

## Digitising the modern archive

Katrina Dean\*

**Dr Katrina Dean** is University Archivist at the University of Melbourne. She was Curator of the History of Science at the British Library (2006–2010), a post-doctoral researcher in the School of Geographical Sciences at the University of Bristol (2004–2005) and previously worked at the National Archives of Australia and the Australian Science Archives Project. She has published several articles in scholarly journals, and is co-editor of a new book on nineteenth-century inventor of photography William Henry Fox Talbot: *William Henry Fox Talbot: Beyond Photography* (Yale University Press, 2013).

In 1979 French sociologist Jean-François Lyotard predicted ‘that anything in the constituted body of knowledge that is not translatable’ into computer-readable packages of information ‘will be abandoned and that the direction of new research will be dictated by the possibility of its eventual results being translatable into computer language’.<sup>1</sup> Despite the efforts of archivists and (digital) scholars, most of the archival legacy of the modern period remains untranslated into computer-readable language and accessible only to those with traditional archival research skills or specialist reference services. Archivists remain the keepers of important stores of evidence and authentic information that are becoming dangerously irrelevant. The postmodern, digital archive only compounds the fundamental flaws of the modern archive, itself an incomplete project.

### Fundamentally flawed?

Archives are modern institutions which gained currency following the democratic revolutions in Europe at the end of the eighteenth century, appropriating earlier public and private collections.<sup>2</sup> Their evolution in the English-speaking worlds of Sir Hilary Jenkinson<sup>3</sup> and Theodore R Schellenberg<sup>4</sup> emphasised elements of modern, rational bureaucracy: public accountability, systems and hierarchies of classification and professional expertise.

Tendencies of the modern state to control information, enforce categories and claim dimensions of experience are often emphasised.<sup>5</sup> Very recent examples emphasise the risk of distortion in the modern archive, not least by obscuring what is included and what is left out.<sup>6</sup>

Yet the potential of modern archives as both narratives and sites of political contest has been widely demonstrated. Dramatic political events such as the opening of the Stasi Archives<sup>7</sup> and access to archives of the Apartheid state<sup>8</sup> underline this potential, as well as its limitations. Routine discoveries of social, economic, and cultural complexity characterise research using archives of the modern era.<sup>9</sup>

---

\*Email: [kjdean@unimelb.edu.au](mailto:kjdean@unimelb.edu.au)

### **Dangerously irrelevant**

Bearman's critique that modern archives were inefficient, bureaucratic and irrelevant to consumers' needs might be seen in the context of not only postmodernism but also neo-liberalism. The traditional provinces of archival activity were focused on ill-targeted and unachievable goals.<sup>10</sup> Some archivists belatedly addressed obvious shortcomings of modern archival practice, for example the More Product Less Process approach to processing archives.<sup>11</sup>

Digital information, ephemeral, fragile and poorly controlled, detached from the implied physical contexts of traditional repositories, only accentuates many of the problems outlined by Bearman regarding the importance and value of the information, its fragility, its ability to be adequately described within available resources and its ability to connect with users. These practical problems are intensified by the digital economy, shifting content from the public realm to commercial ownership, either through the domination of commercially generated content or the commercial ownership and exploitation of third party and user-generated content. The act of publishing archival documents online and thus bringing them out of a controlled and moderated public sphere and into a digital economy has also brought new questions of intellectual property and rights, including privacy, to the fore and has accentuated issues of the ownership of cultural property in general.

In response, digital archival processes are if anything more interventionist and labour intensive. The level of descriptive, technical and administrative metadata required to manage, preserve and discover digitised archives generally exceeds the level of metadata required to manage and make accessible their physical counterparts. Increased standards of explicit evidential value or significance, description, preservation, discoverability and compliance required to bring archival collections online cause us to revisit the inadequacy of our attempts to meet these standards in the traditional archive in the first place. The only justification for this increased investment to make digital content available can be the potential of the postmodern archive to reach much larger audiences, and to be accessed in more independent and interactive ways. This promise makes the modern, traditional archive seem increasingly irrelevant in a digital world.

