 E. Casanova, *Archivistica*, Siena 1928.

10 K. Konarski, *Nowożytna archiwistyka polska i jej zadania,* dodatek do czasopisma „Archeion", Warszawa 1929.

The more visionary of these archivists, such as Konarski, were also already contemplating what the nature and pace of change in bureaucratic production as well as in the form, nature and volume of the record being produced were likely to mean for those archival practices. New documentary and reproduction technologies were proliferating in the first half of the 20th century. By the 1930s, archivists were nervously asking questions about whether they should be acquiring film-based materials or whether those should be left as the purview of a new specialist field of film archives since they required different kinds of expertise and technology to preserve and play back.

By the 1960s and 1970s, in concert with political awakenings in various countries, as well as intellectual shifts in history away from modern scientific history focused on elites toward "history from the bottom up", archivists began to open up the possibilities of a playing a proactive documentary role in response to questions and criticisms from these new historians and ethnic and women's studies scholars relating to the sufficiency, and later the equitability, of extant archival holdings. What did the archival record not include? And might audiovisual documentation, oral histories and electronic records have some capacity to fill in gaps in the record and the archive? These impulses presaged both Australian approaches to the appraisal of electronic records articulated from the 1990s onwards, and from the 2000s, the societal engagement and community archiving movements referenced by Cook. They also provoked a debate that had already been vehemently addressed by Jenkinson in the 1940s: how involved should archivists be in shaping the historical record?

Electronic recordkeeping: from machine readable records to AI algorithms

Computing technology began to take on a more prominent profile in administrative and research activities around 1945 after John Maunchly and John Presper Eckert at the Moore School at the University of Pennsylvania in the United States developed ENIAC, a computer based on vacuum tubes rather than telephone relays, and in the next year founded the Eckert-Maunchly Computer Corporation that was acquired by Sperry Rand in 1950. From then on, computing developed rapidly and widely, evolving from large analog to digital mainframe systems, to microcomputers, to networked systems, to the Internet

and then the World Wide Web and the Cloud accessible from anywhere through all sorts of increasingly small personal devices, and culminating in where we are today in this incipient age of AI.

The first byproducts of computing systems, usually punchcards, printouts or datatapes, were referred to, in the US at least, as "machine readable records" and up until the 1980s archives with such materials largely adopted the management practices of the emerging social science data archives community because it was felt that the archival paradigm could not readily be applied to these kinds of materials. There was also a tendency to assert that whatever was valuable as a record could be printed out and kept in that form, a tendency that in several countries by the 1990s was definitively quashed by legal decisions that ruled that while the printout might have retained the informational content of the record, for various reasons, it did not retain its evidentiary value. As bureaucracies increasingly adopted automated record-keeping, and the systems that they deployed moved beyond what could readily be printed, existing approaches for managing machine-readable records were no longer viable or appropriate. We see at this point the professional terminology shift to refer instead to electronic records and electronic recordkeeping, thus marking the transition of the world of recordkeeping into a predominantly digital rather than a physical paradigm, and indeed today we tend to refer to "digital recordkeeping" and "born-digital records". Now having to work with a virtual rather than a physical record, archivists were compelled to address a cluster of questions fundamental to the conceptual foundations of both archival practice and theory, the most fundamental being "what is a record?". To answer this question, it was necessary to identify the characteristics and capacities of a record absent of any particular physical manifestation, and how to preserve those over time. Without a conceptual understanding, it was becoming impossible to identify what constituted a complete record in complex and linked systems, engendering an associated question "where is the record?". It was also impossible if one were taking the appraisal approach that was promoted by Australian archivists through the records continuum model from the mid-1990s onwards, to design new record-keeping systems that would be able not only to generate a complete and preservable record, but potentially additional documentation of events and processes that otherwise might go undocumented. The latter approach raised 2 questions that resonate with earlier concerns: What is the appropriate role of archivists in ensuring that records, and good records at that, are created in

the first place? And if archivists are to be involved in designing systems to create records, should they advise entities to create records that have not traditionally been part of the business process?

