Towards International Perspectives on Collection Data Infrastructure Development

Marco Humbel[1], Nina Pearlman[2], JD Hill[3], Andrew Flinn[1], and Julianne Nyhan[1,4]

[1]Department of Information Studies UCL, London [2]UCL Art Collection, London [3]The British Museum [4]Technische Universität Darmstadt

This paper brings together international perspectives on the interests and needs of libraries, archives and museums ('collecting and heritage organisations') in collection data infrastructure developments. To date a typical model for digital infrastructures in the cultural heritage sector are national data aggregation platforms, such as Trove, the Deutsche Digitale Bibliothek (German Digital Library), the Digital Public Library of America or Japan Search (Paltrinieri, 2021: 4–7). Given the recent surge of investments in building and enhancing cross-border digital infrastructures, such as the Common European Data Space for Cultural Heritage[1], this presentation makes a timely contribution by unpacking the latent challenges in international collection data infrastructure development. We ask: What are the experiences of collecting and heritage organisations of participating in national and international digital infrastructure projects? Which factors enable and impede collecting and heritage organisations in unifying siloed collections, and how do these factors differ between countries? What can be learned from national collection data infrastructure programmes, like the UK's Towards a National Collection (TaNC) programme, for connecting digital collections internationally?

After decades of mass digitisation in the cultural heritage sector funders seek to find the most "suitable infrastructure components and methods" (Ahnert et al., 2023: 23–24) for reaping the perceived benefits of virtually unified collections and data at scale. The UK Arts and Humanities Research Council (AHRC) envisions for instance that its £18.9 million TaNC programme "[…] will allow researchers to formulate radically new research questions, increase visitor numbers, dramatically expand and diversify virtual access to our heritage, and bring clear economic, social and health benefits to communities across the UK" (TaNC, 2023). From 2021 to 2023 the Australian Research Data Commons (ARDC) invested parts of a 8.9 million AUD grant to improve data driven access via the Trove aggregation platform for humanities and social science researchers (ARDC, 2023a; ARDC, 2023b). In 2022 the European Commission awarded the Europeana foundation with a multi-million service contract for deploying the Common European Data Space for Cultural Heritage, which is imagined to become in conjunction with other data spaces (e.g. for health or agriculture) "a genuine single market for data, open to data from across the world" (European Commission, 2020: 4–5).

Key stakeholders of these programmes are collecting and heritage organisations as they are asked to mobilise collection data at scale for digital infrastructures.

---

[1] See: https://digital-strategy.ec.europa.eu/en/news/commission-proposes-common-european-data-space-cultural-heritage (accessed 09 December 2023).

Accordingly collecting and heritage organisations were of interest in a number of international surveys to benchmark their progress in digitisation an data dissemination (Nauta et al., 2017; McCarthy and Wallace, 2018; Estermann, 2018). Qualitative research about collecting and heritage organisations' experiences in respect to digital infrastructure development took mostly a nation-specific focus however, with attention given to small and independent organisations. Michelle Caswell and Bergis Jules explored for instance the perspectives of US-American community archives for becoming part of a 'National Digital Platform' (2017). In the context of TaNC Gosling et al. facilitated focus group discussions with representatives of small organisations on their capacities to join a UK collection data infrastructure (2022). To date there is a lack of qualitative research which compares the experiences, perspectives and needs of collecting and heritage organisations regarding digital infrastructure development internationally.

Due to the limited availability of qualitative studies exploring collecting and heritage organisations' perspectives on digital infrastructure development internationally we identified semi-structured interviews and focus group discussions as appropriate data collection methods. These qualitative data collection methods consolidate as invaluable instruments for contextualising and extending information which is otherwise not available to the public at all, or dispersed across an abundance of grey literature and technical reports (Hauswedell et al., 2020: 140). We identified interviewees through a "key knowledgeable" sampling strategy. The approach is a purposeful sampling strategy "[…] to create a group of cases that provide information-rich data-gathering and analysis possibilities" on "highly specialized" subjects areas (Patton, 2015: 405; 408–09), such as collection data infrastructures. Our sample consists of national institutions, university collections, community archives and regional heritage organisations. The UK, Germany and Australia are of focus in our study because of the described major funding programmes running in these regions now, and our language expertise. All interviews and focus groups were audio recorded and transcriptions returned to interviewees for approval.[2] We analysed the transcriptions following the 'Miles, Huberman and Saldaña method' in iterative cycles of thematic and In Vivo coding (Saldaña, 2016: 102–10; Miles et al., 2014). Table 1 gives an overview on the data collection method and sample size for each of the countries we surveyed.