Not least owing to the resource-intensiveness of these processes, there is increasingly a gap between the information that is born digital or has been translated into computer-readable form and the larger and arguably more important reserves not yet translated. In 2012, a sample of 403 European archives/records offices had digitised an estimated 12% of their collections, and aimed to digitise a further 55%.<sup>12</sup> Across a sample of 200 cultural institutions, 7% of archival records had been digitised with the ambition to digitise a further 52% and no intent to digitise 40% (presumably not least owing to barriers to digitisation outlined above). A follow-up study in 2014 projects it will take 30 years to digitise the target for Europe's cultural heritage.<sup>13</sup> A 2010 study by the Collections Trust estimated it would cost 100 billion euros to digitise the collections of Europe's museums, archives and libraries in addition to up to 25 billion euros for preservation of and access to the digitised collections over 10 years.<sup>14</sup>

In the meantime, while potential uses of digitally available sources may be richer, the scope and potential of digital sources currently available for research remains impoverished in comparison to traditional archival sources and it cannot be assumed that the most important questions can be addressed using only digital sources.

### Unfailingly important

Whatever its failings, which are no doubt considerable, the modern archive is one of the richest sources of information available and documents sides of life (bureaucratic, social, cultural and personal) not repeatable in other sources, oral or public. Born-digital records preserved as archives have scarcely begun to replicate the richness of modern archives, whether originating from government, personal or community collections. The focus of the digital economy and digital scholarship on digitally available information can miss the best content, and the full scope and potential of the archive.

Efforts in digital scholarship to date often focus on metadata registries or rich descriptions,<sup>15</sup> with limited online access to archival records; or case studies using specific sets of digitised archives. Digital resources that do engage more deeply with modern archives take years to create, with researchers often performing the bibliographical work of cataloguing, transcribing or extracting information from and interpreting published and archival sources, in collaboration with archivists and other information professionals.<sup>16</sup> While the digital findings or output of this research are important, arguably more important are the efforts to bring the archives into computer-readable form and thus enable their transmission.

Despite the language adopted by the digital preservation community, it may be somewhat misleading to think about archives in terms of computer-readable information packages. Archives are about relationships and for their evidence and informational value to be fully explored and exploited they must reveal relationships between contexts and records and among sources. There must remain the possibility to expose what is missing. In this sense, archival exegesis remains a largely un-automated task relying on 'trained minds', whether individuals or collaborations between producers (or keepers) of archives and users (researchers).

In the modern archive, many of these relationships were implicit; in the architecture of buildings, storage configurations, old registers, administrative files, draft lists, curatorial and researcher knowledge, and the character of collections and their histories. Short of digitising whole collections and transposing these contexts into metadata, it seems unlikely that collections in their present configurations will be transmitted into the future knowledge economy.<sup>17</sup> New digital collections will be built on top of the old archives, carrying forward texts, images, sounds and artefacts of a forgotten civilisation. It makes sense to me to take stock of modern archives, their shortcomings as well as their unrealised potential, and to try to better understand how to bring forward this potential into digital culture.

### Endnotes

1. Jean-François Lyotard, *The Postmodern Condition: A Report on Knowledge*, translation from the French by Geoff Bennington and Brian Massumi, Manchester University Press, Manchester, 1984, p. 4.
2. Lucie Favier, *La Mémoire de l'État: histoire des Archives nationales*, Fayard, Paris, 2004.
3. Hilary Jenkinson, *A Manual of Archive Administration*, Clarendon Press, Oxford, 1922 (revised editions 1937 and 1965).
4. TR Schellenberg, *Modern Archives, Principles and Techniques*, FW Cheshire, Melbourne, 1956.
5. Thomas Richards, *The Imperial Archive: Knowledge and the Fantasy of Empire*, Verso, London, 1993.