The question of whether practice drives theory or vice versa, at this point is no longer merely an intellectual debate. In order to "find" the record for the purposes of records scheduling, appraisal or ongoing stewardship, archivists now had to determine conceptually, what makes something a record, and then find a way to mandate, instantiate and preserve that record, whether that be through legislation and regulation, systems design, or digital preservation or most probably some combination of all of these. This was a particular concern in the US context, where registry systems have never been implemented, even for paper records, and thus predetermined scheduling approaches have had been less effective than in many European countries. In 1999, therefore, we see David Bearman argue that a record is something that crosses a transactional barrier, from human to human, human to computer, or computer to computer. His reasoning was that records are by-products of transactions associated with particular activities, and when designing a computer system, a transaction is a kind of event that can be flagged and captured by a system with relative ease:

> "All (!) it requires is that we can capture all transactions entering and leaving the system when they are created, ensuring that the original context of their creation and content is documented, and that the requirements of evidence are preserved over time"[11].

A year later, drawing upon the considerable electronic records research that had taken place in the 1990s, Philip Eppard and I identified 3 sets of characteristics of records that spoke to the more expansive question: what is a trustworthy record? The first of these was a more detailed delineation than Bearman's that was intended to convey to systems designers and records creators more specifically what it was that archivists wished a system to be able to generate. The second addressed one of the characteristics of records that can make it difficult to identify the moment when they should be captured, as well as how they might be described at different points in their active and archival lives. The third was drawn from the work of

[11] D. Bearman, *Realities and Chimeras of Electronic Records Preservation*, "D-Lib Magazine" 1999, vol. 5, no. 4, https://www.dlib.org/dlib/april99/bearman/04bearman.html [access: 5.10.2025].

the InterPARES projects and spoke to procedural requirements for ensuing that records are trustworthy:

> "Records are heterogeneous distributed objects comprising selected data elements that are pulled together by activity-related metadata such as audit trails, reports, and views through a process prescribed by the business function for a purpose that is juridically required [...].
>
> Records are temporally contingent – they take on different values and are subject to different uses at different points in time. Records are also time-bound in the sense that they are created for a specific purpose in relation to a specific time-bound action.
>
> The degree to which a record can be considered reliable is dependent upon the level of procedural and technical control exercised during its creation and management in its active life. Authenticity, by contrast, is the responsibility of archival management of inactive records, and is an absolute concept"[12].

Amelia Acker in 2017 built upon the idea of records as temporally contingent when she notably added another consideration when she asked: when is a record? Her work also added back in a consideration of the role played by the material aspects of digital records creation in considering the nature of the record. Based on her examination of time-based digital communications such as text messages sent from a mobile phone on a train that was moving between communication networks, she offered a framework that could be used "to account for the transmission and materiality of electronic records [...] by locating them within the infrastructures of contemporary, networked communication"[13].

It is this definitional work around what constitutes a record and its various affordances that has built the intellectual muscle that will help archivists today to cope with a shift perhaps even more profound than that from physical to digital – a shift into the era of machine intelligence bringing with it what we might dub the "AI paradigm". This new paradigm yet again opens up questions about trust in the record and the characteristics that might help us determine what, where,

---

[12] A.J. Gilliland-Swetland, P.B. Eppard, *Preserving the Authenticity of Contingent Digital Objects*, "D-Lib Magazine" 2000, vol. 6, no. 7/8, https://www.dlib.org/dlib/july00/eppard/07eppard.html [access: 5.10.2025].

[13] A. Acker, *When is a Record?* [in:] *Research in the Archival Multiverse*, eds. A.J. Gilliland, S. McKemmish, A.J. Lau, Melbourne 2017, pp. 288–323.

when is a record. This time, however, we also have to pay particular attention to how the record was created and which inputs and what kinds of business, cultural and other assumptions and perspectives were associated not only with the algorithmic reasoning, but also the training of the Large Language Model that has been applied. This opens up some new questions such as: Can an algorithm be a record? If so, how can the genesis and evolution of that algorithm be traced and preserved? What mechanisms can be put in place to ensure the accountability and transparency of algorithmic processes that generate records?

By the 1990s, it had become clear that addressing these kinds of conceptual questions and assertions in practice necessitates capturing, creating and managing extensive metadata, a concept that in the library world at the time referred primarily to cataloging information, but that in the computing world invoked the contextual information surrounding the record itself. For archivists, the concept of metadata would take on a wider scope still[14]. While it is analogous, to some extent, with the contextual information archivists were already accustomed to incorporating into their descriptions of physical records, the introduction of the concept of metadata was still seen as an innovation in the archival field. It would become an important bridge between the largely autonomous areas of electronic records management and archival arrangement and description. That bridge was actually made by David Wallace in a landmark article in 1995 in which he argued that:

> "Traditional archival description [...] has limited our ability to capture crucial contextual and structural information throughout the life cycle of record-keeping systems that are essential for fully understanding the fonds in our institutions. This shortcoming has resulted in an inadequate knowledge base for appraisal and access provision. Such complications will only become more magnified as distributed computing and complex software applications continue to expand throughout organizations"[15].

