



“Born digital” shedding light into the darkness of digital culture

Larry Stapleton¹ · Lise Jaillant²

Published online: 20 February 2022

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

The *AI and Society* community’s fascination with technology, culture and the custodians of cultural heritage stretches back decades. In a special issue of the journal from the turn of the millennium, Victoria Vesna (2000) asserted that “Databases are Us”, drawing attention to the relationship between emerging data processing systems and cultural heritage, citing the extraordinary difficulties that libraries and librarians face when trying to place structure upon complex cultural datasets. Vesna robustly addressed the interactions between the creators of cultural materials (artists and musicians for example) and those who are charged with curating these artefacts (librarians and archivists for example). She traced a direct connection from the printing press to what she calls the “information age”. Vesna concluded that a rethink of “our relationship between consciousness and our organisation and dissemination of data” was needed (p. 157). Creators, curators and information architects now inhabited a shared online space, co-creating cultural artefacts using data as the new raw material. Gutenberg’s ink was replaced by bits and bytes. She contrasted moving through a physical exhibition space with traversing datasets, comparing the architecture of the museum with the architectures of emerging data processing machines.

Vesna was keenly aware of the historical backdrop against which these tensions unfolded. In the early twentieth century, Marcel Duchamp rejected painting in favour of curating as an artistic endeavour. He became a freelance librarian in Paris, parodying the museum as a sacred institution in which cultural works are secreted in dark alcoves, away from ordinary people. In 1934, Duchamp published a volume of over 300 special editions of “Green Box”, foreshadowing both

pop-art and the digital revolution in which cultural artefacts are endlessly copied in an ever-expanding cultural universe of data (Bloch 1974). A digital version of this collection is available for view at the New York Metropolitan Museum of Modern Art website where the original is now housed.¹

For centuries, elites have attempted to control and “own” important cultural materials. In 1994, Bill Gates famously bid over \$30 million for the Codex Leicester, a 1510 CE notebook of Leonardo Da Vinci, amidst fears that this priceless object would be unavailable to the public gaze. These fears were only allayed when the publisher, Continuum, obtained the intellectual property rights to the notebooks’ images. The contents of this codex later became available to the public in a digital version on CD-ROM and were also included as a desktop theme in Microsoft’s Windows 95 operating system. The hoarding of cultural materials in private repositories is nothing new, but digital adds a new twist, in this case revealing content that might otherwise have remained beyond the reach of the interested scholar.

Digital may also hide cultural materials in an impenetrable archaeology of technological obsolescence or in inaccessible vaults forged from wire and signal. In his book *New Dark Age*, James Bridle explored the ways we have been conditioned to think of computers and databases as technologies which reveal a clearer and simpler world, reducing complexity and expanding human agency. However, for Bridle, these assumptions were far from the truth. He contended that the advent of digital technology renders the world opaque, and artefacts that begin their lives in the digital world, may eventually be lost to history.

Preserving access was also a pressing concern for Brewster Kahle. As an MIT graduate fascinated by the potential of thinking machines, Kahle embarked on a project in the mid-1990s that was an attempt to create a digital library which offered “universal access to all knowledge”. He called it “Alexa Internet” after the great Library of Alexandria (the first digital “Alexa” was not a smart home system!). Recognising the dangers of losing early internet content to the darkness of technical obsolescence, he curated out-of-print

✉ Larry Stapleton
larrystapleton@knewfutures.com; lstapleton@wit.ie

Lise Jaillant
l.jaillant@lboro.ac.uk

¹ INSYTE-Coolley Research Lab, Waterford Institute of Technology, Waterford, Ireland

² Loughborough University, Loughborough, UK

¹ <https://www.metmuseum.org/art/collection/search/492389>

web pages which he would make available for free to anyone who wished to view them. Kahle saved 85 million web pages from digital obscurity. For this pioneering work of curating the early internet Kahle has been rightly enrolled in the “Internet Hall of Fame.” In the spirit of *AI and Society*, this computer scientist with an interest in AI, recorded for posterity the cultural phenomenon that was the nascent world wide web. Without Kahle, these born-digital cultural artefacts would have been lost to history (Ximm 2013).

