

## Socio-cultural challenges in collections digital infrastructures

Marco Humbel<sup>1</sup>, Julianne Nyhan<sup>1,2</sup>, Nina Pearlman<sup>3</sup>, Andreas Vlachidis<sup>1</sup>, JD Hill<sup>4</sup> and Andrew Flinn<sup>1</sup>

<sup>1</sup>Department of Information Studies UCL, London <sup>2</sup>Technische Universität Darmstadt <sup>3</sup>UCL Art Collection, London <sup>4</sup>The British Museum

Published in: Journal of Documentation

DOI: 10.1108/JD-12-2023-0263

Date of acceptance: 17-Jul-2024

License: CC BY-NC 4.0

### Abstract

**Purpose:** This paper explores the accelerations and constraints libraries, archives, museums, and heritage organisations ('collections-holding organisations') face in their role as collection data providers for digital infrastructures. To date digital infrastructures operate within the cultural heritage domain typically as data aggregation platforms, such as Europeana or Art UK.

**Design/methodology/approach:** Semi-structured interviews with 18 individuals in eight UK collections-holding organisations and two international aggregators.

**Findings:** Discussions about digital infrastructure development often lay great emphasis on questions and problems that are technical and legal in nature. Important as technical and legal matters are, more latent, yet potent challenges exist too. Though less discussed in the literature, collections-holding organisations' capacity to participate in digital infrastructures is dependent on a complex interplay of funding allocation across the sector, divergent traditions of collection description, and disciplinary idiosyncrasies. Accordingly, we call for better social-cultural and trans-sectoral (collections-holding organisations, universities, and technological providers) understandings of collection data infrastructure development.

**Practical implications:** Sustainable financial investment across the heritage sector is required to address the discrepancies between different organisation types in their capacity to deliver collection data. Smaller organisations play a vital role in diversifying the (digital) historical canon, but they often struggle to digitise collections and bring catalogues online in a first place. In addition, investment in existing infrastructures for collection data dissemination and unification is necessary, instead of creating new platforms, with various levels of uptake and longevity. Ongoing investments in collections curation and high-quality cataloguing are prerequisites for a sustainable heritage sector and collection data infrastructures. Investments in the sustainability of infrastructures is not a replacement for research and vice versa.

**Research implications:** The authors recommend developing more understanding of the social-cultural aspects (e.g., disciplinary conventions) and their impact on collection data

dissemination. More studies on the impact and opportunities of unified collections for different audiences, and collections-holding organisations themselves are required too.

**Social implications:** The authors recommend establishing networks where collections-holding organisations, technology providers and users can communicate their experiences and needs in an ongoing way and influence policy.

**Originality:** To date, the research focus on developing collection data infrastructures has tended to be on the drive to adopt specific technological solutions and copyright licensing practices. This article offers a critical and holistic analysis of the dispersed experience of collections-holding organisations in their role as data providers for digital infrastructures. The paper contributes to the emerging understanding of the latent factors that make infrastructural endeavours in the heritage sector complex undertakings.

## 1. Introduction

This paper explores the accelerations and constraints libraries, archives, museums, and heritage organisations ('collections-holding organisations')[1] face in their role as collection data providers for digital infrastructures. To date such infrastructures operate within the cultural heritage domain typically as data aggregation platforms of national, international, or domain-specific scope. Examples are for instance Japan Search, Cultura Italia, Europeana or Art UK (Paltrinieri, 2021). We address the following research questions:

1. What are collections-holding organisations' experiences of participating in digital infrastructure projects?
2. Which factors enable and impede collections-holding organisations in unifying siloed collections?
3. Which risks and opportunities do collections-holding organisations perceive in large-scale infrastructure investments, such as the UK's Towards a National Collection (TaNC) programme?

After two decades of investments in mass digitisation projects collections' metadata, digital images of objects, and machine-readable full text documents are available at unprecedented scale. Scholars thus speak about "collections as data – a conceptual orientation to collections that renders them as ordered information, stored digitally that are inherently amendable to computation" (Padilla et al., 2019: 7). Accordingly, funders' attention is moving towards finding the most "suitable infrastructure components and methods" for accessing collection data at scale and, crucially, reaping their potential (Ahnert et al., 2023: 23–24). The literature shows that this potential is being sublimated in different ways by different stakeholders (McGillivray et al., 2020: 11–13). Researchers are asking questions latent to data, or they are using digital collections to develop critical understandings of the capabilities and impact of technology on collections (Van Strien et al., 2020; Beelen et al., 2023). For collections-holding organisations, aggregated mass digitisation opens opportunities to research collection histories and biases in new ways (MacDonald, 2023). In the creative sector digital collections are used for artistic interventions, inspiration, and content production (Terras et

al., 2021). There is also a trajectory within governmental policies to justify mass digitisation investments as a means for boosting the digital economy. For the European Commission heritage collections are for instance “[...] an important contributor to the European economy, fostering innovation, creativity and economic growth [...] and the reuse of such [mass digitised] content can generate new jobs not only in the cultural heritage sector but also in other cultural and creative sectors, including for instance the video game and film industries” (European Commission, 2021: para. 4). To bolster research and the commercial potential of digital collections there is a strong push towards Open Access to cultural heritage from governmental and non-governmental bodies.[2] A prime example is the OpenGLAM initiative which advocates for Open Access to cultural heritage and calls collections-holding organisations to adopt open licenses for metadata and digital surrogates of objects.[3] The OpenGLAM initiative received financial support from the European Commission, the Open knowledge Foundation and Creative Commons (OpenGLAM Initiative, 2023). Similarly, Open Access has increasingly become a condition for receiving funding for digitisation, as in case of the UK National Lottery Heritage Fund for instance.[4] Yet in tandem, some UK governmental policies call collections-holding organisations to attract external funding for mass digitisation by collaborating with commercial enterprises in so-called private-public partnerships. This practice places digital collections either behind paywalls entirely or under restrictive licensing conditions and creates a certain cognitive dissonance for collections-holding organisations who are implicitly called on to pursue two apparently contradictory paths simultaneously (Ahnert et al., 2023: 24–25).

For those who seek pathways to unlocking the perceived latent creative, economic, and cultural value of the mass digitisation activities undertaken in the past decades, digital infrastructures for unifying digitised collections have taken on a decisive significance. Accordingly significant investments in digital infrastructures are being made by players in the governmental-, private-, heritage- and academic-sector. The Australian Research Data Commons invested part of an 8.9 million Australian Dollar programme (2021 to 2023) in the Trove aggregation platform to improve data driven access to collections for humanities and social science researchers (ARDC, 2023a; ARDC, 2023b). In 2022, a consortium led by Europeana was awarded a multi-million Euro service contract by the European Commission for deploying the so-called ‘common European data space for cultural heritage’. The aim is to: expand the existing Europeana aggregation infrastructure and its interoperability with the wider planned data space ecosystem [5]; invest in high-quality digitisation and data enrichment; capacity building and training; and public engagement (European Commission, 2022). Europeana and the data space are also expected to become important dissemination platforms for adjacent digital infrastructures, such as the Horizon Europe funded ‘European Collaborative Cloud for Cultural Heritage’ – a planned cloud infrastructure aimed at heritage professionals for data sharing and leveraging digital research tools (European Commission et al., 2022: 8;24; European Commission, 2024). Within these projects and schemes, aggregations are imagined as spaces to bring together two decades of mass digitised collections and to facilitate cross-search. Yet they are more significant than this: aggregations promise to satisfy the needs and expectations of communities of interest that cut across disciplines, sectors, and society. Thus, aggregation infrastructures are also spaces where uses and boundaries converge and even collide. Given the gaps and omissions in mass digitisation efforts (Zaagsma, 2023; Mak, 2014) aggregation infrastructures will also

shape who and what is remembered, and how digital collections can and cannot be used in future.

The UK context illustrates the scale of investment being put into stimulating access to mass digitised collections through digital infrastructures. At the time of writing in 2023, at least two UK research and Innovation (UKRI) funded schemes are under way. Both have overlapping and potentially competing ambitions: The five year (2019-2024) £18.9 million Arts and Humanities Research Council's (AHRC) TaNC programme seeks to "[...] take the first steps towards creating a unified virtual 'national collection' by dissolving barriers between different collections – opening UK heritage to the world" (TaNC, 2023). The stakes in the programme's deliverables are high. After all, AHRC claims the programme to be "[...] the largest investment of its kind to be undertaken to date, anywhere in the world [...]".[6] Through TaNC, the UK aims to secure a pioneering position in interdisciplinary research by "set[ting] global standard[s]" for the creation of digital collections (UKRI, 2021). But TaNC is not the only major investment seeking to unify collections at scale in the UK. The non-profit company, Open Data Institute (ODI) gained £8 million from UKRI's Innovative UK for setting up a research and development programme "[...] to support innovation, improve data infrastructure and encourage ethical data sharing" (ODI, 2023). As part of this programme the ODI commissioned the UK Collections Trust from 2020 to 2021, to research how museum collections could be "joined-up" and access provided "seamlessly" (Himmelsbach, 2021). Art UK, a major aggregation platform with artworks from over 54,000 artists, takes together with the University of Leicester a leading role in the project. Their endeavours will result in the Museum Data Service (MDS) which is described as a "real world digital infrastructure" which "[...] will act as a data repository for tens of millions of raw object records drawn from UK museums and other public collections" (Art UK et al., 2022). The research presented in this article is conducted as part of 'The Sloane Lab: Looking back to build future shared collections,' which is one of TaNC's five Discovery Projects. In this paper we address however issues about connecting and using collections at scale which go beyond the immediate focus of TaNC.

Given the fundamental role collections-holding organisations are ought to have in these digital infrastructures as those who curate collections and bring them into machine processable forms in a first place, surprisingly little research is concerned with their perspectives and capabilities to do so. A number of surveys benchmark the progress and extent of digitisation and data dissemination in the heritage sector (Nauta et al., 2017; McCarthy and Wallace, 2018; Gosling et al., 2022; Estermann, 2018). Some research has focussed specifically on small and activist-led collections-holding organisations' perspectives on digital infrastructure developments (Caswell and Jules, 2017; Gosling et al., 2022: 29–32; Humbel, 2023). A recent study identified the technical requirements of institutions with collections from the natural sciences in the context of the planned Swiss Virtual Natural History aggregator (Petrus et al., 2023). To date the experiences of collections-holding organisations in their role as data providers for collection data infrastructure development is not well known and is undertheorized. Moreover, information pertinent for such an endeavour is dispersed across an abundance of case-study reports and grey literature (see section two).