The period of intensive electronic records management research and development that largely occurred in the 1990s and early 2000s was based

---

[14] A.J. Gilliland, *Setting the Stage* [in:] *Introduction to Metadata. Third edition*, ed. M. Baca, 2016, https://www.getty.edu/publications/intrometadata/setting-the-stage/ [access: 5.10.2025].

[15] D. Wallace, *Managing the Present: Metadata as Archival Description*, "Archivaria" 1995, vol. 39, pp. 11–21, https://archivaria.ca/index.php/archivaria/article/view/12064 [access: 5.10.2025].

strongly around concerns of legal admissibility, bureaucratic accountability and transparency, and best practices for recordkeeping. It was characterized at the time as a return to evidence-based practice from the documentary turn taken by the field in the 1970s and 1980s. Instead of the emphasis being on how arrangement and description activities could retain and highlight evidence as was the case in the original evidence paradigm, it focused on the record in the active business context and how that record might be appraised and subsequently preserved by archivists. It surfaced a plethora of difficult questions such as: at what point should the archivist be engaged with the electronic record? Is the concept of an original record still viable? Is fixity still a necessary or even feasible characteristic of an electronic record? What makes an electronic record worth preserving? Who should be responsible for that preservation? Will archives take physical custody of electronic records or will they leave them in situ, on the original system and instead steward the conditions under which it is maintained? And even more fundamentally, what is to be archived – is it the record, or is it the capacity to recreate or render an authentic copy of the record?

Bearman introduced 2 possible frameworks to help make the case for organizational electronic records management and to support archival decision-making. The first, and the one that had the most immediate uptake, was risk management – what were the potential risks to records creators of not addressing these issues with their various categories of electronic records and did those risks justify the cost and effort of electronic records management? Where those risks were seen to be legal, financial or reputational consequences, this approach fit well with the evidentiary (re)turn. Recognizing that not all organizations are equally motivated by mandates and these types of negative consequences, however, Bearman proposed a second framework – enterprise management. This was a more positive approach that fit well with a digital asset management approach and appealed to entities such as the creative industries. In this context, where the product of such enterprises might be a digital publication or a media production, the question of what a record might be took on another dimension – how to discern what was the product, what was the record, and what was the metadata that documented the production of both, and then how to manage each most effectively? Of course, when viewed through a digital asset management lens, these delineations could also overlap. The granularity at which digital asset management and content management systems function also raised new questions about whether something might exist that is smaller than a record

about which archivists should be concerned – its component parts. The notion of records components had also come up in InterPARES research as it examined different types of composite records. As archivists contemplate different levels of aggregation, therefore, it is no longer sufficient to think in terms of aggregates and individual items. Those items themselves can be decomposed into various components, in turn begging a question for both describing and delivering the record of: how to ensure that the item or component is not taken out of context?

Automating archival description: building information infrastructures

In apposition to electronic records management, "archival automation" in the 1980s and 1990s tended to refer to efforts to develop and disseminate digital description of archival holdings, efforts that necessitated both standardizing archival practices and participating in the development of national and international information infrastructures, often with other so-called information professions. In his landmark 1945 article *As We May Think*, American engineer and science administrator Vannevar Bush laid out a bold holistic vision for structuring the scientific record to make "our bewildering store of knowledge" more accessible:

> "A record, if it is to be useful to science, must be continuously extended, it must be stored, and above all it must be consulted. Today we make the record conventionally by writing and photography, followed by printing; but we also record on film, on wax disks, and on magnetic wires. Even if utterly new recording procedures do not appear, these present ones are certainly in the process of modification and extension"[16].