Country	Data Collection Method	Location	Data Collection Period	Sample Size
United Kingdom	Semi-structed interviews	Online	June to July 2022	15 individuals in 8 organisations
Germany	Focus group discussion	Sloane Lab Knowledge Exchange Event Europe, hosted by Technische Universtät Darmstadt	September 2023	10 individuals in 7 organisations
Australia	Focus group discussion	Researching the Future of Museum Collections Symposium hosted by Deakin University Melbourne	November 2023	8 individuals in 7 organisations

Tabel 1 Overview on data collection method and sample size per country

Throughout our consultations we observe that collecting and heritage organisations' capacity to participate in digital infrastructures is dependent on a complex interplay of resource allocation across the heritage sector and within collecting and heritage organisations, including divergent traditions of collection description, and disciplinaries idiosyncrasies. Accordingly, we call for better social-cultural and trans-sectoral (Bowker et al., 2009: 100–01) understandings of collection data infrastructure development. Latent issues in collections as data infrastructures development include:

- Heterogeneous and unsustainable funding structures for establishing digital infrastructures.
- A lack of understanding of disciplinary conventions' impact of for framing, describing and using collection data.
- 'Information flattening' through digital infrastructures due to the lack of methods for representing tacit knowledge and interpretation about collections.
- The need for situated ethical frameworks to address the affordances emerging from collections' histories, cultural functions and significance.

Acknowledgements

References

Ahnert, R., Griffin, E., Ridge, M. and Tolfo, G. (2023). *Collaborative Historical Research in the Age of Big Data: Lessons from an Interdisciplinary Project*. (Cambridge Elements: Elements in Historical Theory and Practice). Cambridge: Cambridge University Press doi:10.1017/9781009175548. https://www.cambridge.org/core/product/identifier/9781009175548/type/eleme nt (accessed 4 May 2023).

ARDC (2023a). Trove Enhancements https://ardc.edu.au/project/trove-researcher-platform-for-advanced-research/ (accessed 1 September 2023).

ARDC (2023b). HASS and Indigenous Research Data Commons https://ardc.edu.au/program/hass-rdc-indigenous-research-capability/ (accessed 1 September 2023).

Bowker, G. C., Baker, K., Millerand, F. and Ribes, D. (2009). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In Hunsinger, J., Klastrup, L. and Allen, M. (eds), *International Handbook of Internet Research*. Dordrecht: Springer Netherlands, pp. 97–117 doi:10.1007/978-1-4020-9789-8_5. http://link.springer.com/10.1007/978-1-4020-9789-8_5 (accessed 11 November 2021).

Caswell, M. and Jules, B. (2017). *Integrating Community Archives into a National Digital Platform: Challenges, Opportunities, and Recommendations*. (A White Paper Reporting on the 2016-2017 "Diversifying the Digital Historical Record" Forums) https://escholarship.org/uc/item/8r10h3tw (accessed 7 December 2021).

Estermann, B. (2018). Development paths towards open government - an empirical analysis among heritage institutions. *Government Information Quarterly*, **35**(4): 599–612 doi:10.1016/j.giq.2018.10.005.

European Commission (2020). Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and the Committee of the Regions: A European strategy for data https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066 (accessed 8 November 2023).

Gosling, K., McKenna, G. and Cooper, A. (2022). *Digital Collections Audit*. Zenodo doi:10.5281/ZENODO.6379581. https://zenodo.org/record/6379581 (accessed 23 March 2022).

Hauswedell, T., Nyhan, J., Beals, M. H., Terras, M. and 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**: 139–65 doi:https://doi.org/10.1007/s10502-020-09332-1.

McCarthy, D. and Wallace, A. (2018). Survey of GLAM open access policy and practice https://docs.google.com/spreadsheets/d/1WPS-KJptUJ-o8SXtg00llcxq0IKJu8eO6Ege_GrLaNc/edit#gid=1216556120 (accessed 27 February 2022).

Miles, M. B., Huberman, A. M. and Saldaña, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook*. Third edition. Thousand Oaks, Califorinia: SAGE Publications, Inc.

Nauta, J.G., Van Den Heuvel, W., Teunisse, S. and DEN Foundation (2017), *Europeana DSI 2–Access to Digital Resources of European Heritage: D4.4.Report on ENUMERATE Core Survey 4*. Europeana https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/ENUMERATE/deliverables/DSI-2_Deliverable%20D4.4_Europeana_Report%20on%20ENUMERATE%20Core%20Survey%204.pdf (accessed 11 February 2022).

Paltrinieri, C. (2021). *International Benchmarking Review: A Towards a National Collection Report*. (Towards a National Collection Directorate Report) doi:10.5281/ZENODO.5793173. https://zenodo.org/record/5793173 (accessed 16 February 2022).

Patton, M. Q. (2015). *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*. Fourth edition (e-book). Thousand Oaks: SAGE Publications, Inc.

Saldaña, J. (2016). *The Coding Manual for Qualitative Researchers*. Third edition. Los Angeles: SAGE.

TaNC (2023). About us https://www.nationalcollection.org.uk/about (accessed 7 December 2023).