To move towards a critical and holistic understanding of collections-holding organisations' perspectives on digital infrastructure projects the Sloane Lab conducted a series of semi-structured interviews with curators, archivists, librarians, and IT specialists who are responsible for collection management, digitisation, and collections aggregation. Digital infrastructure projects have a history of being stalled (see section five). In this paper we demonstrate that if mass digitisation projects' communities of interest want to better understand the potentials, limitations, and implications of collection data aggregation then it is vital to listen to the people who are in collections-holding organisations placed to facilitate digital infrastructure development. Doing so is indispensable to move towards an "infrastructural inversion" of heritage aggregation projects, "[...] recognizing the depths of interdependence of technical networks and standards, on the one hand, and the real work of politics and knowledge production on the other" (Bowker and Star, 1999: 34).

The paper is organised as follows: section two unpacks aspects of the current discourse on digital infrastructure development in the heritage sector that remain underdeveloped. Section three describes methodology; section four is the analysis of interview consultations. Section five responds to the research questions by discussing the findings of the interviews. This paper concludes in section six with a set of recommendations for future digital heritage infrastructure development and research.

## 2. Literature review

In the cultural heritage domain, the term 'aggregator' refers to a system where data of participating collections-holding organisations are centrally collected, structured into a common format (e.g. the Europeana Data Model - EDM), and published on an online platform (Europeana, 2010: 2). In principle 'aggregation' does not signpost a particular technological approach beyond the centralised unification of collections. An aggregator may crawl data through bots or import data through a protocol such as the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).[7] Collections-holding organisations may also pro-actively push data to aggregators through an Application Programming Interface (API) (Collections Trust, 2019: 8–9).[8] Decentralized data aggregation models can be also implemented, involving agencies serving as data aggregators at either the national or domain level. Examples are Cultura Italia or the UK legacy aggregator Culture Grid (see section 5), which bundle digital collections on a national level and push them to their own platform and Europeana.[9] The Sloane Lab is an aggregator too. Data of contemporary and historical catalogue records are converted into the Resource Description Framework (RDF) format and ingested in a central Knowledge Base (Nyhan et al., 2023: 24).

It is well known that a lack of sustainable funding, expertise, equipment, personnel and sector-wide strategies for coordinating digitisation projects impact collections-holding organisations' capacity to digitise (Pandey and Kumar, 2020: 29–30), and by extension to participate in aggregation infrastructures. However, the question of how collections can be unified at scale is mostly discussed as a question that can be solved by finding the appropriate technology and legal policy. The research agenda of TaNC's Foundation Projects had for instance a strong emphasis on the potential of technologies and standards such as International Image Interoperability Framework (IIIF), Persistent Identifiers (PIDs) and Linked

Open Data (LOD) for creating aggregated collections (Padfield et al., 2022; Winters et al., 2022; Kotarski et al., 2022); and means for enhancing collections discovery through Artificial Intelligence (AI) or geo-spatial information (Rees et al., 2022; Angelova, 2021). Another strand of TaNC looked at the implications of restrictive licensing schemes for reusing collections and copyright management at scale (McNeill, 2022; Wallace, 2022). TaNC’s research emphasis [10] is symptomatic of the core challenges mass digitisation and subsequent aggregation projects keep grappling with: the harmonisation of intellectual property licensing practices as well as developing interoperable technical tools, standards, and systems (Thylstrup, 2018: 66–67).

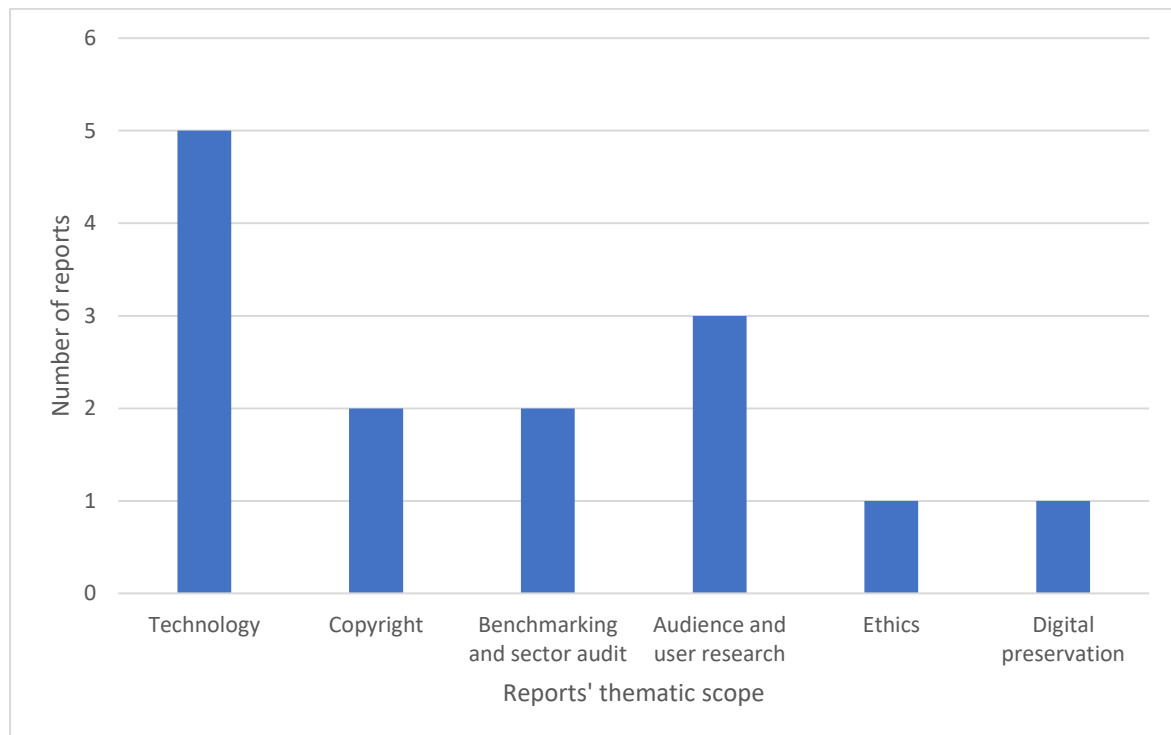


Fig.1: Thematic scope of TaNC’s Foundation Projects and commissioned reports autumn 2022 (figure by authors).[11]

However, examples like TaNC’s Provisional Semantics project suggest that research on “structures, practices, policies, collections and cultures” open crucial research arenas for enriching infrastructure development beyond the legal and technical sphere (Pringle et al., 2022: 10). This can extend to contesting institutionalised racism and injustices embedded in collections, argues the project. Indeed, even in the context of the seemingly more mundane and practical concerns of how to unify collections at scale, Science and Technology Studies established that sociocultural factors “[...] often are greater determinants of interoperability than is technology, per se [...]” (Borgman, 2015: 47). In this context Edwards et al. observe how infrastructures tend to be conceptualised as “laundry lists” which aim to facilitate interdisciplinary information management and work first and foremost through technological means (2007: 5). The planned MDS is for instance anticipated to work as follows:

“The technical solution offers a sustainable, low-cost mechanism to aggregate, manage, and index collection metadata from UK collections by 'harvesting' this 'raw' or 'native'

data and storing it in a cloud-based Data Repository. Initially, this will not include images. MDS will provide tools for public users and authorised partners to find, select and make use of the data in third-party systems and services, subject to rights and licensing restrictions. There are five key parts to the process: HARVEST, STORE & MANAGE, INDEX, DISCOVER, USE [capital letters in the original]" (Art UK et al., 2022).

Finding ‘technical solutions’ for infrastructure development is important. What gets lost with this focus are the challenges of crossing temporal, institutional, and disciplinary boundaries (Bowker et al., 2009: 100–01). In the academic and grey literature on collections unification, the sociocultural factors that can limit collections-holding organisations’ ability to exchange and aggregate data are present, but rarely the focal point of discussion. At best, non-technical challenges are mentioned incidentally, for example, with brief reference to different levels of collections description, disciplinary or local variations of metadata standards, working habits, staff knowledge, available resources for cataloguing, collection management systems’ support of standards or the formalising of partnership agreements (Darcovich et al., 2019: 9; Kostelic, 2017: 68–69; Maron and Feinberg, 2018: 688–90; Moulaison Sandy and Freeland, 2016: 48–49; Renshaw and Liew, 2021: 708–10). In other words, collections’ interoperability is shaped by the histories of collections’ arrangement and organisation (Sloan and Nyhan, 2021: 209–14). And yet the focus of the discourse on aggregation projects tends to be on which tools and licensing policies enable the unification of collections. This can obscure the labour, agency, conventions, and practices of the people who are responsible, behind the scenes, for collection interpretation, management and care (Sherratt, 2015: 7–8).

Yet the need to shift emphasis in infrastructure development is starting to be recognised. Interviews with TaNC’s Foundation Project Investigators brought up that: “Most of the current calls for digital projects seem to push for developing tools that are sometimes not linked to actual needs; there needs to be more reflection around these tools and the practices behind them, and tools [sic] development should be an integral part of research and curatorial practices” (Paltrinieri, 2023: 23). What, then, could a move towards a more holistic research agenda for infrastructure development look like? And what can be the gain from this shift of focus?

Star and Ruhleder argue that infrastructure development entails a complex interplay between the social and technological, the local and the global. Crucially, functional infrastructures emerge from the bottom up. Global technology and standards need to be able to accommodate local practices and conventions. The standing question is thus not ‘what’ an infrastructure is (e.g., an aggregator), but ‘when’ something becomes an infrastructure (1996: 112–14). Ideally, an infrastructure is characterised by being ‘embedded’ in social and technological arrangements, not built on top of them. By extension, a working infrastructure becomes ‘transparent’ and visible only when it breaks. Infrastructures also have some important caveats that need to be considered in their development and deployment. Conventions are hallmarks of infrastructures and familiarisation; they urge new members to become part of existing communities of practice. These existing conventions and practices are often inherited and may have strengths, such as allowing connections with other infrastructures. But inherited conventions may also be weaknesses that lock infrastructure participants in certain practices and at worst may force the adaptation of

practices favoured by more powerful participants (Star and Ruhleder, 1996: 113). Thus, for infrastructure development sociocultural questions need to be considered alongside technical ones. Cognizance of this is important for infrastructural actors and agents who wish to be more responsive towards the local and situated needs of an infrastructure's stakeholders, rather than setting boundaries in advance (Edwards et al., 2007: 7). Understanding infrastructures as more than technologies and standards allows us to push forward in developing nuanced understandings of the interests, limitations, and potentials that converge in them, and that must in turn support the interests of numerous diverse stakeholders.

While outside the scope of this paper such examinations call on us “[...] to reflect critically on what kind of social values and ways of thinking and working are embedded in planned infrastructures [...]” and crucially how “[...] to contribute to the reconfiguration of the global representation of digital knowledge” (Pawlicka-Deger, 2021: 540). Recent examples of such approaches are for example Tzouganatou’s analysis of APIs as accelerators and barriers for participatory memory practices (2021), or the study of the craft and precarity of maintenance work for collection databases (Thomer and Rayburn, 2023).