Bush's vision was highly influential on the library and information science fields, and it also invoked archives. And so began a different quest that was related to optimizing how stored information might be organized, collated and

[16] V. Bush, *As We May Think*, "The Atlantic" 1945, July, https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/ [access: 5.10.2025].

reformatted in ways that the knowledge it contained could be more accessible for researchers. While the electronic records archivists were working within an evidence paradigm, archivists who specialized in arrangement and description came together with colleagues in the library and information science fields to work within what we might call an "information" paradigm. By the 1980s, the so-called "Information Age" began to become a reality as computing began to permeate government, business and research activities and microcomputers became widely available to the broader public. As early as the 1960s, while still working with early mainframe systems, librarians had begun to develop national and international "information infrastructure" with the goals of efficiencies of effort and cost effectiveness through the standardization of description and shared cataloging of bibliographic materials, while information scientists focused on optimizing information search and retrieval in online catalogs and other information systems. By the early 1980s, the infrastructure developed around MaRC (Machine-readable Cataloguing) had been widely adopted by libraries. Thought leaders within the archival profession were encouraging archivists to think of themselves not so much as history or evidence professionals, but rather as information managers and specialists. They were asking whether archives could also benefit from standardizing what, despite general adherence to foundational archival principles, remained disparate and often idiosyncratic descriptive practices. While archivists continued to develop detailed narrative finding aids, they asked whether a shorter form catalogue record containing standardized elements of description could be devised and disseminated online? If so, should archival description be shoehorned into the MaRC infrastructure that was already in place within many libraries nationally and internationally, or should archivists devise their own standard? Led by some of the institutions that had major research collections, the American archival profession decided on the former, and in 1984, adopted the MaRC Archival and Manuscripts Control format, although it should be noted that that format was never designed to accommodate electronic records. The drivers were different from those of libraries, however. They included: how can a researcher find out that a fond or a collection exists and make an informed decision about whether to travel to see its contents? How can we bring archival descriptive practices into sufficient conformity that researchers don't have completely different experiences at different archives? And would descriptive standards help to bring archives everywhere up to the same basic level of description? Moreover, it again raised

the question of how to balance standardized practices with local descriptive needs, and as with developments in electronic records management, it opened up whole new questions about how to prepare archivists with the technical skills to work in these new areas. It also introduced a strong tendency toward subject access and arguably a deprioritization of provenancial access within online descriptive systems that would be further exacerbated by the possibilities of searching not only online descriptions but also digital archival materials by keyword on the World Wide Web.

In 1999, following a lengthy process begun in 1988, ICA adopted ISAD(G), the General International Standard Archival Description, as a standard for cataloging archival material. It defined elements that should be contained in a finding aid but did not prescribe the exact form they should take, thus trying to enable different archival jurisdictions to find their own balance between local needs and standardized practices. Some went on to develop their own ISAD(G) compliant standards, such as Encoded Archival Description. From this point on, archival description became increasingly sophisticated, with additional standards being developed for other facets of archival description, crosswalks with other metadata standards used to describe digitized materials, and standards to support the integration of additional kinds of heritage materials held by other cultural, heritage and memory institutions such as galleries and museums, and possibly one could also argue that yet another paradigm – a "heritage paradigm" – emerged associated with convergences of so-called GLAM (galleries, libraries, archives and museums) institutions. However, there perennial concerns have remained within the archival field about whether the evidence function of archives is being subordinated to the information function, or indeed, some kind of heritage umbrella, and also whether the standards being used by archivists are sufficiently capable of addressing the complex descriptive needs of what have come to be termed "born-digital" records.

Having reached this point, however, the critical turn in the archival field has challenged and, in some cases, repudiated the standards-based approach. New questions have been raised such as: what are the downsides of standards? Who was party to the development of those standards, with which kinds of uses and users in mind? Whose perspectives and needs were omitted? How can standardized description be sensitive to diverse cultural and historical practices and divergent or even incommensurable perspectives and experiences? Should we always assume that repositories are open to sharing descriptive information?

How can small or community-based archives acquire the technical expertise and infrastructure necessary to participate in wider information infrastructures? Speaking to one of the earliest professional questions about balancing local and wider imperatives in archival description, we now see widespread efforts by archives in retrospective "reparative" description as well as initiatives such as the TK (Traditional Knowledge) Labels that can be used by Indigenous and other marginalized communities to augment or comment upon existing standardized descriptions prepared by non-Indigenous repositories.