6. Richard Drayton, 'The Foreign Office Secretly Hoarded 1.2 Million Files. It's Historical Narcissism', *The Guardian*, 27 October 2014, available at <<http://www.theguardian.com/commentisfree/2013/oct/27/uk-foreign-office-secret-files>>, accessed 3 March 2014.
7. Simon Menner, 'Images From the Secret Stasi Archives, Or: What Does Big Brother See, While He is Watching?', available at <<http://www.simonmenner.com/pages/Stasi-Index.htm>>, accessed 3 March 2014.
8. Albie Sachs, 'Archives, Truth and Reconciliation', *Archivaria*, no. 62, Fall 2006, pp. 1–14, available at <<http://journals.sfu.ca/archivar/index.php/archivaria/article/view/12887/14118>>, accessed 3 March 2014.
9. The great post-war social panopticon of the Mass Observation Archive at Sussex University is the paradigm example, see <<http://www.massobs.org.uk/index.htm>>, accessed 19 June 2014. Also see John Lack, 'Prest's Social Survey of Melbourne', in *eMelbourne: The City Past and Present*, July 2008, Department of History, University of Melbourne, Melbourne, available at <<http://www.emelbourne.net.au/biogs/EM01186b.htm>>, accessed 2 March 2014.
10. Kate Cumming, 'Revisiting Bearman', presentation to workshop Reinventing Archival Methods, Sydney, 29 November 2012, available at <<http://recordkeepingroundtable.files.wordpress.com/2012/12/revisiting-bearman-cumming.pdf>>, accessed 2 March 2014.
11. Mark A Green and Dennis Meissner, 'More Product, Less Process: Revamping Traditional Archival Processing', *American Archivist*, vol. 68, no. 2, 2005, pp. 208–63.
12. Natasha Stoker and Rene Vogels, Panteia (NL) on behalf of the ENUMERATE Thematic Network, 'ENUMERATE, Survey Report on Digitisation in European Cultural Heritage Institutions', May 2012, p. 11, available at <<http://www.enumerate.eu/fileadmin/ENUMERATE/documents/ENUMERATE-Digitisation-Survey-2012.pdf>>, accessed 2 March 2014.
13. Natasha Stoker and Rene Vogels, Panteia (NL) on behalf of the ENUMERATE Thematic Network, 'Survey Report on Digitisation in European Cultural Heritage Institutions 2014', January 2014, available at <<http://www.enumerate.eu/fileadmin/ENUMERATE/documents/ENUMERATE-Digitisation-Survey-2014.pdf>>, accessed 2 March 2014.
14. Nick Poole, the Collections Trust, 'The Cost of Digitising Europe's Cultural Heritage. A Report for the Comité des Sages of the European Commission', November 2010, available at <[http://ec.europa.eu/information\\_society/activities/digital\\_libraries/doc/refgroup/annexes/digiti\\_report.pdf](http://ec.europa.eu/information_society/activities/digital_libraries/doc/refgroup/annexes/digiti_report.pdf)>, accessed 2 March 2014.
15. See, for example, Humanities Networked Infrastructure, <<http://huni.net.au/>>, accessed 2 March 2014.
16. For examples, see Old Bailey Online, <http://www.oldbaileyonline.org/>; Founders and Survivors, Australian Life Courses in Historical Context (1803–1920), <<http://foundersandsurvivors.org/>>; Board of Longitude, <<http://cudl.lib.cam.ac.uk/collections/longitude/>>; Codebreakers: Makers of Modern Genetics, Wellcome Library, <<http://wellcomelibrary.org/using-the-library/subject-guides/genetics/makers-of-modern-genetics/>>, all accessed 2 March 2014.
17. Research tools like Trove's collection profiler or visualisations can preserve some of this information, represented in familiar or unfamiliar terms.