### 3. Methodology

The Sloane Lab determined semi-structured interviews as the most appropriate data collection method. This decision was based in our recognition of the limited availability of qualitative studies focussing on perspectives of collections-holding organisations and their role as data providers for aggregation infrastructures. Semi-structured interviews have the potential to contextualise and extend information that is otherwise either not available to the public at all, or dispersed across the grey literature (Hauswedell et al., 2020: 140). The format of semi-structured interviews provides a framework for a focussed conversation between interviewer and interviewee. But in contrast to more structured interview approaches, interviewees have in the semi-structured format the possibility to gain a more active role in the process of “knowledge production” by having the space to expand on issues they identify as relevant (Brinkmann, 2018: 579–80).

We used a “key knowledgeable” sampling strategy to identify interviewees. 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 digital collection aggregations and infrastructures. In total we identified and contacted 18 collections-holding organisations and aggregators based in the United Kingdom, the European Union and Australia. Of these 18 organisations 10 agreed to participate in an interview. Table I presents a pseudonymised overview of the interviewed organisations.



<b>Pseudonym</b>	<b>Abbreviation</b>	<b>Organisation description</b>	<b>Location</b>	<b>Number of interviewees</b>
Aggregation project 1	AP1	Collection research aggregation project	Australia and UK	2[12]
Aggregation project 2	AP2	Europeana aggregator	European Union	1
Collections-holding organisation 1	CHO1	Community Archive[13]	UK	1
Collections-holding organisation 2	CHO2	National Organisation (library)	UK	1
Collections-holding organisation 3	CHO3	National Organisation (library)	UK	2
Collections-holding organisation 4	CHO4	University Collection	UK	1
Collections-holding organisation 5	CHO5	National Organisation (historic environment)	UK	2
Collections-holding organisation 6	CHO6	National Organisation (plants and biodiversity conservation)	UK	3
Collections-holding organisation 7	CHO7	National Organisation (museum)	UK	2
Collections-holding organisation 8	CHO8	University Collection	UK	3

Table I: Overview on the interviewed organisations (table by authors)

The interview questionnaire[14] was derived from a review of the literature (partly formalised in section two) and through consultation with Sloane Lab team members with extended experience conducting qualitative data collection methods (Humbel, 2022; Nyhan and Flinn, 2016). The questionnaire covered the following thematic areas that are lesser explored in the literature:

- Rationale for participating in unified collections and infrastructures
- Experience with former collaborations
- Technical capabilities and data interoperability
- Legal aspects

- Anticipated use and benefits of unified collections and infrastructures

Our research is approved by the Research Ethics Committee of University College London (ethics ID: 22509/001). All interviews were facilitated by the paper’s first author and were held in June and July 2022 online via Microsoft Teams. Interviewees received the questionnaire, an information sheet and consent form prior to the interview. All interviewees gave consent to participate, to record and transcribe the interview. For the recording we used a H4n Pro recording device. The interview recordings were transcribed manually. Subsequently, the full transcriptions were returned to interviewees who were asked to inform us within four weeks about clarifications, amendments or whether they wished to withdraw from transcripts. This approach does not imply that interviewees agree with our findings or the arguments we present in this paper.

For our data analysis we applied what some call the ‘Miles, Huberman and Saldaña approach’ for qualitative data analysis (Punch, 2014: 173). The approach offers a robust framework for qualitative data analysis without being bound “to any one particular genre of qualitative research” (Miles et al., 2020: 6). This framework consists of three simultaneous streams of activities: “data condensation”, “data display” and “drawing and verifying conclusions” (Miles et al., 2020: 8–10). We used the ‘first-cycle’ coding methods of descriptive and In Vivo coding (verbatim codes derived from responses) to explore the content of each interview. The descriptive coding served for the initial familiarisation with the data, and In Vivo coding for gain deeper understanding of the interviewee’s perspectives. Following In Vivo coding’s ethos of honouring the interviewees’ voice we use direct quotes in section four relatively frequently (Saldaña, 2016: 102–10). In a second step we grouped the codes into thematic categories. This process of data condensation was accompanied by reflective reading of the subject-specific literature and narrating the findings deriving from the codes into analytic memos. Data display in mind-maps supported the process of data condensation. We verified emerging observations by checking how patterns repeat or deviate across the interviews (Miles et al., 2020: 79–91; 296–303). Table II exemplifies first cycle codes and thematic units.

Example thematic codes	Example In Vivo codes	Thematic units
Aggregation rationale Aggregation audience Aggregation potentials	“many forms of aggregation and sharing collections” "make our data as findable as possible"	Motivation for aggregated collections
Sustainability Funding structures Stalled projects	“[aggregation] very important nice-to-have” “reinventing the wheels”	Aggregation fatigue

Discrepancy in technical capability	"[aggregation] technical and ontological exercise"	Technical challenges
Legacy systems	"no automation"	
Aggregation prerequisites: vocabularies		
Copyright and collection type	"it's not just about copyright"	Legal and ethical responsibilities
Data protection		
Donor relationships		
Authority files	"the concept of aggregation is outdated"	Centralised vs. decentralised data sharing
Linking collections		
Data set dissemination		

Table II: Example first cycle codes and thematic units (table by authors)

## 4. Data

### 4.1 Motivation for aggregated collections

Most of our interviewees acknowledge the potential of aggregators to enhance collection discoverability, and to reach people "[...] who would never think of coming to this type of [herbaria] collection" (CHO4), for instance. Most interviewees participate in a wide range of aggregation infrastructures. These platforms are either domain specific (e.g.: the Global Biodiversity Information Facility - GBIF, Art UK, Early English Books Online) or dedicated to a specific geographic scope (e.g.: Scran, Archives Hub or Europeana). Some interviewees understand the concept of aggregation broadly and mentioned platforms developed outside the cultural heritage domain for sharing bespoke dataset types, such as Flickr Commons for images or Hugging Face for machine learning models. Our interviewees' core user groups are still expected to use the institutional catalogue. There was scepticism whether aggregations are of relevance for their core audiences beyond offering enhanced collections discoverability (CHO2; CHO3; CHO6; CHO8). This is particular the case for the community archives domain. CHO1 observes that reaching out to new audiences and promoting collections through aggregations are potential benefits for community archives. However, because community archives' main target group tends to be "their local community" it is, according to our interviewee, questionable as to whether aggregations are of relevance to community archives and whether they can withstand a cost benefit analysis. Some community archives may also not be aware that aggregations could be a way of collection dissemination (CHO1).

Some interviewees recognise the potential of aggregation for facilitating high-level analysis, creative use of historically dispersed collections or mapping phenomena like biodiversity. Such applications of collection data are, however, seen as an area of interest for research-oriented audiences with technical skill sets, such as the Digital Humanities domain. Meta analysis of historic collections is an area of interest mirrored in the ambitions of AP1 who describe the opportunities of aggregated collection data in “[...] better illuminate[ing] past relationships that have been obscured through institutional records”. For AP2, in contrast, the potential of collection data extends beyond the research domain, into education, tourism, information services like maps, and infotainment systems. AP1’s experiences with their project partners suggests that there could be an interest in aggregated collections for heritage organisations themselves, beyond serving just their audiences. Not discussed directly in interviews, this could be realised in terms of organisations leveraging data-driven insights on questions relating to collection policy, such as the extent of cultural diversity reflected within collections, for example.

Some of our interviewees illuminated how, for smaller collections-holding organisations, aggregations are the sole means of allowing their collection to be searched online (CHO1; CHO5). Indeed, infrastructures like the HOPE aggregator (Heritage of the People's Europe), hosted by the International Association for Labour History Institutions (IALHI), was explicitly built with the intention of supporting organisations with limited capacity for making collections available online (Siebinga et al., 2012: 5). The same applies for regional initiatives. Ulf Preuss describes, in the context of a regional aggregator for the German Digital Library, a “cooperative approach” where larger organisations support smaller partners in the provision of expertise and infrastructure, as well as in skill building. Large project partners also took over administrative tasks and funding acquisition (2016: 6; 10–11). Benefits of aggregation infrastructures for large collections-holding organisations can go beyond improving collection discoverability.

CHO2 explains how motivations for participating in aggregation infrastructures include facilitating ways of data access and display which go beyond the possibilities of their inhouse catalogue. Information enrichment through other collections or other third-party resources, such as authority files, via linked data is seen as a benefit too (CHO4; CHO6). Reports commissioned by TaNC and Collections Trust suggest that services provided by an aggregator, like data enrichment, long-term digital preservation, or copyright clearance could act as participation incentives for organisations (Gosling et al., 2022: 34–35; Collections Trust, 2019: 50; Wallace, 2022: 55–58; McNeill, 2022: 24–26). Some aggregators like PHOTOCONSORTIUM or the Digital Repository of Ireland (DRI) advertise such offers.[15] In response to the question of whether such additional services are part of aggregators’ responsibilities, interviewees expressed a need for aggregators to lower the threshold for data submission as much as possible. Everything else beyond facilitating data unification and dissemination would for most interviewees depend on the aggregator’s remit, however, and may be better served by other bodies in the heritage sector. One explanation for the reluctance that exists in seeing aggregation infrastructures as something more than data dissemination platforms relate to the sustainability of past initiatives.

## 4.2 Aggregation fatigue

As the range of aggregation initiatives listed in section 4.1 suggests, feeding into these infrastructures is for some interviewees a common practice (CHO10; CHO7). For most collections-holding organisations we spoke with, participating in aggregations is however seldomly part of a larger, sustained strategy and often takes place on an ad-hoc basis. It is rather a “nice-to-have” than a “must-have” (CHO5). Contributing to aggregators is not necessarily part of ongoing digitisation programmes. The competitive environment for gaining funding for digitisation and participating in infrastructure projects was identified as an important factor that led to the development of siloed collections (CHO2). Interviewees expressed that if making collection data available in digital infrastructures is something collections-holding organisations ought to be doing then there needs to be a more sector-wide strategic approach with appropriate funding structures (CHO3).

