



# 'Digitalising a National Archive': interview with John Sheridan, Digital Director at The National Archives, UK

John Sheridan<sup>1</sup> · Clare Foster<sup>2</sup>

Received: 22 November 2021 / Accepted: 25 May 2022 / Published online: 4 July 2022  
© The Author(s) 2022

## Abstract

John Sheridan talks with Clare L E Foster, sharing some wider observations about the challenges of the digital transformation of The National Archives. (<https://www.nationalarchives.gov.uk/about/our-role/executive-team/john-sheridan/>).

**Keywords** Archives · National memory · Digitalisation · Curation · Digital humanities · Digital research · National records

*CF: You speak widely around the UK about the issues raised as a national archive moves from paper records to digital, and how this highlights the challenges many archives are confronting. Could you repeat here some of the points you made in discussion at CRASSH - about the fundamental questions around repetition and persuasion that have come up for you in this enormous digitizing project?*

JS: Well, all archives are both enablers and disrupters of repetition. As we digitise our collections, we are enablers in the sense that what we know about our collections, what has been digitally described through our catalogues and can be readily searched, or what has been digitised and can be readily accessed online, typically tends to support historical research and analysis down previously well-trodden paths. Popular historians or the media often tell the public broadly familiar stories, and have a sizeable audience. Providing the evidence that enables this storytelling is part of the archive's public value. The evidence helps underpin existing narratives about who we are—the 'national stories' as it were.

We are disrupters in the sense that our collections are large bodies of primary evidence—far more evidence than a single person could ever consume. There is extraordinary potential to throw fresh light on an issue, to develop a deeper

understanding, uncover new perspectives, challenge orthodoxies, and tell new stories through the evidence an archive holds. That potential has significantly expanded through digital tools, capable of processing vast amounts of information—and it is growing as we digitise more.

Repetition in a socially-forming sense is foregrounded by a 'national' archive. Repetition plays a role in all three of its primary purposes—guiding appraisal and selection decisions about what records are kept; preserving records; and managing access. These are all being profoundly challenged and changed by digital technologies.

The shift in form of the record from the physical to the virtual is a momentous change for all archives in general. It is a challenge to every part of an archive's practice and work, from appraisal and selection, to transfer, preservation and access. Archives have now been living with the challenges of digital records for several decades. However the temptation to apply paper era assumptions and thinking to digital challenges remains strong, especially when those assumptions still hold true for tangible records. We know that traditional archival thinking—for example, the lifecycle model of record-keeping—cannot carry us through the decades ahead. We see being a digital archive not only as a technological issue, but also as a challenge to archival practices.

The practices of the national archive really matter. As an archive of the state, The National Archives fulfils a unique role in our democracy. To govern, the state has to collect information: Individuals, organisations, society and the economy all need to be legible to the state in some way. In particular, the executive gathers data so that it can understand what is going on and intervene, by changing laws or taxes or spending money. The state

✉ Clare Foster  
clarefoster@gmail.com

John Sheridan  
john.sheridan@nationalarchives.gov.uk

<sup>1</sup> The National Archives, Kew, Richmond, UK

<sup>2</sup> Centre for Research in the Arts, Social Science and Humanities (CRASSH) and Faculty of History, University of Cambridge, Cambridge, UK

produces records of its actions, the decisions that led to those actions, and the analysis, options and evidence underpinning those decisions. The resulting records might be laws, or a wide variety of administrative documentation. Transparency is important to gaining and retaining public trust. In turn, public trust confers legitimacy on the state's institutions. The archive of the state is part of this infrastructure of transparency, trust and legitimacy.

The national archive's next job is to help make the state more legible to its citizens—albeit often only after some time has passed. Firstly, by providing access to the records, so everyone can see the evidence of the state's decision-making. The records tell us who was involved and why decisions were made. Secondly, the records the state amasses provide an opportunity to look back, to see the past through the state's eyes. This is always an instructive perspective, however partial, selective, subjective or biased at the time.