Disseminating digital content: digitizing the archives

Inevitably, at the same time that archival description was being automated and going through standardization, users of archives were asking whether it might not be possible to make versions of the archives themselves available digitally. The development of Project Gutenberg at the University of Illinois at Urbana-Champaign in 1971 is often pointed to as beginning of the era of transforming non-digital documents into electronic texts. At first, the technology was heavily circumscribed by technical limitations and could only support text transmission with small file sizes. The philosophy of the initiative was one of "Replicator technology" – the ability for texts to be endlessly reproduced and made available to the general public at little to no cost "in forms a vast majority of the computers, programs and people can easily read, use, quote, and search"[17]. This was an information rather than an evidence approach, and if one wanted to have an image of the original document, one still had to resort to microfilm or microfiche. Developments in data compression in the mid-1980s and of the Hypertext Transfer Protocol in 1989 as well as personal computers and compact disk technology set the stage for the digitization and digital dissemination of archival content first on CD ROM and then over the Web. When digitization really began to be employed by archives for accessibility purposes, it was to overcome a major limitation posed by the physicality of unique fonds that had meant they could only reside in one place at a time. One of the pioneers of archival digitization was The General Archive of the Indies in Seville, which began digitizing its vast

[17] Project Gutenberg, The History and Philosophy of Project Gutenberg, by Michael Hart, https://www.gutenberg.org/about/background/history_and_philosophy.html [access: 5.10.2025].

holdings relating to the Spanish empire in the Americas and Asia in order to help to make them available to those in its far-flung former colonies. Another major motivation for putting archival content online has been to encourage its use by non-traditional audiences, and especially schoolchildren, although doing so has required archivists to develop new skills in developing accompanying pedagogical materials.

Making a digital copy available meant that the researcher no longer needed to travel in person to the archive. However, it also put more pressure on the digital finding aid and the design of the archival information system to explain the context and structure of the records in question, since in most cases there was no longer a reference archivist to serve as a mediator between the researcher and the record. While this situation afforded the researcher more privacy in consulting the records, it also engendered archival concerns about the record being misunderstood, taken out of context, or re-used in some inappropriate way. There was also an increased possibility of someone browsing the online materials using simple keyword searches of accidentally encountering sensitive information, or of a researcher compiling disparate information from different archives or applying new computational tools to them in ways that might compromise the privacy or security of individuals, institutions or knowledge mentioned in the records. One interesting consequence of putting descriptive information online, and especially of also making the associated records digitally available, is that it can illuminate different perspectives on shared histories such as diplomatic events, political and demographic movements and major international conflicts. Furthermore, when the records pertaining to such matters are subject to different closure and opening periods in each jurisdiction, then it can result in public pressure on the archives that hold related but still closed records.

An enduring question for archives in the face of inadequate or inconsistent resources has been how to set priorities for digitization. However, as more and more content is digitized and archives have come more onto the radar of politicians, the general public and tech developers, more troubling questions have arisen such as: how to prevent archival content being taken out of context and used to show individuals in a negative light, especially through social media? With the advent of deep fakes, how can end users know whether what is disseminated is a true representation of what is in the archive? And how can archivists ensure

that archival content is not being used to train Large Language Models absent explanations of its provenance, purpose and various built-in biases?

Conclusion

As this very brief review has hopefully indicated, bringing archives up to full digital capacity is not merely a matter of resources or technical innovation. It is also a matter of balancing motivations and shifting priorities. But at its core, it involves the ongoing recalibration of some of the archival profession's most fundamental concepts, practices and concerns without ceding its unique role *vis-á-vis* the stewardship of the record and its care for the wellbeing of the communities associated with that record. Many of the questions that have arisen are both complex and perduring, but there are important new twists also, and addressing them in their future manifestations will require strong conceptual understanding and foresight as well as a thorough knowledge of archival practice and its trajectories. Each of the various paradigms at work in the field centres important perspectives and priorities, but they are often working at odds with each other. It is not to be expected that this will change. Rather, we need to acknowledge that today we are a multi-paradigm, multi-priority field, and we need to prepare archivists who are cognizant of each and who can work within and across them, even if they are primarily invested in one more than others. The next generation archival workforce will need to be able to do the core work of archival processing as well as implement second-level processing workflows to make archival holdings data/digital computation ready and much of the latter work will itself likely be done with the assistance of computational methods[18]. They will also have to put protections in place for those mentioned in the digital record who might be vulnerable and it may well be that archives will forgo some

[18] See, for example, D. Canning, L. Jaillant, *AI to review government records: new work to unlock historically significant digital records*, "AI & Society" 2025, vol. 40, no. 6, pp. 4433-4445, https://doi.org/10.1007/s00146-025-02221-0; P. Murietta-Flores, R. Vega-Sánchez, A. Sánchez-Díaz, F. Cruz-Ríos, *Unlocking Colonial Records with Artificial Intelligence: Achieving the Automated Transcription of Large-Scale 16th and 17th-Century Latin American Historical Collections*, "STAR: Science and Technology of Archaeological Research" 2025, vol. 11, no. 1, pp. 1–13, https://doi.org/10.1080/20548923.2025.2484828; M. Hedges, R. Marciano, E. Goudarouli, *Introduction to the Special Issue on Computational Archival Science*, "Journal of Computing and Cultural Heritage" 2022, vol. 15, no. 1, pp. 1–2, https://doi.org/10.1145/3495004.

aspects of digital accessibility in order to do this. In these latter respects, I would offer just a few more questions to ponder: Are we preparing new archivists to do this work? Do we know what that workflow needs to be? Do we understand how to identify vulnerability in an era of algorithmic processing, and do we know how to reduce that vulnerability as we make our holdings computation-ready?