Within the UK context, the funding landscape for digital projects is indeed patchy and TaNC is a public investment unprecedented since the late noughties (Terras, 2011: 14–15). In the last few years some of the biggest investments in digitisation projects have come from commercial private partnerships, favouring collection types that would correspond more intentionally with market needs. Market logics can mean that commercial companies can have less incentive to create sustainable infrastructures and to offer content with non-restrictive licenses (Ahnert et al., 2023: 24–25; Hauswedell et al., 2020). Among interviewees there is a strong frustration in respect of public investments made in projects that were partially or fully stalled after funding ran out (CHO1; CHO3; CHO5; CHO7). The motivation for participating in something new that has no commitment to long-term sustainability is low: “[...] we don't want to invest in an aggregator that is going to be static in another year [...]” (CHO3). Some interviewees also note the potential negative effect of unsustainable infrastructures and their communities of interest: “if there’s finite funding, it can be quite damaging for the institution when that funding ends and the community feels abandoned” (CHO5). TaNC is designed to be entirely explorative, with a limited number of project funding calls and no commitment to ensure long-term sustainability. Some interviewees are sceptical as to whether TaNC can fulfil its ambitions with such an approach: “[...] one will come up against the buffers of internal sort of politics, the need for funding, and the competition between organisations in terms of creating an entire national collection” (CHO4). Given the abundance of existing initiatives for sharing collection data, our interviewees questioned the relevance of creating a new aggregation infrastructure. Interviewees expressed concerns that different projects seem to be “reinventing the wheels” (CHO3), and do not communicate well among each other: “There's presumably parallel spending of money on similar things, similar mistakes, and equally there's no sharing of experiences in most instances. [...] the money spent on these things and the efforts on these things aren't actually then becoming available. So, yes, it seems like a wasteful exercise [...]” (AP1). Interviewees note that there is generally an “[...] underestimation about the complexity of trying to bring national collections together” (CHO3). In the next two sections we unpack why unifying collections at scale is challenging from a technical and legal perspective.

### 4.3 Technical challenges

It is to be expected that for small organisations with few financial and personnel resources the ambitions for collections as data are, without dedicated support, out of scope. Our interviewee from the community archive domain observes a dissonance between aggregations and the practical reality of many community archives to "get their catalogues online in the first place" (CHO1). Throughout our interview consultations we observe however that collections-holding organisations' asymmetric capabilities for publishing collection data does not only play out between major and small organisations, but also between different collection types within the organisations themselves.

Our interviewees from CHO5, an organisation for historic environment conservation, report how data from spatial data information systems can be exported straight forwardly through programming routines. Within the same organisation the data export for the archival collection would in contrast require a significant amount of manual intervention, as the data is kept within a legacy SQL (Structured Query Language) database system with limited processing capacity. With 1.5 million collection records and over 1 million images stored in the database any data export at scale requires support from the organisation's IT department (CHO5). In the case of CHO6 the state of digitisation varies across the library and archive collection vs. the herbarium and living plant collection. The herbaria's digitisation programme runs since 2004, resulting in approximately three million plant specimen records and 600,000 images. For the living plant collection, metadata and images are complete and updated on an ongoing basis. Herbarium and living plant collection data are also regularly pushed to GBIF. In contrast, the focus of the library and archive was to date not on digitisation, but on metadata creation. While the library made collection information available in union catalogues, the capacity for participating in aggregations was facilitated only through the provision of external IT support, as the in-house IT-service did not have capacity for this kind of work (CHO6). These are not exceptional experiences. At the time of the interview some of our other interviewees just recently updated and harmonised their Digital Asset Management Systems (DAMS) across the organisation (CHO2; CHO8). Our interviewee from CHO2 made clear that such work needs careful planning and always needs to be weighed up against the costs for collection conservation and management. Legacy- or "homegrown" systems are also common within the community archive domain and limit their potential of integration in infrastructures (CHO1).

Given the huge discrepancies in technical capacity across and *within* collections-holding organisations, aggregators need to be able to accommodate a wide range of different data export formats and standards. Metadata mapping is for aggregators a core activity, and data import often needs to be customised towards the capabilities of individual data providers (Brugman et al., 2016: 1278; Butigan et al., 2020: 61). Both aggregation projects we interviewed reported how essential it is to keep the technical threshold for content providers as low as possible by taking the lead on tasks such as data massaging, cleaning, and mapping to facilitate the participation of organisations of different sizes and types. But reducing the challenges of collection data unification at scale on technological and capacity considerations alone obscures some of the more subtle, yet potent challenges. As one of

our interviewees from the research aggregation project puts it, collection unification at scale “[...] becomes a story about the inevitable differences between the conventions of recording digital information about archives, libraries, objects and art.” Crucially, this is a “technical *and* ontological exercise” (AP1). Our interviewee from CHO4, an organisation based in the field of natural sciences, observes how pushing data to GBIF is for the organisation a straightforward process as the data and records are relatively homogenous and confined to domain-specific taxonomical descriptors. At least for specialist audiences, the common search terms and query types can be anticipated. Difficulties emerged when attempting to bring herbaria collections together with other domains such as archaeology or art (CHO4). Similarly, CHO7 report how in their experience standards for data sharing are much more established within the natural sciences than in the humanities. Finding common denominators for cultural collections presents a challenge as the humanities disciplines are traditionally more focussed on interpretative work (Borgman, 2007: 217–19).[16]

#### *4.4 Legal and ethical responsibilities*

The challenges collections-holding organisations face when identifying collection copyright status, adopting online licensing schemes and signposting public domain works as Open Access are known to be grounded in the complexity of copyright legislation, contractual agreements, requirements to create revenue, or because sometimes copyright holders cannot be identified or contacted anymore (orphan works) (Martinez and Terras, 2019; Peters and Kalshoven, 2016; Wallace, 2022: 81–83). In practice, collections-holding organisations navigate these challenges by assessing the risk of copyright infringement, the consequences of releasing sensitive and personal information, or damaging donor relationships (Wallace, 2020b: 8–9; Deazley, 2017: 4–5; Stobo, 2016: 281–85). All these themes were raised by our interviewees. Our analysis suggests that in addition to these issues, collections-holding organisations’ disciplinary practices, cultures and collection affordances play out in these risk assessments, and ultimately privilege certain collection types being made available online.

Art collections belong to the most celebrated collection types for Open Access frameworks, and the practices of major institutions like the Rijksmuseum, the Danish Statens Museum for Kunst, or the Metropolitan Museum of Art significantly raised expectations in respect of open licensing (Halperin, 2017; McCarthy and Wallace, 2020; Pekel, 2013; Sanderhoff, 2014). One of our interviewees observes how major art collections lend themselves to Open Access frameworks because their creators tend to be identifiable and well-known (CHO8).[17] For certain object types the risk of copyright infringement is low. Interviewees exemplified this in the context of herbaria collection where plant specimens qualify as non-human made works (CHO4; CHO6). However, other collection types present significant copyright assessment challenges. Copyright assessment is sometimes impossible due to the mass of records an archive needs to process (CHO5). The impact of collection types on legal assessments can be illustrated further. While for CHO6’s herbaria collection the legal assessment is relatively straight forward, the case for their archival collection is not, and more uncertainty characterises their online access policy for this part of the collection. In the cases of CHO5 and CHO6, interviewees report how in the past collections were

commonly deposited without specifications relating to intellectual property or access permissions. But even in the case where depositor agreements exist there is a dissonance between legacy agreements and the changed expectations of access, as donor agreements often predate the internet and unification of collections at scale:

“I think, from a legal point of view, we would have to be very cautious because we didn’t state we were going to do that aggregation work, so therefore we have to return to the owners and work through that and get a new signature on a piece of paper” (CHO5).

Depending on the organisation and collection type, legal considerations blur with responsibilities which go beyond the sole assessment of copyright or collection ownership. CHO4 reports on withholding geographic data about certain herbaria collections that fall under the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), as this information could be used to deplete species. Archival collections typically consist of photographs, diaries, letters, or administrative records which are prone to contain information that fall under data protection legislation (CHO6). Even if data protection does not apply anymore, interviewees report that they still need to consider that some collection information could be associated with living descendants and family members (CHO6; CHO7; CHO8).

Interviewees emphasise that the legal assessment of collection access protocols are insufficient where intra-cultural injustice needs to be addressed. Digitisation and online dissemination can entrench the violent appropriation of heritage from people who were dispossessed through colonialism, imperialism, the enslavement of human beings and other iniquitous acts and processes (Pavis and Wallace, 2019: 6–7; Wallace, 2021: 15; Ortolja-Baird and Nyhan, 2022). Crucially, cultural injustices are not only about images but also about descriptive information, potentially including metadata, as our interviewee from AP1 points out. They estimate that at least a decade will be necessary for building meaningful relationships with affected groups to address cultural injustice within their collection data aggregation project. Important steps forward are in this context the CARE (Collective Benefit, Authority to Control, Responsibility, and Ethics) principles, which seek to “[...] reposition[s] Indigenous Peoples, nations, and communities from being subjects of data that perpetuate unequal power distributions to self-determining users of data for development and wellbeing” (Carroll et al., 2020: 2).

#### *4.5 Centralised vs. decentralised data sharing*

Throughout our conversations we observe scepticism towards the concept of unifying siloed collections in a centralised manner because “[...] one size isn’t going to fit all for the legal, technical, organisational [...] conditions” of participating organisations (CHO5). For many of our interviewees the issue of “dissolve[ing] barriers between different collections” (TaNC, 2023) is not resolved by creating a digital national collection. To tackle the technical and legal idiosyncrasies described in section 4.3 and 4.4, they call instead for investment in the development and enhancement of resources which allow for cross-disciplinary information management, such as vocabularies and authority files (CHO1; CHO4; CHO6; CHO7; CHO9).



Closely related is the observation by some of our interviewees that the concept of aggregation no longer represents the state-of-the-art, which presently tends towards the linking of collections without an intermediary platform (AP2; CHO4; CHO5; CHO6). One interviewee from CHO3 argued, for instance: “It feels to me like it's counterintuitive to still be bringing huge amounts of digital stuff together in one place just so people can search it [...] it should be like: you're in one place and it points you off into all these other related resources”. Linked Data, “[...] a set of best practices for the publication of structured data on the web” (Van Hooland and Verborgh, 2014: 3), presents an alternative to the data aggregation model. The TaNC Foundation Project ‘Heritage Connector’ leveraged for instance the LOD database Wikidata for data enrichment and linking between siloed institutional collections catalogues (Dutia and Stack, 2021: 2–3).[18] However, LOD is not a common approach within the sector. The few interviewees who mention experience with LOD projects describe them to be of an experimental character. Indeed, scholars found that LOD applications within the cultural heritage sector tend to remain in “the prototype stage” (Winters et al., 2022: 10) and are mostly facilitated by big institutions in explorative manners (Davis and Heravi, 2021: 12).

Rather than disseminating collections via an aggregator, some interviewees experiment with publishing their collection as datasets for download (CHO2; CHO3). The advantage of this approach is that people can use the material in different ways, including integrating collections in existing aggregations and platforms such as Wikipedia or the Internet Archive, without the organisation actively needing to invest resources in “managing those relationships” with third party aggregators (CHO2). CHO8 see similar potential in offering access to collection data via an API. But the experiences they made suggest that the technical tool of an API does not resolve the fundamental issues that impede collections-holding organisations in sharing their collection data. Policies and workflows need to be in place to ensure data is shared in accordance with copyright and data protection legislation. Their servers would not be capable of handling a mass of queries at the moment (CHO8). The statements by CHO8 echo the findings by Winters et al. who point out that the current design of the APIs used by collections-holding organisations does not support large-scale use (2022: 14–16). Our interviewees were also clear that dissemination of data sets alone serves only a very small cohort of potential users with advanced technical skills or the appropriate equipment to deal with large data sets: “having terabytes of data [...] isn't the most accessible form for people who don't have a computer that can open terabytes worth of files” (CHO2). Similarly, CHO4 observes how for the general public the dissemination of data alone is of little use and for them “a lot of investment into the interpretation of the objects” is necessary. In terms of the limited technical capacities and skills of smaller organisations, purely data driven approaches for sharing collections favour large institutions with appropriate resources, bolstering existing privilege in data-contexts and further marginalising less-well-resourced players.