The seminal example of this is the Domesday Book, one of the earliest and most important records we hold at The National Archives. The Domesday Book is essentially a data set - the first national data set. This survey was carried out not long after the Norman Conquest. England was divided into areas and or circuits, and royal commissioners were appointed to each. A series of questions were asked, the information was laboriously distilled, organised and encoded, then made available in a certain format. For hundreds of years the resulting book was taken to represent something important—ownership of land. But questions about what the book now represents are complicated by several things. First, the passage of time (we need context to understand the value of any information); second, by the gap between who the book was for at the time, and what meanings it could take on for future generations.

These questions are ones the digital archive particularly finds itself in the business of managing. Archivists play a key role in determining what the archive represents through two main processes, which they are bound to influence or decide: selection, and category (or code).

Even in the case of the Domesday Book, before the questions could be asked, there first had to be agreement about categories: for example what counted as a “villain”? Categories not only shape and describe contents, and direct readers' attention—they also determine what stories can be told by the material, now and in the future. This influence is different in a digital context, where categories are chosen based on likely end-user search terms, rather than deriving from patterns in the content itself.

*CF: Yes, and how you categorise becomes a reifying event, becomes the thing itself. That's a lot of power.*

*JS: Archives are agents of the powerful, a national archive overtly so, as we serve the government of the day. We are conditioned by our context and that will bring inevitable*

*biases. Nevertheless, objectivity and impartiality are values we have always striven for.*

The National Archives dates back to the 1830s and the founding of the Public Records Office. The institution has evolved alongside what today we might call ‘archive science’. The archive of paper records was a set of practices framed and codified in the early 1920s by the great civil servant archivist Sir Hilary Jenkinson. Undoubtedly an establishment figure and a product of his time, he ran the Public Records Office for many years, and in 1922 wrote the *Manual of Archival Administration*, which set out principles and practices still recognisable in our archive today. And Jenkinson's viewpoint started from the notion of objectivity of the archival record. He recognised the importance of provenance, but he also put a lot of emphasis on custody, intellectual control, and what he called “the moral defence of the record”, conceiving the archivist as an impartial custodian and guardian.

In Jenkinson's paradigm, it is for record-creators to decide what has value and should be kept, not archivists. His legacy still has huge influence in our profession, and his ideas are still quite deeply embedded in the law around public records and the processes for deciding what is kept. Yet in a digital age, even our conception of the record creator is challenged. Consider synthetic content. Essentially any type of digitally-encoded information humans might create, from the sublime (poetry, music) to the prosaic (meeting minutes), can now be synthesised by a machine using an AI deep network. Through this process information is being endlessly re-versioned, re-cycled and re-purposed. Who is the creator? Who is left to decide what has value and should be kept? The scale of this transformation means we are in a different era.

That said, even with Artificial Intelligence, a digital archive still has to be organised in some way—arranged, sorted and labelled—if nothing else, to make it functionally available to potential future users.

*CF: And the gatekeepers for that process are now not experts or authorities, but end-users?*

*JS: This is another aspect to both enabling and disrupting repetition. The World Wide Web has given archives a scale of audience, reach and relevance that was unimaginable 30 years ago. Archives are and should be for everyone. Thanks to the web, far more people can engage with archival collections. The public can find and access the evidence for themselves, unmediated. Of course, in reality we have simply gained new mediators to our collections, not just historians or journalists, but search engines like Google, social media channels like Facebook (both of course, underpinned by new forms of repetition and advertising-based business models) or Wikipedia editors.*

*In the old days, there were index cards describing the contents held. These were really designed for historians or researchers to find their way around. Now we have to decide*

what categories members of the public in the future might want to search for, so we have to think in terms of what our users might already think, or know. Which people, events, and so on might people want to know more about? So our categories follow that logic, rather than only describing contents. The decisions we make are coming more from what we think people might already be looking for than from what is in the material.

This can be a disincentive to discovery, to the preservation and repetition of accidental, or incidental information, which is where some of the greatest value can lie. The records that we have for example from the nineteenth century operation of the Poor Law provide hugely valuable insights into the lives of the poorest people in our society at that time, as do the records of the state's oppression of gay, lesbian, bi- and transsexual people in the early 1920s, with things like the raids on the Caravan Club in Soho. These materials offer huge insight into the lives of people at those times.