In his contribution to the “21st Birthday” special issue of *AI and Society*, Banerjee (2007) offered a vision in which AI impacts society in ways that are liberating, ethical and able to shift attitudes in positive directions. Banerjee was keen to see a discourse from which a new roadmap would point us towards a deeper understanding of the ethical and aesthetic implications of contemporary machine intelligence in its relation to human society and culture. For him, a knowledge of culture and technology and how they interpenetrate each other is both “ethical and liberating” (p. 418). Digital technology at once offers new possibilities whilst, simultaneously, propagating very complex problems. Opportunities and challenges that digital calls forth are not confined to the technical domain alone. Instead, there is a dense jungle of interwoven technical and non-technical risks and rewards. Making sense of it all requires a cross-disciplinary discourse that brings together a range of scholarly traditions and professional perspectives and experience.

This special issue of *AI and Society* includes the voices of scholars from the computer sciences and the humanities alongside insights from professional custodians of cultural artefacts. By bringing these voices together, we hope that this special issue will satisfy Banerjee’s desire for a rich ethical and liberating discourse about the relationship between technology and culture.

The first contributions included in this special issue are in an Open Forum and include papers from practitioners, and others, working with important collections in Ireland and the United Kingdom. They outline current approaches to archival processing with Artificial Intelligence (AI).

In ‘Creating a Linked-Data Thesaurus for Irish Traditional Music,’ Treasa Harkin offers her insights into the complexities of presenting Irish traditional music records. It is an excellent and novel application of smart databases and semantic web technologies at the Irish Traditional Music Archive (ITMA). These solutions provide an effective means of categorising songs and musicians in machine-readable knowledge models. Harkin explains how advancements into linked-data models and thesauri are allowing for a more inclusive categorisation of these variations. Into the future, the Irish traditional music community looks to further develop user-focussed practices based on smart data and related technologies.

Discussing the broader picture for archival processing, Bunn et al. identify the current disparity between scholarly researchers and archivists. The challenge of accessing dark archives, both for data processors and researchers, comes together in their searching article “Dark Archives or a Dark Age for Reasoning over Archives?” Bunn et al. stress the need to find common ground between these two, once disparate, communities. Shared goals are needed to leverage AI. While advances in AI are helping to reduce the intensive labour behind such processes, the authors remind us that the current use of machine learning (ML) in cultural organisations is still evolving.

In ‘Managing and Accessing Web Archives: Irish Practitioners’ Perspectives,’ Keating and Finnegan discuss the historical necessity of preserving the Web Archive at the National Library of Ireland (NLI). This article addresses the problems and restrictions the NLI is facing with regards to their born-digital resources. Keating and Finnegan present the shared nature of issues national archives need to address across the globe, and the necessity and yet difficulty of archiving the web in real time. Keating and Finnegan recognise the positive steps made by Archives Unleashed and Archive-IT to develop user-friendly tools and datasets. They call for further advancements and collaborations between knowledge bases for processing born-digital archives.

Next, we have two Curmudgeon papers. These are short, opinion contributions focussing on issues of concern to those in the AI/archive world.

In ‘Will Archivists use AI to Enhance or to Dumb Down our Societal Memory?’ Titia and Bram van der Werf ask if archivists will “use AI to Enhance or to Dumb Down our Societal Memory?”. This is framed within the challenge of the enormity of the data involved and the systems needed to process this data. They remind us that marginalised identities—namely those excluded from the historical narrative—have been neglected in all this. In our second curmudgeon paper by Ennals and Jenkins entitled ‘Born Free: A Tale of Two Rivers,’ they address the pressing challenge of the exclusion of indigenous peoples from historical discourses. They reflect upon the events surrounding a catastrophic fire at the Jagger University Library in South Africa in April 2021. The article draws attention to the critical importance to humanity of African cultural assets, especially in the light of colonising forces down through history. They also consider Amazon’s proposal to build their headquarters in Cape Town on a sacred burial site. While Amazon offers books and music from South Africa to the world, their proposed headquarters could destroy a sacred burial site of central importance to indigenous communities in Africa. This is an all-to-common paradox when we consider digital cultural assets and points to a discourse in which different priorities, values, narratives and histories collide.