## 5. Discussion

In this section we respond to the research questions, by discussing the findings from the interviews with reference to the wider literature. In this article we asked:

1. What are collections-holding organisations' experiences of participating in digital infrastructure projects?
2. Which factors enable and impede collections-holding organisations in unifying siloed collections?
3. Which risks and opportunities do collections-holding organisations perceive in large-scale infrastructure investments, such as the UK's Towards a National Collection (TaNC) programme?

Regarding collections-holding organisations' experiences of participating in digital infrastructure projects, a complex picture emerged from the interviews. Interviewees acknowledge broadly positive experiences with digital infrastructure projects. Interviews contain discussion on the merits of these infrastructures and the new opportunities they can open in respect of data enrichment, collection exploration, display and making collections available to wider audiences. Existing infrastructures are also important platforms for some smaller organisations to make their collections available online in a first place. The potential of analysing collections at scale through their unification is acknowledged by interviewees, though mostly seen as an interest confined to a specialised audience, a position echoed in the external literature (Hauswedell et al., 2020: 155–56). The potential benefits of unification are understood by interviewees to pertain to audiences. Research shows however that usage profiles of digital infrastructures' audiences tend to be not well understood and articulated (Bettivia and Stainforth, 2023: 709; Bailey-Ross, 2021: 25). While there could be a potential for institutions themselves to explore their own histories and collections through virtual unification, this avenue does not seem to be a core interest for the organisations we interviewed. Overall, the value and opportunities of aggregated collections need to become clearer (Poole, 2015). Those who seek to fund the provision aggregators have been known to communicate their rationale in language that emphasises novelty, and accordingly scarcity, and is furthermore peppered with references to the technological state of the art. TANC is a case in point:

“By seizing the opportunity presented by new digital technology, it 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).

Notwithstanding what is characterised above as a broadly positive attitude to aggregators, interviewees' responses often seem to question whether collections-holding organisations experience such infrastructures as novel, scarce and technologically cutting-edge. In fact, to date, collections-holding organisations have been able to choose from an abundance of aggregators or other solutions to share collection data outside of their own catalogues. In the UK context we find many examples which aspired to unify collections at scale. Since 1995, Scran has been bringing together digitised audio-visual material “[...] from over 300

[UK] museums, galleries, and archives including the V&A [The Victoria & Albert Museum], National Galleries of Scotland, Glasgow Museums and The Scotsman to name a few” (Historic Environment Scotland, 2023a; Historic Environment Scotland, 2023b). Another example is the UK Archives Hub – a discovery service that aggregates the collections of over 350 archival organisations – and which emerged in the early noughties out of “[t]he vision of a National Archives Gateway” (Hill, 2002: 239). With Culture Grid the UK had until relatively recently another infrastructure in place for “[...] joining and opening up UK collections for more use by more people in more ways” (Culture Grid, 2010). While formally still operating as “the *de facto* national aggregator for UK museums [italics in the original]” and pushing content to Europeana, Culture Grid closed down for new data providers in 2015 (Gosling, 2019). TaNC’s aspiration of unifying collections on a national level is not new. Strong frustration across collections-holding organisations surfaced in interviews in respect to infrastructure projects with similar objectives and uncertain longevity.[19]

Interviewees also questioned the presentation of aggregators as technologically state of the art. Rather than established aggregation models, interviewees expressed more interest in decentralised approaches to collection unification. What such an approach could practically look like remains unclear. LOD approaches have not had wide up take in the sector despite having had some presence there for over a decade. The distribution of data sets alone is likely to serve a specialised technical audience only. Lightweight interfaces and toolboxes for collection data exploration could offer ways forward (Winters et al., 2022: 16)[20] but would require a certain level of technical expertise from the users.

The European Commission’s ambition in respect of the ‘common European data space for cultural heritage’ point towards decentralised approaches too. The creation of a ‘data space for cultural heritage’ was proposed by the European Commission in 2021 (2021: 5) and is situated within the European data strategy to develop “a genuine single market for data, open to data from across the world” that is subject to EU legislation, norms and governance (European Commission, 2020: 4–5). Accordingly, the EU funded ‘Data Spaces Support Centre’ defines the data space as: “[a]n infrastructure that enables data transactions between different data ecosystem parties based on the governance framework of that data space. Data space [sic] should be generic enough to support the implementation of multiple use cases” (Poikola et al., 2023: 5). Given these elusive definitions commentators point out that there is a lack of consensus as to what the concept of the ‘data space for cultural heritage’ is ought to be in practice (Dobrevá et al., 2022: 493). Ambiguous also is the role of Europeana (an aggregator of aggregators, after all) which is expected to become in a rather “recursive” manner “[...] the core of the emerging Common European Data Space for cultural heritage” (Keller, 2021). Despite advances in decentralised technologies such as Linked Data there is to date no robust framework for a coordinated approach for facilitating interlinking of collection data in a decentralised manner.

Interviewees noted serious risks for small collections-holding organisations in infrastructure programmes. Current funding models, and indeed strategy and foresight policies surrounding aggregation infrastructures are not only deemed inadequate but, from the perspective of smaller organisations (who are the most commonly occurring heritage entity

within the wider sector),[21] structurally flawed. Participation in aggregation infrastructures often takes place in an ad-hoc manner, not as part of long-term strategies. Our interviewees identify the absence of a sector-wide strategy and competitive funding environment as key drivers of the fragmented digital collection' landscape. Infrastructure funding within the heritage sector is commonly organised through competitive funding calls.[22] It is known that the nature of major funding schemes presents significant hurdles for smaller collections-holding organisations, such as community archives. Funding applications are labour-intensive, require a certain skill set and domain expertise as well as good connections within professional networks. Funding schemes can come with the prerequisite to collaborate in partnerships that may compromise the local practices of smaller heritage bodies (Jules, 2019: 7–8). By definition the nature of major funding schemes reinforces the dominance of a handful of well-resourced major institutions in the digital canon as a whole (Kizhner et al., 2020: 12–14; Gosling et al., 2022: 6; Wallace, 2020c: 5).

Regarding factors that enable and impede collections-holding organisations to unifying what are otherwise siloed collections, the tension between innovation and sustainability arises. The limited capabilities of even relatively well funded collections-holding organisations to disseminate collection data sits much at odds to the scope of current innovation driven infrastructure investments. Interviewees observed that there is commonly an underestimation of the complexity entailed in building large-scale infrastructure projects, and that the capabilities of many organisations to deliver data is often limited. Responses from our interview consultations suggest that legacy systems with limited capacities for data management and export are a common phenomenon across the sector. While it is unsurprising to find that small collections-holding organisations can struggle to make collection data available, we observe throughout our interview consultations that the asymmetries in technical capacities play out not only between major and small organisations, but also within organisations themselves. It is worth noting that the issue of legacy technology is not only something that impedes collections-holding organisations in making data available, but aggregators and other infrastructure providers too. OAI-PMH - first published in 1999 - is widely considered to be an outdated technology, proposed to be replaced by alternatives such as IIIF (Van de Sompel and Nelson, 2015; Freire et al., 2020). However, OAI-PMH is still widely used in the heritage sector, including by Europeana, the Digital Public Library of America, or sub-repositories of the Digital Research Infrastructure for the Arts and Humanities (DARIAH-EU), such as ROSSIO (Silva et al., 2022: 3). The technology is difficult to replace because it is considered to be the “de-facto standard” now (Butigan et al., 2020: 63–64). Technological hypes often suggest a radical change by adopting a particular technology and obscure how moving from one technology to another is rarely clear cut. Instead, the ‘old,’ ‘new’ and sometimes even alternatives, co-exist. In addition, the choice of one technology over another is in many cases less rational as commonly assumed. Instead, economic-, cultural-, and social factors in addition to pure chance, pragmatism, or ideology have far greater impact on which technologies become established (Edgerton, 2006: 1–3; 8–11).[23]

The dynamics shaping technological choice and its ongoing impact forces us to think about the interplay between innovation and sustainability. Contrary to the rhetoric on innovation

surrounding infrastructure investments, there is a strong risk that infrastructures become static once they are set up, rather than spaces where research and development happens (thus the upkeep of technologies like OAI-PMH). Unsurprisingly a common lack of funding for sustaining infrastructures reinforces this tendency (Rockwell, 2010: 620–21). Sustained infrastructure investment is not a replacement for research and innovation and vice versa. Equally collections-holding organisations need more support for addressing limited technical capacities (e.g. updating legacy databases and skill development) to be able to deliver collection data in a first place.

Divergent disciplinary traditions across the so-called ‘G-L-A-M’ sector tend to be obscured when speaking of infrastructures which aim to cater for all types of collections-holding organisations. Larger organisations, for example, commonly have specialised departments with bespoke collection management systems for different domains (e.g., for humanities and natural science collections) or different collection types: library-, archive- or museum collections. The unification of collections at scale is not simply a matter of finding the appropriate technical tool. It is shaped and hampered by the inherited norms, practices and cultures of information management which developed with the formation of Western scientific disciplines in the 17<sup>th</sup> and 18<sup>th</sup> century onwards (Edwards et al., 2007: 5–7). Museum collection description emerged out of the need to register and manage objects internally, not for external information retrieval. Information management via cataloguing standards and vocabularies gained within the museum domain for this reason a less prominent role than in libraries for instance. Consequently the linking of museum collections internally or externally with other collections is hampered (Zoller and DeMarsh, 2013: 58–64). Another distinction is that libraries and museums usually describe collections on an item-level, whereas archives describe on a collection- or folder level. This is because archives’ focus is to contextualise the relationships between administrative or individual’s activities of record creation (Timms, 2009: 74). The insufficiency of bringing collections together on an object-level only, as many aggregation projects do, is recognised. It can, to some extent, be mitigated by allowing for data ingest on different levels: such as institution, collection, catalogue record and digital surrogate (Collections Trust, 2019: 22). However, querying such aggregates will inevitably be skewed towards organisation types which record information on an item-level as their content will surpass others in quantity and detail. Accordingly aggregates of digital collections do not present ontological surrogates of their physical collections’ counterparts. Aggregators are data assemblies, where the information management practices from library-, archive-, museum-studies and other disciplines play out in ways that are different to the in-house digital and physical catalogues.