On a final note, to return to my opening comments, whether the majority of archives ever get to the point of being fully digital in all their activities, or whether that is even desirable, in my opinion, remain open questions. What hopefully is clear from this brief review, however, is that the years spent by the archival field in grappling with the evolving conceptual as well as practical challenges that the digital has continuously presented have put the profession in a much stronger position not only to survive, but also to play a key leadership role in ensuring that the 21st century AI era is one that not only creates a legacy of trustworthy and inclusive historical records, but also is equipped with the ability to illuminate and hold accountable the provenance, processes and motivations by and through which those records were created.

Bibliography

Acker A., *When is a Record?* [in:] *Research in the Archival Multiverse*, eds. A.J. Gilliland, S. McKemmish, A.J. Lau, Melbourne 2017, pp. 288–323.

Bearman D., *Realities and Chimeras of Electronic Records Preservation*, "D-Lib Magazine" 1999, vol. 5, no. 4, https://www.dlib.org/dlib/april99/bearman/04bearman.html [access: 5.10.2025].

Bush V., *As We May Think*, "The Atlantic" 1945, July, https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/ [access: 5.10.2025].

Canning D., Jaillant L., *AI to review government records: new work to unlock historically significant digital records*, "AI & Society" 2025, vol. 40, no. 6, pp. 4433–4445, https://doi.org/10.1007/s00146-025-02221-0.

Casanova E., *Archivistica*, Siena 1928.

Cook T., *Evidence, Memory, Identity and Community: Four Shifting Archival Paradigms*, "Archival Science" 2013, vol. 13, pp. 95–120.

Fournier P., *Conseils pratiques pour le classement et l'inventaire des archives et l'édition des documents historiques écrits*, Paris 1924.

Gilliland A.J., *Setting the Stage* [in:] *Introduction to Metadata. Third edition*, ed. M. Baca, 2016, https://www.getty.edu/publications/intrometadata/setting-the-stage/ [access: 5.10.2025].

Gilliland-Swetland A.J., Eppard P.B., *Preserving the Authenticity of Contingent Digital Objects*, "D-Lib Magazine" 2000, vol. 6, no. 7/8, https://www.dlib.org/dlib/july00/eppard/07eppard.html [access: 5.10.2025].

Hedges M., Marciano R., Goudarouli E., *Introduction to the Special Issue on Computational Archival Science*, "Journal of Computing and Cultural Heritage" 2022, vol. 15, no. 1, pp. 1–2, https://doi.org/10.1145/3495004.

Jenkinson H., *A Manual of Archive Administration*, London 1922.

Konarski K., *Nowożytna archiwistyka polska i jej zadania,* dodatek do czasopisma „Archeion", Warszawa 1929.

Konarski K., *On the Issues of Modern Polish Archival Science,* translated and edited by Bartosz Nowożycki, "American Archivist" 2017, vol. 80, no. 1, pp. 213–229.

Murietta-Flores P., Vega-Sánchez R., Sánchez-Díaz A., Cruz-Ríos F., *Unlocking Colonial Records with Artificial Intelligence: Achieving the Automated Transcription of Large-Scale 16th and 17th-Century Latin American Historical Collections*, "STAR: Science and Technology of Archaeological Research" 2025, vol. 11, no. 1, pp. 1–13, https://doi.org/10.1080/20548923.2025.2484828.

Project Gutenberg, The History and Philosophy of Project Gutenberg, by Michael Hart, https://www.gutenberg.org/about/background/history_and_philosophy.html [access: 5.10.2025].

Time Machine, About us, https://www.timemachine.eu/about-us/ [access: 5.10.2025].

Wallace D., *Managing the Present: Metadata as Archival Description*, "Archivaria" 1995, vol. 39, pp. 11–21, https://archivaria.ca/index.php/archivaria/article/view/12064 [access: 5.10.2025].