

Research guide: digital resources in the humanities

Dr Hannah Smyth

1. Reflect on the kinds of decisions that must be made by institutions, companies and individuals who build digital archives
2. Reflect on how those decisions shape digital archives and the kinds of questions that can and cannot be asked of them
3. Examine how and whether the makers of digital archives inform the users of digital archives about the selection rationale they have used
4. Identify the different kinds of partnerships (e.g. public-private) that often underpin digital archives
5. Go beyond question of what the archive is – what does it do? How is it being used? Impact?

1. Decisions and how they are shaped:

- (All interrelated)
- What or how much to include and digitise / to exclude?
- What is its purpose? / How will it be used?
- Who will use it?
- What format/functionality?
- Is there a specific timeframe for delivery? Dictated by who/what?
- **Context:**
 - Is the digitiser or digital resource creator a public or cultural heritage institution / academic / commercial enterprise / community / society?
 - Political contexts
 - National; international; post-colonial; settler-colonial; post-conflict
 - Historical
 - prehistory of the digital archive
 - history of the institution or company and their track record on access, ethics and digital projects
 - Social & cultural context
 - open / closed information culture
 - attitudes towards things like access, rights in records, open government, freedom of information?
 - significance of 'heritage'
 - Academic / intellectual context
 - 'public history' or 'history from below'
 - decolonial practice
 - historiography/historiographical debates
 - Commemorative/memorial
 - timing; funding; political context; national cultural heritage; popular or 'high value' collections
- Intended or expected use/purpose:
 - Who/what will this resource serve? E.g.:
 - Genealogical tool

- diaspora
 - Academic research
 - Educational (primary/secondary/general public)
 - Justice / evidentiary / healing
 - Rights-based – memory, identity
 - Restorative justice
 - Reparation – displaced archives
 - Peace and reconciliation
 - Community building, engagement, collaboration
 - Commemorative
 - Law enforcement
 - Open Government
 - Private enterprise
- **Financial:**
 - Institutional budgets
 - Governmental heritage/cultural budgets - national economy – prosperous? crisis? recovering economy?
 - External funding / donations – who and what influence?
 - Usage and social media reporting – link between demonstrable use and funding justifications
 - Labour – conservation, cataloguing, staffing levels
- **Legal:**
 - National legislation
 - Who donated?
 - When?
 - Under what circumstances?
 - What stipulations regarding use / - re-use permissions?
 - Sensitive information – risk management
 - Copyright restrictions?
 - Digitised but *onsite* digital access only?
 - Re-use permissions
- **Selection rationale:**
 - Possibly indicated in administrative reporting or other documentation
 - Usually the general process rather than the specifics
 - Selection criteria?
 - can be implicit and/or explicit
 - see for example: Hauswedell, T., Nyhan, J., Beals, M. H., Terras, M., & Bell, E. (2020). Of global reach yet of situated contexts : an examination of the implicit and explicit selection criteria that shape digital archives of historical newspapers. *Archival Science*, 20, 139–165. <https://doi.org/10.1007/s10502-020-09332-1>
 - ‘Significance’
 - What significance or value frameworks are applied (implicit or explicit)?
 - see for example: Russell, R., & Winkworth, K. (2009).

Significance 2.0: a guide to assessing the significance of collections (2nd ed., Vol. 370, Issue 8362). Collections Council of Australia Ltd.

- **Use of technologies:**
 - Format and functionality
 - Modes of search and discovery, of data manipulation
 - Use of AI, Machine Learning, Linked Open Data techniques etc.
 - Mordell, D. (2019). Critical questions for archives as (big) data. *Archivaria*, 87(Spring 2019), 140–161.
 - Data available for analysis (e.g., via an open API)
 - Is it Open Source (licence to use, study, change, and distribute software and its source code to anyone and for any purpose)?
 - Algorithmic bias?
 - Risk or harm mitigation?
 - See for example: Almeida, D., Shmarko, K. & Lomas, E. The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: a comparative analysis of US, EU, and UK regulatory frameworks. *AI Ethics* (2021). <https://doi.org/10.1007/s43681-021-00077-w>
- **Ethics:**
 - Who are the record creators or owners?
 - Where do the records / data come from?
 - Is the data collection / digitisation / access-level appropriate?
 - What archival description / descriptive practices?
 - Is the use of emergent technologies appropriate or problematic? Is there transparency about their use?
 - How are ethical considerations accounted for? (transparency and accountability)
- **Partnerships:**
 - Inter-institutional
 - International
 - University research centres / university archives
 - Learned Societies
 - Educational
 - State / State commemorative
 - Participatory or Collaborative
 - public memorabilia projects, public collection days, collective metadata and description creation
 - public transcription projects
 - community partnerships
 - Private enterprise, e.g., genealogical archives, newspaper archives e.g. Gale, Ancestry, FindMyPast etc.
- **Financial:**
 - Institutional budgets
 - State/Governmental heritage/cultural budgets - national economy – prosperous? crisis? recovering economy?