Differences in collection types and their description also play out in respect to collections-holding organisations’ ability to comply with copyright licensing standards. Copyright clearance requires the identification of the creator of a piece of work. But depending on the cataloguing practice, for example, the taxonomical classification of biological specimens, the identification of authorship may not be of central concern. “Name of creator” in the archival standard ISAD(G) [24] refers for example to “[...] the name of the organization(s) or the individual(s) responsible for the creation, accumulation and maintenance of the records in the unit of description”, which can but not necessarily has to be the collection’s copyright holder (ICA, 2000: 10;18). UK Archives Hub no longer require ‘creator’ as a mandatory metadata field for data submission because for many archival organisations it is not possible

to identify records or collection's creators (Stevenson, 2019: 98). A case study presented by the German National Library demonstrates how for copyright assessment of books each work needs to be physically checked because illustrations or photographs may have different copyright holders than the book's author. Overall the report concludes that the costs of rights clearance for one work (between €1.39 and €27) "[...] often exceeds the economic value of the work themselves" (Peters and Kalshoven, 2016: 1–5). If the economic value of Open Access collections is to be reaped, as promoted by some policy makers, considerations need to be put in place to lower the threshold for copyright clearance through legislative frameworks and funding support.

Increasingly Open Access to cultural heritage as a blanket approach for stimulating access and use of collection data turns out to be insufficient. Scholars point out how the violent appropriation of collections from their communities of origin render those collections as inappropriate for digitisation and Open Access frameworks (Pavis and Wallace, 2019: 6–7). Other collection types prone to replicate trauma and harm through collection data dissemination include among others medical records, LGBTQ+ and Feminist publications (Wernimont, 2021; Cowan and Rault, 2018: 124–25). The CARE principles are for affirming Indigenous data governance and sovereignty. While Carroll et al. see the potential of the CARE principles' wider application the authors also call for caution for the principles' co-option in non-Indigenous environments, until the principles matured fully (Carroll et al., 2021: 5). Important contributions on addressing data inequities are also made in the field of Critical Archival Scholarship. The Feminist Ethics of Care, for instance, call collecting-organisations to take up "affective responsibilities" towards the lived-experiences who are the most vulnerable collections stakeholders (depending on the context: creators, subjects, users, and the organisation's wider community) (Caswell and Cifor, 2019). Accordingly the Ethics of Care consolidate as a framework for addressing digital collections' uncertain ethical afterlives (Ziegler, 2020; Agostinho, 2020: 69–70). Recent research building on Social Movement Archives' trajectory of informing critical archival theory and praxis, argues to move towards social-justice based frameworks for collections as data (Humbel, 2023: 181–86).

Finally, notwithstanding the potential of aggregators, interviewees still expect their core audience to use the institutional catalogue for collection access, and call for investments for improving the quality of catalogues, such as authority files. Collection curation and cataloguing are prime examples of infrastructural – and thus invisible and typically feminized – work (Shirazi, 2018; Caswell, 2016: 10–13; Star and Strauss, 1999). The common touting of digitisation and digital infrastructures as the solutions for unlocking the past misses the point that significant amounts of collections are not even digitally searchable (Zaagsma, 2023: 845). Catalogue data consolidate as distinct sources for data driven research (Baker et al., 2022; Bagnall and Sherratt, 2021; Havens et al., 2023), and the importance of collections curation and cataloguing is meanwhile even recognised outside the immediate heritage and Digital Humanities domain. So highlighted in summer 2023 The Guardian's chief culture writer Charlotte Higgins the fundamental role of professional cataloguing work for institutional transparency and accountability in context of the stolen, missing and damaged objects from the British Museum (2023).[25] Cataloguing cannot be thought as an addition to existing curators' job descriptions, but require specialised roles with distinct skill sets for

information management, standardisation, and access provision (Zoller and DeMarsh, 2013: 64). Ongoing investments in collections curation and high-quality cataloguing are a prerequisite for a sustainable heritage sector and collection data infrastructure development.

## 6. Conclusion

This article highlights institutional perspectives and experiences in relation to large-scale, third-party funded infrastructural investments which aspire to aggregate collections at scale. Through an analysis of a series of interview consultations with experts in the aggregation scene, and a critical analysis of the dispersed academic and grey literature, we contribute to the emerging understanding of the latent factors that make infrastructural endeavours in the heritage sector complex undertakings. To date the research focus has tended to be on the drive to adopt specific technological solutions and copyright licensing practices. In addition, projects like TaNC and MDS seek to build new platforms that unify collections at scale. In this article we argue that the question of how to unify collections at scale cannot be reduced to finding the appropriate technology or licensing framework, nor by building new aggregators. Instead, collections-holding organisations' capacity to participate in digital infrastructures is dependent on a complex interplay of resource allocation across the sector and within organisations, divergent traditions and requirements of collection description, and different disciplinary histories and aims. By conceptualising infrastructures as a "machine to be built" there is a risk to "[...] downplay[ing] the importance of social, institutional, organizational, legal, cultural, and other non-technical problems developers always face" (Edwards et al., 2007: 7). In this article we went beyond the developers' views by bringing together the perspectives of curators, archivists, librarians, and IT specialists. Through a critical analysis of their responses and of the scholarly and grey literature, including the stated rationales of aggregators and third-party funders, we exposed holistic levels of potential and yet unresolved difficulty in collection data infrastructure development not apparent when the issue is seen from the perspective of one actor only.

Accordingly, we call for better social-cultural and trans-sectoral (collections-holding organisations, universities, and technological providers) understandings of collection data infrastructure development. Attention to collections-holding organisations' local affordances and the differences in disciplinary conventions is likewise salient. For funders, policy makers, researchers, and practitioners of infrastructure projects we recommend:

- Sustainable financial investment across the heritage sector is required to address the discrepancies between different organisation types in their capacity to deliver collection data. Smaller organisations play a vital role to diversify the (digital) historical canon, but they often struggle to digitise collections and bring catalogues online in a first place. Ongoing investments in collections curation and high-quality cataloguing are a prerequisite for a sustainable heritage sector and collection data infrastructures.
- More studies on the impact and opportunities of unified collections for different audiences, and collections-holding organisations themselves.

- Investments in existing infrastructures for collection data dissemination and unification, instead of creating new platforms. This may include bringing existing systems together.
- Investments in the sustainability of infrastructures is not a replacement for research and vice versa.
- Establishing networks where collections-holding organisations, technology providers and users can communicate their experiences and needs in an ongoing way and influence policy.

## Acknowledgements

This research was part of ‘The Sloane Lab: Looking back to build future shared collections,’ a collaborative project led by UCL and the Technische Universität Darmstadt with the Natural History Museum (London) and the British Museum. This project was supported by the UK Arts & Humanities Research Council funded *Towards a National Collection* programme. AHRC project [AH/W003457/1]. We wish to thank all interviewees for their invaluable contributions. We also wish to thank the anonymous peer reviewers for their helpful feedback.

## Notes

---

1. GLAMs (Galleries, Libraries, Archives and Museums) is a common acronym used to refer to cultural organisations which curate and provide access to collections. Not all organisations we consulted are covered by this acronym, however. Some with herbaria collections might be better designated as scientific research institutions for instance (see table 1). We decided to use ‘collections-holding organisations’ as a short-hand to refer to the organisations covered by our research.

2. Open Access to cultural heritage is defined as “[...] a policy or practice that allows reuse and redistribution of materials for any purpose, including commercial” (Wallace, 2020a: 4)

3. For the OpenGLAM principles see: OpenGLAM Initiative (2024). OpenGLAM Principles [WWW Document]. URL <https://openglam.org/principles/> (accessed 18 June 2024).

4. See: The National Lottery Heritage Fund (2020). Advice: Understanding our licence requirement [WWW Document]. URL <https://www.heritagefund.org.uk/stories/advice-understanding-our-licence-requirement> (accessed 27 June 2023).

5. Other planned European data spaces include for instance the domains of health, industrial manufacturing, finance and agriculture (European Commission, 2020: 22–23).

6. The claim highlights the weight the UK – as a single country - puts on digital heritage infrastructure investments. Compare for instance the ICT PSP digital libraries programme, in which the European Commission made €30 million funding available (European Commission, 2010: 13).



---

7. “OAI-PMH [...] is a client/server architecture protocol specification that facilitates the diffusion of metadata e.g.: Resource description of the resource (*title, author, date of publishing, publisher, etc.*) [and] Resource location on the Internet (*indicated by the URL*) [italic in the original]” (IFLA, 2023).

8. A close relative of the ‘aggregation approach’ are the federative union catalogues developed in the library domain (Collections Trust, 2019: 8). The focus in this paper is however on the aggregation model because these systems are typically used to bring collections together across archives, libraries, museums and heritage organisations.

9. For a list of Europeana aggregators see: [WWW Document]. URL <https://pro.europeana.eu/page/aggregators> (accessed 14.12.23)

10. As the bar chart indicates, TaNC also commissioned research on engaging audiences or the affordances of born-digital material. The point is that research specifically addressing the question on collection unification at scale typically takes a focus on technical outputs and copyright.

11. Categorisation based on thematic analysis of reports’ content. All TaNC reports are available via Zenodo: [WWW Document]. URL <https://zenodo.org/communities/tanc?q=&l=list&p=1&s=10&sort=bestmatch> (accessed: 17.10.23)

12. The interview had to be facilitated in two separate sessions due to the unforeseeable unavailability of one of the interviewees.

13. We use the term ‘Community Archive’ “[...] to characterise a non-traditional archival collection specifically tied to a particular group, often one that may be undocumented or under-documented by traditional archival institutions” (Bastian and Flinn, 2018: XX).

14. The questionnaires were developed by Marco Humbel, Nina Pearlman, JD Hill, Andreas Vlachidis, Daniele Metilli and Julianne Nyhan and are available in the appendix.

15. See: [WWW Document]. URL <https://www.photoconsortium.net/services/> and <https://dri.ie/dri-membership/> (accessed 8.6.23).

16. For an exemplary case study see for instance the challenges of identifying the boundaries of object descriptions in historical museum catalogues (Ortolja-Baird et al., 2019: 21–28).

17. On the impact of copyright on smaller art collections see the ongoing work by UCL Art Museum: [WWW Document]. URL <https://www.ucl.ac.uk/culture/projects/ucl-rightsholder-clearance-project> (accessed 2.12.23).

18. Fundamental principles of Linked Data include Uniform Resource Identifiers (URI) to anchor machine readable data on the web, and the Resource Description Framework (RDF) to describe resources, and the relationships among them, using triple statements (subject – predicate – object). If Linked Data is published under an open license, it becomes Linked *Open* Data (LOD) (Berners-Lee, 2010).