*CF: Andrew Prescott was saying something similar in his discussion with you as part of the series convened by Anne Alexander at Cambridge Digital Humanities called 'Re-Reading the Archive': that these issues of selection and category are profoundly changed by digitisation, giving archivists a new creative, or if you like, persuasive role.*

*JS: The National Archives come to the issue of persuasion from a very particular perspective. As an archive of the state, we are a government department. We are Civil Servants. That means serving the government of the day, whatever its political persuasion, to the best of our ability, bound by values of integrity, honesty, objectivity and impartiality. The civil servant archivist is not an overt persuader.*

However, given the scale of digital information-, more informationally-vast yet at the same time more physically compact - decisions about the selection or destruction of material remain fundamental. The act of digitising any archive is heavy with significance and consequence. There are key decisions to be made around what records are chosen to be kept, and who does the choosing. What records are sufficiently valuable to be kept? Who or what makes those evaluative decisions? The archivist has always been a mediator, facilitating access to the collections, but one of the big changes in the digital era is that the archivist is undeniably an active decision-maker in this question of what evidence the future will get to have of 'now'.

*CF: So people might think end-user-driven search categories are more democratic than various authorities deciding what gets saved, and under what headings, but it isn't necessarily so?*

*JS: Yes. And for me, this speaks to two things of central importance that archives and archivists need to cling onto, as the challenges of digital record-keeping collapse so much*

of our previous practice. The first is to foreground the value of the archive as a place where you can answer the question 'how do we know?'; and the second is the joy of discovery, finding something new, or learning something that is new to you (the joy of discovery can be something new to a person, not necessarily new to the world).

Our primary schools outreach programme focuses on these two goals. The 5 and 6-year-olds who visit us at The National Archives at Kew, for example, have almost all heard about the Great Fire of London. So we gather them round and ask them to tell us how the fire started. Pretty soon they mention the main elements: there was a baker, called Thomas Farriner, who had a bakery in Pudding Lane, and there was a fire in his oven, and the city burned down. Then we ask the children how they know those facts. They say things like their teacher told them, or their brother told them, or their mum. Then we point out that their teacher, brother and mum weren't alive in 1665: they weren't first-hand witnesses of those events, so how did *they* know? And you can see the realisation flickering through these young minds, to the point where they start wondering how anybody knows anything. And that's when we introduce the idea of the record. We tell them we have a tax return for Pudding Lane that has the name Thomas Farriner, baker, on it. We have evidence that this man existed, that he was a baker, and that he was doing this activity at that time. And the children get very excited as they realise that now they know *why* they know anything.

*CF: That's a helpful reframing of what an archive is, a place where you find out not what you know, but why you know.*

*JS: Providing the why we know, through keeping primary sources of evidence, is exactly how the digital archive can provide most value.*

However, access is problematic for us in an increasingly complex information rights landscape: there's a world of difference between making something available for inspection in a reading room, and publishing it on the web.

We also have obligations to the subjects of records, as well as to the users of records. These sorts of changed conditions are forcing archives to rapidly re-develop our practices in order to be able to cope with a fundamental shift at the heart of our business, a shift that touches on our role as narrative-shapers for future generations. The archivist is now inescapably a decision-maker, actively involved in curation, in issues of gathering and keeping evidence, whose future use is much more problematic than it used to be. Someone needs to do the keeping of our collective memory, and someone needs to pay for the recordkeeping. But under what framework of rights or privileges? It is by no means certain

that the record-keeping of the future will be done by the memory institutions of today.

We may see radical disruption, and very different kinds of record-keepers keeping records for very different kinds of reasons. If the archive is to retain legitimacy and maintain its role as the collective record-keeper, how does it balance participation, privileges and obligations to the fundamental right of people in both the present and the future to *know*? There's never been a more exciting time to be working in the field or a more relevant time to be asking and answering some of these questions.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

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