- External funding / donations – who and what influence?
- Usage and social media reporting – link between demonstrable use and funding justifications
- Labour – conservation, cataloguing, outsourcing, partnerships
- **Legal:**
 - National archives legislation
 - Data protection legislation (varies across jurisdictions)
 - Who donated/contributed?
 - When?
 - Under what circumstances?
 - What stipulations regarding use and re-use permissions?
 - Sensitive information – data protection and risk management
 - Balancing rights in records with data privacy/data protection
 - Copyright restrictions?
 - Digitised but *onsite* digital access only (e.g., national web archives)?
 - Re-use permissions

2. How to research these:

How do we address these different frameworks?

Where do we look, what can we know of them?

Almost a process of ‘reverse engineering,’ working back from the finished product.

Suggested reading:

Mak, B. (2014). Archaeology of a digitization. *Journal of the Association for Information Science & Technology*, 65 (8), 1515–1526. <http://10.0.3.234/asi.23061>

Jensen, H. S. (2020). Digital Archival Literacy for (all) Historians. *Media History*. <https://doi.org/10.1080/13688804.2020.1779047>

- **Context of collection/archive/digitisation/institution/organisation:**
 - History/context in which the archive/collection/resource originated or was created
 - Information provided through website, exhibition texts, research pages, FAQs, associated publications, news media
 - Published collections guides, where they exist
 - Catalogue search, acquisition records
 - Debates or controversies around their creation/collection/digitisation/access/use
- **Interface affordances or conditions of use:**
 - Curation / presentation / interpretation / functionality
 - Can we trace the intended use/usability from the resulting digital archive or digital resource?
 - Bespoke website/microsite or integrated to regular catalogue
 - Database/browsing/exhibition-style
 - Searchability/findability/readability – file formats and viewing functions
 - Metadata and description

- **Paradata:**
 - Operate on the assumption that a lot of the information about the constitution of a digital archive/data repository may not be found within the digital archive itself, and may be difficult to discern
 - Administrative decisions and contextual data
 - see for example: Fyfe, P. (2016). An Archaeology of Victorian Newspapers. *Victorian Periodicals Review*, 49(4), 546–577. <https://doi.org/10.1353/vpr.2016.0039>
 - University or research heavy projects usually provide much more technical detail and have descriptive research outputs
 - Amount of data from public reporting varies from institution to institution
- **Annual reports and reviews**
 - Often found in 'Freedom of Information' or 'Policy' sections of a website
 - Institutional activities: planning and implementation of project
 - Financial reporting
 - Social media policy
 - Digitisation or digital imaging policy
 - Strategic planning reports
 - Partnerships and collaborations
 - Online usage reporting – website analytics
 - Outreach activities
 - Social media analytics reports
 - Bequests/donations ; ringfenced funds/stipulations of use
 - Reporting on labour involved in creating a digital archive or resource e.g., temporary project-specific hires, timeline of the project
- **Website versioning**
 - Web archive collections/databases e.g., *Internet Archive WayBack Machine* and national web archive collections
 - Some documents may only be available in archived versions of a website
- **Associated media**
 - Is it highly mediatised? What does this tell us about its perceived/ascribed value?
 - News media, documentaries, social media
 - Third-party media or linked resources
 - Press releases, blogs
 - Parliamentary debates and select committee reports, departmental reports (usually available online)
- **Literature search:**
 - Published research on/about the digital resource or archive (archival, digital humanities, digital heritage journals etc.). Usually more available for university projects.