19. For examples of stalled UK infrastructure see: Gosling, K., 2019. Nothing new except what has been forgotten [WWW Document]. URL <https://collectionstrust.org.uk/blog/nothing-new-except-what-has-been-forgotten/> (accessed 3.6.23). In addition, also consider the fate of the JISC-AHRC

---

funded national research data infrastructure 'Arts and Humanities Data Service' (AHDS). Established in 1995 the service ceased to receive funding in 2008. Ultimately the service was decommissioned in 2017 (Greenstein, 1996; AHDS History, 2007; King's Digital Lab, 2024).

20. For an example for this approach see: Tim Sherratt. (2021). GLAM Workbench (version v1.0.0). Zenodo. [WWW Document]. URL <https://doi.org/10.5281/zenodo.5603060> (accessed 2.12.23).

21. According to Candlin et al. in 2017 large and huge museums formed for instance together 15,36% of all UK museums, whereas small museums 56% (2020: 26). For a discussion on the underrepresentation of the majority of small archives, libraries and museums in the UK's digital canon see (Gosling et al., 2022: 33).

22. See for instance DIGITAL-2022-CULTURAL-02 Data for cultural heritage (deployment): [WWW Document]. URL [https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/digital/wp-call/2022/call-fiche\\_digital-2022-cultural-02\\_en.pdf](https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/digital/wp-call/2022/call-fiche_digital-2022-cultural-02_en.pdf) (accessed: 21.07.23).

23. See for instance the famous VHS vs. Betamax case (Ponte and Camussone, 2013).

24. The International Standard Archival Description (General) "[...] provides guidelines for creating descriptions of archival materials, establishing a model based on the principle of respect des fonds within a multi-level description (SAA, 2023).

25. See the British Museum's press release on the case: [WWW Document]. URL [https://www.britishmuseum.org/sites/default/files/2023-08/Announcement\\_regarding\\_missing\\_stolen\\_and\\_damaged\\_items.pdf](https://www.britishmuseum.org/sites/default/files/2023-08/Announcement_regarding_missing_stolen_and_damaged_items.pdf) (accessed: 17.11.23).

## References

- Agostinho, D.** (2020). Care. In Thylstrup, N. B., Agostinho, D., Ring, A., D'Ignazio, C. and Veel, K. (eds), *Uncertain Archives: Critical Keywords for Big Data*. Cambridge, Massachusetts: The MIT Press, pp. 67–75.
- AHDS History** (2007). AHDS History - Depositing Data with the AHDS <https://web.archive.org/web/20130413012920/http://www.ahds.ac.uk/history/depositing/index.html> (accessed 7 July 2024).
- 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/element> (accessed 4 May 2023).
- Angelova, L.** (2021). *Deep Discoveries: A Towards a National Collection Foundation Project Final Report*. (Interim Report Foundation Projects Towards A National Collection). London: TaNC doi:10.23636/1214. <https://www.nationalcollection.org.uk/sites/default/files/2021-10/Deep%20Discoveries%20Final%20Report%20.pdf> (accessed 1 December 2021).
- 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).
- Art UK, Collections Trust and the University of Leicester** (2022). Museum Data Service <https://artuk.org/about/museum-data-service>.
- Bagnall, K. and Sherratt, T.** (2021). Missing Links: Data Stories from the Archive of British Settler Colonial Citizenship. *Journal of World History*, **32**(2): 281–300 doi:10.1353/jwh.2021.0025.
- Bailey-Ross, C.** (2021). *Online User Research Literature Review: UK Gallery, Library, Archive and Museum (GLAM) Digital Collection*. Zenodo doi:10.5281/ZENODO.5779826. <https://zenodo.org/record/5779826> (accessed 9 June 2023).
- Baker, J., Salway, A. and Roman, C.** (2022). Detecting and Characterising Transmission from Legacy Collection Catalogues. *Digital Humanities Quarterly*, **16**(2) <http://www.digitalhumanities.org/dhq/vol/16/2/000615/000615.html> (accessed 15 November 2023).
- Bastian, J. A. and Flinn, A.** (2018). Introduction. In Bastian, J. A. and Flinn, A. (eds), *Community Archives, Community Spaces: Heritage, Memory and Identity*. London: Facet, pp. xix–xxiv doi:10.29085/9781783303526. <https://www.cambridge.org/core/product/identifier/9781783303526/type/book> (accessed 29 April 2021).
- Beelen, K., Lawrence, J., Wilson, D. C. S. and Beavan, D.** (2023). Bias and representativeness in digitized newspaper collections: Introducing the environmental scan. *Digital Scholarship in the Humanities*, **38**(1): 1–22 doi:10.1093/llc/fqac037.
- Berners-Lee, T.** (2010). Is your Linked Open Data 5 Star? *Linked Data* <https://www.w3.org/DesignIssues/LinkedData.html> (accessed 7 November 2023).
- Bettvia, R. S. and Stainforth, E.** (2023). Negotiating digital public spaces: context, purpose and audiences. *Journal of Documentation*, **79**(3): 703–17 doi:10.1108/JD-04-2022-0079.
- Borgman, C. L.** (2007). Disciplines, Documents, and Data. *Scholarship in the Digital Age*. The MIT Press, pp. 179–226 doi:10.7551/mitpress/7434.003.0011.
- Borgman, C. L.** (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, Mass.: MIT Press.
- 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](http://link.springer.com/10.1007/978-1-4020-9789-8_5) (accessed 11 November 2021).
- Bowker, G. C. and Star, S. L.** (1999). *Sorting Things out: Classification and Its Consequences*. (Inside Technology). Cambridge, MA: The MIT Press <http://cognet.mit.edu/book/sorting-things-out> (accessed 15 July 2020).
- Brinkmann, S.** (2018). The interview. In Denzin, N. K. and Lincoln, Y. S. (eds), *The SAGE Handbook of Qualitative Research*. Fifth edition (e-book). Los Angeles: SAGE, pp. 576–99.
- Brugman, H., Reynaert, M., Van Der Sijs, N., Van Stipriaan, R., Sang, E. T. K. and Van Den Bosch, A.** (2016). Nederlab: Towards a single portal and research environment for diachronic Dutch text corpora. Portoroz, pp. 1277–81 <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84996923652&partnerID=40&md5=c0acfc159740a25e2bd04fa138c679d1> (accessed 7 March 2022)
- Butigan, T., Cassidy, K., Jeszke, N., Kondratjev, P., Nowak, A., Parkoła, T., Truyen, F., Žogla, A., Davies, R., Felmayer, L., Scholz, H.** (2020). Landscape of National Aggregation in Europe and Establishment of Emerging National Aggregators. *Europeana Common Culture* [https://pro.europeana.eu/files/Europeana\\_Professional/Projectpartner/EuropeanaCommon](https://pro.europeana.eu/files/Europeana_Professional/Projectpartner/EuropeanaCommon)

- 
- CultureProjectFiles/MS3%20Landscape%20of%20national%20aggregation%20in%20Europe%20Report.pdf (accessed 12 January 2022).
- Candlin, F., Larkin, J., Ballatore, A. and Poulouvassilis, A.** (2020). *Mapping Museums 1960-2020: A Report on the Data*. London: Birkbeck, University of London  
<https://museweb.dcs.bbk.ac.uk/static/pdf/MappingMuseumsReportMarch2020.pdf> (accessed 28 February 2022).
- Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Jarita Holbrook, Lovett, R., Materechera, S., Parsons, M., et al.** (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, **19**: 43 doi:10.5334/dsj-2020-043.
- Carroll, S. R., Herczog, E., Hudson, M., Russell, K. and Stall, S.** (2021). Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Scientific Data*, **8**(1): 108 doi:10.1038/s41597-021-00892-0.
- Caswell, M.** (2016). 'The Archive' is Not An Archives: Acknowledging the Intellectual Contributions of Archival Studies. *Reconstruction*, **16**(1): 1–21.
- Caswell, M. and Cifor, M.** (2019). Neither a beginning nor an end: Applying an ethics of care to digital archival collections. In Lewi, H., Smith, W., Lehn, D. vom and Cooke, S. (eds), *The Routledge International Handbook of New Digital Practices in Galleries, Libraries, Archives, Museums and Heritage Sites*. 1st edition. London: Routledge, pp. 159–68.
- 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).
- Collections Trust** (2019). *Mapping Digitised Collections in England: Final Report*. Department for Digital, Culture, Media & Sport  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/974903/Mapping\\_digitised\\_collections\\_in\\_England\\_-\\_scoping\\_report\\_V2.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/974903/Mapping_digitised_collections_in_England_-_scoping_report_V2.pdf) (accessed 12 January 2021).
- Cowan, T. L. and Rault, J.** (2018). Onlining queer acts: Digital research ethics and caring for risky archives. *Women & Performance: A Journal of Feminist Theory*, **28**(2): 121–42  
doi:10.1080/0740770X.2018.1473985.
- Culture Grid** (2010). About *Culture Grid* <http://www.culturegrid.org.uk/about/> (accessed 17 February 2023).
- Darcovich, J., Flynn, K. and Li, M.** (2019). Born of Collaboration: The Evolution of Metadata Standards in an Aggregated Environment. *VRA Bulletin*, **45**(2): Article 5.
- Davis, E. and Heravi, B.** (2021). Linked Data and Cultural Heritage: A Systematic Review of Participation, Collaboration, and Motivation. *Journal on Computing and Cultural Heritage*, **14**(2): 1–18 doi:10.1145/3429458.
- Deazley, R.** (2017). Copyright and Digital Cultural Heritage: Introduction  
<https://copyrightcortex.org/files/copyright101/1-CDCH-Introduction.pdf> (accessed 2 June 2021).
- Dobрева, M., Stefanov, K. and Ivanova, K.** (2022). Data Spaces for Cultural Heritage: Insights from GLAM Innovation Labs. In Tseng, Y.-H., Katsurai, M. and Nguyen, H. N. (eds), *From Born-Physical to Born-Virtual: Augmenting Intelligence in Digital Libraries*, vol. 13636. (Lecture Notes in Computer Science). Cham: Springer International Publishing, pp. 492–500  
doi:10.1007/978-3-031-21756-2\_41. [https://link.springer.com/10.1007/978-3-031-21756-2\\_41](https://link.springer.com/10.1007/978-3-031-21756-2_41) (accessed 23 December 2022).
- Dutia, K. and Stack, J.** (2021). Heritage connector: A machine learning framework for building linked open data from museum collections. *Applied AI Letters*, **2**(2) doi:10.1002/ail2.23.  
<https://onlinelibrary.wiley.com/doi/10.1002/ail2.23> (accessed 27 September 2021).