Summarising, the Curmudgeon articles challenge us to ensure AI is used appropriately if we are to avoid discriminatory biases. Indeed, if we are not careful, important connections to our past and digital futures which benefit all people, rather than a small chosen few, could be threatened.

In the next section, the special issue presents regular academic “born digital” related research.

The first paper in this section is ‘Unlocking Digital Archives: Cross-Disciplinary Perspectives on AI and Born-Digital Data’ authored by Lise Jaillant and Annalina Caputo. They survey note that many born-digital records are inaccessible due to a range of issues, including privacy rules, copyright laws, and technical issues including methodologies and approaches which need to be revisited. They consider the possibility that AI could be used to bring these records out of the darkness and in to the light. New technologies can also create new challenges. The article outlines various solutions to accessing currently inaccessible materials, proposing digital consortiums, AI sensitivity reviews, and restricted access with identification as important approaches for consideration.

Viewing digital cultural collections as data, Kirsten Carter, Abby Gondek, Ted Randby, Richard Marciano and William Underwood present the “Future of Archives and Records Management initiative” case study of The Morgenthau Holocaust Collections Project at the Roosevelt Library. Consulting WWII records, the project explores how AI and ML are being applied to make this hard-to-reach archival content accessible to the public. Here, the authors consider the challenge of interoperability and survey innovative technologies which can potentially incorporate multiple digital formats. Combining human expertise and machine learning this paper presents the Morgenthau Holocaust Collections Project interface as “a living, growing, tangible product, one that invites public exploration and scholarly conversation as ML and AI experiments continue.”

In ‘Finding Light in Dark Archives: Using AI to Connect Context and Content in Email,’ Stephanie Decker, David Kirsch, Santhilata Venkata and Adam Nix offer invaluable insight into the methodologies and proposed practices for discovering email content within large, multi-custodian archives. Acknowledging that archival users will come with their own particular technical skill sets, the article proposes innovative routes to connect content and context. This will help accommodate multifarious approaches for archive users and their individual research needs.

In ‘Jumping into the Artistic Deep End: Building the Catalogue Raisonné,’ Todd Dobbs, Aileen Benedict and Zbigniew Ras discuss the challenge of artistic authentication in the catalogue raisonné (a resource assembled by art scholars to hold information about artist’s works). “Recent advances in machine learning and imaging have outperformed humans in tasks of image classification,” they claim and they offer

evidence of a significant improvement in authentication using the techniques they propose. With further collaboration, Dobbs et al. envisage greater accuracy and a more tailored approach to authentication of historical art works.

Current data standards for sharing cultural artefacts across libraries, archives and museums are woefully inadequate, given the complex nature of the metadata associated with these artefacts. For example, the Machine-Readable Cataloguing standard (MARC), which is widely used in Europe, is not designed to manage the rich variety of objects now available in collections. In their contribution ‘Digital Cultural Heritage Standards: From Silo to Semantic Web,’ Brenda O’Neill and Larry Stapleton highlight the need for a symbiosis between human knowledge and machine learning. They note a problem of interoperability due to the inadequacy of metadata standards for sharing digital cultural data on the web. They propose an extended Metadata Encoding and Transmission Standard (METS) as a machine-readable metadata standard appropriate for sharing complex digital data on cultural artefacts across the internet. This extended METS standard could be used to create machine-readable knowledge models of cultural knowledge which would be amenable to machine reasoning using descriptive logic. This, in turn, lays the foundation for a machine reasoning system developed using linked data and semantic web technologies by which people could explore and engage with digital cultural materials in a new and exciting way.

