Letters to the Editor

THE ARCHIVAL WORLD

Volume:8 - Issue:2
2021/Winter

DISRUPTIVE DIGITAL ARCHIVES

Archives of the state fulfil a unique role in society. In order to govern the state collects information about us. It produces records of its actions, the decisions that led to those actions, and the analysis, options and evidence underpinning those decisions. Society and the economy needs to be legible in some way to the state, so that the government can intervene through law making, taxation, or spending. This has long been the case. It is not an accident that the English word "statistics" is derived from the Latin *statisticum collegium* or council of state. One of the most notable records held by The UK National Archives is the Domesday Book – an extensive survey of property ownership in England, compiled shortly after the Norman Conquest in the 11th century. It provides a unique insight into who owned what, before and after the conquest. The Domesday Book is an effort to make Saxon and then Norman England legible to its new rulers.

If as citizens we are read in some way by the state, through the information it holds about us, it falls to the archive of the state to turn the tables in favour of the citizen. The state becomes legible to its citizens through the archive. We help provide people with two insights. Firstly, through the records of government we can see who was involved in decisions, how and why those decisions were made. Secondly, through the records the state amasses, we can look at society through the state's eyes. We can see what the state saw in previous times. It is always an instructive perspective, however partial, selective or biased that viewpoint might be. Through the state's eyes we can find out about the past, including the lives of our family ancestors.

The digital transformation of the state over the last four decades poses some major challenges for state archives, like The National Archives in the UK. New forms of computational use of archival collections provide both opportunity and risk. It can be disorientating for the archivist to know how to respond in the face of so much rapid change.

Archivists have been living with the various challenges posed by digital records for some time. It has long been apparent that archival thinking of the Jenkinson era (which predates electronic computers by some decades) cannot carry us through the decades ahead. At The National Archives, Jenkinson's archive so to speak, we see the issue less in technological terms and more as a challenge to archival practice. That perspective motivates our strategic aim to become the disruptive digital archive.

It is not new thinking. In the late 1980's the archival theorist David Bearman wrote "*Traditional methods employed in archives for appraisal, description, preservation, and access to records fail to meet archival needs, because the demands exceed the capacity of the profession - and by more than an order of magnitude*"[1]. Bearman advocated radical post-custodialism.

In the years since, custody of digital records has proven itself a vital part of the value proposition for the digital archive. Today custody matters perhaps more than ever, with

[1] Bearman, D. (1989). Archival Methods. *Archives and Museum Informatics* 3, 28.

Geliş Tarihi/Received: 10.12.2021
Kabul Tarihi/Accepted:28.12.2021

archival storage increasingly moving to the cloud. By custody we mean that the archive takes active responsibility for managing the risks to the records over time, wherever they are physically stored.

On the whole the risks to digital records are quite different from their physical counterparts. In some ways there is less risk – it is relatively cheap and easy to make and keep copies of data files. Indeed, digitisation is a good strategy for preserving deteriorating papers, as well as enabling wider access. In other ways there is much more risk – you have to know what you have, know what capability you need to realise the informational value of what you have, and keep making copies over time – being sure not to inadvertently corrupt something along the way. In an increasingly risky cyber security environment the archive must keep its digital collection safe from harm (malicious attacks from ransom ware, viruses and Trojans and accidental damage through human error) whilst making active interventions to reduce other types of preservation risk.

Data does not keep itself. The computer systems we all use to create, store and manage data are highly complex and increasingly so. Some of the risks to digital records are well known, others are emergent. As hardware and software becomes obsolete the risk grows that we lose the ability to access and use data from the past. With technology changing very rapidly, the time periods involved can be very short. This is a problem for everyone and a particular problem for the digital archive, with our long-term horizons.

Digital preservation involves actively managing the risks to digital records over time. There are two aims. Firstly to ensure the archive knows, and will continue to know, what digital assets it has and, given that The National Archives' assets are public records, knows what the records are evidence of (who produced the record? and what were they doing at the time?). This is intellectual control. Secondly, to ensure the informational value of the record can be reliably rendered or produced, so this value can be experienced by users of the archive. This is renderability. Overall, we view digital preservation risk as the combination of the risks to renderability and the risks to intellectual control.

There are many risk factors which affect renderability and intellectual control. To synthesise the evidence around digital preservation risk, The National Archives has developed and published a digital preservation risk model using a Bayesian Network, called DiAGRAM, the Digital Archiving Graphical Risk Assessment Model[2]. This has been jointly developed by the Digital Archiving team at The National Archives with the Applied Statistics and Risk Unit at the University of Warwick and over a dozen experts in digital preservation from a variety of other archives. The DiAGRAM model draws on and develops concepts from other schemes such as the NDSA Levels of Digital Preservation[3] and the Digital Preservation Coalition's Rapid Assessment Model (RAM)[4].

DiAGRAM can be used to measure the level of digital preservation risk. It also allows the archivist to model different scenarios for intervention, and compare the results in terms of the risk. Its main value is as a decision support tool. Given the archive has a finite amount of resource, and the range of possible interventions the archive might take are very broad, digital preservation is about resource utilisation to maximum effect; selecting the most impactful interventions from a range of possibilities.

[2] https://nationalarchives.shinyapps.io/DiAGRAM/
[3] https://ndsa.org/publications/levels-of-digital-preservation/
[4] https://www.dpconline.org/digipres/dpc-ram

The risk landscape the archive operates in continues to evolve. There are emergent threats. For example, new techniques in Artificial Intelligence have opened the door to the convincing synthesis of almost any form of content. The issues and consequences of "fake news" or "fake history" are mainstream topics of conversation.

The disruptive digital archive looks for new techniques to meet these challenges. That was the motivation of our research project, Archangel, which looked at using distributed ledger technology (also known as blockchain) to authenticate records held in the archive, over long timespans. We researched the prospects of creating content fingerprints, that we might write to the ledger, for content like video records, that we are likely to re-encode when rendering or producing the record.

Looking ahead, there is no shortage of work for the disruptive digital archive to do. There are two grand challenges that we need to grapple with. The first is environmental sustainability. With humanity's net data holdings continuing to grow rapidly, selecting important digital content to keep for posterity and passing it to the archive is environmentally friendly. We can maximise the benefit by having much greater regard to the energy consumption of the digital archive. This means being clear that the energy we are consuming is needed in terms of mitigating preservation risk.

The second is around access in the web era. Archives need to be used to be useful. Yet the way we organise and present our collections is far from most people's expectations, shaped by using search engines or asking Alexa. Archival collections are also full of information about people and their lives. We need to find ways of enabling use of digital archives in the public interest, whilst respecting privacy, mitigating the risks of harm and retaining public trust in the archive.

We all have so much to do. There has ever been a more interesting, exciting or important time to be a disruptive digital archivist.

John SHERIDAN
Director of Technology, The UK National Archives
john.sheridan@nationalarchives.gov.uk
*https://orcid.org/*0000-0001-8578-3857