Too often, public and private sector data systems embody marginalising cultural forces in society. This can be a source of anguish for those at the margins who experience their force. Noeleen Donnelly, Larry Stapleton and Jennifer O’Mahoney scrutinise this issue in public data systems in Ireland. In 2015, constitutional changes in Ireland were designed to embrace LGBTQI+ rights and create a more inclusive and just society. This paper provides empirical evidence which shows how born-digital data systems still embody the old order in which many LGBTQ+ people felt marginalised. These systems fossilise gender binaries in ways that are deeply at odds with modern public values as enshrined in constitutional changes. ‘Born Digital and Marginalisation: An Empirical Study of How Born Digital Data Systems Continue the Legacy of Social Violence towards LGBTQI+ Communities in Ireland’ reveals the continued and widespread use of gender binary classification systems, which “keep alive violence and oppression long after civil rights have been enshrined in constitutional law.” Technology development communities must confront the ethical choices associated with the development of advanced, socially-sustainable, digital data systems.

In her article, ‘Using Linked Data to Discover Born-Digital Materials: The Design and Evaluation of the NAISCL Interlinking Framework for Libraries, Archives and Museums’, Lucy McKenna provides an overview of data

interlinking tools, approaches and requirements that can allow libraries, archives and museums to “expose born-digital resources to a large community of potential users.” Acknowledging and addressing the challenge faced by information professionals to implement such tools, McKenna demonstrates the power of the NAISC-L model. The paper also shows how new skill sets are needed to design and deploy linked data models like the NAISC-L solution.

The Magdalene Laundries were Roman Catholic institutions in Ireland from the eighteenth to twentieth century that housed so-called ‘fallen’ women, while the Industrial Schools of the nineteenth and twentieth centuries housed children. Both institutions have since closed and have been subject to public inquiry into the horror of what happened within their walls. In ‘The role of born digital data in confronting a difficult and contested past through digital storytelling: The Waterford Memories Project,’ Jennifer O’Mahoney highlights the importance of born-digital data for the survivor narratives of the Magdalene Laundries and Industrial Schools in the South-East of Ireland. Dr. O’Mahoney notes that access to official Magdalene Laundry data sets is prohibited. These archival embargoes are complicit in the silencing of women. In response, the Waterford Memories Project has created its own archival resource, digitally collecting the testimonies of the women’s experiences as a means of public record. Where history has been locked away, O’Mahoney reminds us that it is so important to empower survivors of institutional abuse who are too often silenced and disempowered by society at large. Digital technologies can be an important way of unlocking meaningful histories and giving voice to the silenced.

The special issue closes with a contribution to the student forum in which the journal presents work from emerging scholars and doctoral candidates. Angeliki Tzouganatou, in her article, ‘Towards a Participatory Digital Culture: Opening up Born-Digital Archives’, asks “how open to the public should born-digital archives be?” Drawing connections with the articles from other authors in this issue, Tzouganatou

investigates open knowledge, AI, and the importance of placing human experience at the centre of future development trends.

This special issue marks an exciting time for AI in cultural institutions. It was born of the AURA (Archives in the UK/Republic of Ireland & Artificial Intelligence) research network.² The network of universities and practitioners explore ways of unlocking cultural assets that are preserved in digital archives but closed to the public or are difficult to access. The community was interdisciplinary, including computer scientists, digital humanists, archivists, and scholars. We all came together to discuss the future of archival processing. Clearly, the work has only just begun.

At a time of ongoing social change and upheaval, digital solutions such as AI present both opportunities and challenges. If deployed in appropriate ways, perhaps it can also aid in our journey towards more inclusive research practices, laying a strong foundation for the future of archives as repositories of our shared history and a basis for deeper, mutual understanding and respect.

References

- Banerjee P (2007) Technology of culture: the roadmap of a journey undertaken. *AI Soc* 21(4):411–419
- Bloch S (1974) Marcel Duchamp’s Green box. *Art J* 34(1):25–29
- Kahle B (2005) The open library. IDB Productions (MP3 file)
- Singh A (2007) Action and reason in the theory of ayurveda. *AI Soc* 21(1–2):27–46
- Vesna V (2000) Databases are us. *AI Soc* 14(2):155–175
- Ximm A (2013) Active personal archiving and the internet archive. In: Hawkins D (ed) *Preserving our digital heritage*. Information Today Inc., USA

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

² www.aura-network.net