- 
- Edgerton, D.** (2006). *The Shock of the Old: Technology and Global History since 1900*. London: Profile Books.
- Edwards, P. N., Jackson, S. J., Bowker, G. C. and Knobel, C. P.** (2007). *Understanding Infrastructure: Dynamics, Tensions, and Design: Report of a Workshop on "History & Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures"*.  
<https://deepblue.lib.umich.edu/bitstream/handle/2027.42/49353/UnderstandingInfrastr?sequence=3> (accessed 11 November 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** (2010). *Competitiveness and Innovation Framework Programme (CIP) ICT Policy Support Programme: ICT PSP Work Programme 2010*.  
[https://web.archive.org/web/20190607220227/http://ec.europa.eu/cip/files/docs/ict\\_psp\\_wp2010\\_en.pdf](https://web.archive.org/web/20190607220227/http://ec.europa.eu/cip/files/docs/ict_psp_wp2010_en.pdf) (accessed 22 November 2021).
- 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).
- European Commission** (2021). Commission Recommendation (EU) 2021/1970 of 10 November 2021 on a common European data space for cultural heritage <https://eur-lex.europa.eu/eli/reco/2021/1970/oj> (accessed 23 December 2022).
- European Commission** (2022). The deployment of a common European data space for cultural heritage <https://digital-strategy.ec.europa.eu/en/news/deployment-common-european-data-space-cultural-heritage> (accessed 9 May 2023).
- European Commission** (2024). The Cultural Heritage Cloud - European Commission [https://research-and-innovation.ec.europa.eu/research-area/social-sciences-and-humanities/cultural-heritage-and-cultural-and-creative-industries-ccis/cultural-heritage-cloud\\_en](https://research-and-innovation.ec.europa.eu/research-area/social-sciences-and-humanities/cultural-heritage-and-cultural-and-creative-industries-ccis/cultural-heritage-cloud_en) (accessed 21 June 2024).
- European Commission, Directorate-General for Research and Innovation, Brunet, P., De Luca, L., Hyvönen, E., Joffres, A., Plassmeyer, P., Pronk, M., Scopigno, R. and Sonkoly, G.** (2022). *Report on a European Collaborative Cloud for Cultural Heritage – Ex – Ante Impact Assessment*. Publications Office of the European Union doi:10.2777/64014.  
<https://op.europa.eu/en/publication-detail/-/publication/90f1ee85-ca88-11ec-b6f4-01aa75ed71a1/language-en#> (accessed 21 June 2024).
- Europeana** (2010). Europeana Aggregators' Handbook [https://admin.biodiversitylibrary.org/wiki-archive/mainSpace/files/Aggregators%20Handbook\\_Europeana.pdf](https://admin.biodiversitylibrary.org/wiki-archive/mainSpace/files/Aggregators%20Handbook_Europeana.pdf) (accessed 26 November 2021).
- Freire, N., Robson, G., Howard, J. B., Manguinhas, H. and Isaac, A.** (2020). Cultural heritage metadata aggregation using web technologies: IIIF, Sitemaps and Schema.org. *International Journal on Digital Libraries*, **21**(1): 19–30 doi:10.1007/s00799-018-0259-5.
- Gosling, K.** (2019). Nothing new except what has been forgotten  
<https://collectionstrust.org.uk/blog/nothing-new-except-what-has-been-forgotten/> (accessed 6 March 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).
- Greenstein, D.** (1996). Serving the Arts and Humanities. *Ariadne*(4)  
<http://www.ariadne.ac.uk/issue/4/ahds/> (accessed 3 July 2024).

- 
- Halperin, J. R.** (2017). New York's Metropolitan Museum of Art releases 375,000 digital works for remix and re-use online via CC0 - Creative Commons  
<https://creativecommons.org/2017/02/07/met-announcement/> (accessed 31 March 2023).
- 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>.
- Havens, L., Alex, B., Hosker, R., Bach, B. and Terras, M.** (2023). Collaboration Across the Archival and Computational Sciences to Address Legacies of Gender Bias in Descriptive Metadata. *Digital Humanities 2023*. Graz, pp. 267–68  
[https://www.pure.ed.ac.uk/ws/portalfiles/portal/364769281/DH2023\\_BookOfAbstracts\\_Havens.pdf](https://www.pure.ed.ac.uk/ws/portalfiles/portal/364769281/DH2023_BookOfAbstracts_Havens.pdf) (accessed 1 December 2023).
- Higgins, C.** (2023). Politicians, not curators, are to blame for the British Museum's woes. *The Guardian* <https://www.theguardian.com/commentisfree/2023/sep/01/british-museum-curators-thefts-funding> (accessed 15 November 2023).
- Hill, A.** (2002). Bringing Archives Online through the Archives Hub. *Journal of the Society of Archivists*, **23**(2): 239–48 doi:10.1080/0037981022000006408.
- Himmelsbach, E.** (2021). Collections Trust – tapping the potential of museum collection data  
<https://theodi.org/article/collection-trust-tapping-the-potential-of-museum-collection-data/> (accessed 6 March 2023).
- Historic Environment Scotland** (2023a). Welcome to Scran SCRAN <https://www.scran.ac.uk/> (accessed 27 February 2023).
- Historic Environment Scotland** (2023b). About SCRAN <https://www.scran.ac.uk/info/aboutscran.php> (accessed 27 February 2023).
- Humbel, M.** (2022). Participatory Action Research for a Digital Humanities research project: Investigating Open GLAM in the context of Social Movement Archives. *Proceedings of the Digital Humanities 2022*. Tokyo, pp. 249–52  
<https://dh2022.dhii.asia/dh2022bookofabsts.pdf>.
- Humbel, M.** (2023). The Digitisation and Open Access Politics of Social Movement Archives London: University College London <https://discovery.ucl.ac.uk/id/eprint/10163973/> (accessed 22 May 2023).
- ICA** (2000). ISAD(G): General International Standard Archival Description  
[https://www.ica.org/sites/default/files/CBPS\\_2000\\_Guidelines\\_ISAD%28G%29\\_Second-edition\\_EN.pdf](https://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_EN.pdf) (accessed 31 March 2022).
- IFLA** (2023). OAI-PMH <https://www.ifla.org/references/best-practice-for-national-bibliographic-agencies-in-a-digital-age/service-delivery/system-interfaces-and-search-protocols/oai-pmh/> (accessed 9 November 2023).
- Jules, B.** (2019). *Architecting Sustainable Futures: Exploring Funding Models in Community-Based Archives*. Shift  
<https://shiftdesign.org/content/uploads/2019/02/ArchitectingSustainableFutures-2019-report.pdf> (accessed 27 April 2021).
- Keller, P.** (2021). Five things I know about Data Spaces <https://openfuture.eu/blog/five-things-i-know-about-data-spaces/> (accessed 23 December 2022).
- King's Digital Lab** (2024). Arts and Humanities Data Service: Enabling Digital Resources for the Art and Humanities <https://ahds.ac.uk/> (accessed 3 July 2024).
- Kizhner, I., Terras, M., Rumyantsev, M., Khokhlova, V., Demeshkova, E., Rudov, I. and Afanasieva, J.** (2020). Digital cultural colonialism: measuring bias in aggregated digitized content held in Google Arts and Culture. *Digital Scholarship in the Humanities* doi:10.1093/llc/fqaa055.

- 
- Kostelic, C.** (2017). Applying the levels of conceptual interoperability model to a digital library ecosystem - A case study. vol. 2017- October. pp. 62–72  
<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85038813338&partnerID=40&md5=28ad429f49055a6ff36e4c6d85a87f66>.
- Kotarski, R., Kirby, J., Madden, F., Mitchell, L., Padfield, J., Page, R., Palmer, R. and Woodburn, M.** (2022). *Persistent Identifiers as IRO Infrastructure: A Towards a National Collection Foundation Project Final Report*. Zenodo doi:10.5281/ZENODO.6359926.  
<https://zenodo.org/record/6359926> (accessed 9 June 2023).
- MacDonald, I.** (2023). Counting when, who and how. *Journal of the History of Collections*, **35**(2): 305–20 doi:10.1093/jhc/fhac034.
- Mak, B.** (2014). Archaeology of a digitization. *Journal of the Association for Information Science & Technology*, **65**(8): 1515–26.
- Maron, D. and Feinberg, M.** (2018). What does it mean to adopt a metadata standard? A case study of Omeka and the Dublin Core. *Journal of Documentation*, **74**(4): 674–91 doi:10.1108/JD-06-2017-0095.
- Martinez, M. and Terras, M.** (2019). ‘Not Adopted’: The UK Orphan Works Licensing Scheme and how the crisis of copyright in the cultural heritage sector restricts access to digital content. *Open Library of the Humanities*, **5**(1) doi:10.16995/olh.335/.  
<https://www.research.ed.ac.uk/portal/files/94014521/MartinezEtal2019NotAdopted.pdf> (accessed 19 May 2021).
- McCarthy, D. and Wallace, A.** (2018). Survey of GLAM open access policy and practice  
[https://docs.google.com/spreadsheets/d/1WPS-KJptUJ-o8SXtg00lxcxq0IKJu8eO6Ege\\_GrLaNc/edit#gid=1216556120](https://docs.google.com/spreadsheets/d/1WPS-KJptUJ-o8SXtg00lxcxq0IKJu8eO6Ege_GrLaNc/edit#gid=1216556120) (accessed 27 February 2022).
- McCarthy, D. and Wallace, A.** (2020). Open access to collections is a no-brainer – it’s a clear-cut extension of any museum’s mission *Apollo Magazine* <https://www.apollo-magazine.com/open-access-images-museum-mission-open-glam/> (accessed 5 March 2021).
- McGillivray, B., Alex, B., Ames, S., Armstrong, G., Beavan, D., Ciula, A., Colavizza, G., Cummings, J., De Roure, D., Farquhar, A., Hengchen, S., Lang, A., Loxley, J., Goudarouli, E., Nanni, F., Nini, A., Nyhan, J., Osborne, N., Poibeau, T., Ridge, M., Ranade, S., Smithies, J., Terras, M., Vlachidis, A., Willcox, P.** (2020). The challenges and prospects of the intersection of humanities and data science: A White Paper from The Alan Turing Institute. figshare doi:10.6084/M9.FIGSHARE.12732164.V5.  
[https://figshare.com/articles/online\\_resource/The\\_challenges\\_and\\_prospects\\_of\\_the\\_intersection\\_of\\_humanities\\_and\\_data\\_science\\_A\\_White\\_Paper\\_from\\_The\\_Alan\\_Turing\\_Institute/12732164/5](https://figshare.com/articles/online_resource/The_challenges_and_prospects_of_the_intersection_of_humanities_and_data_science_A_White_Paper_from_The_Alan_Turing_Institute/12732164/5) (accessed 7 April 2021).
- McNeill, A.** (2022). *Art UK: Opening Up Access to the Nation’s Art*. Zenodo doi:10.5281/ZENODO.6334193. <https://zenodo.org/record/6334193> (accessed 23 March 2022).
- Miles, M. B., Huberman, A. M. and Saldaña, J.** (2020). *Qualitative Data Analysis : A Methods Sourcebook*. Fourth edition. Los Angeles: SAGE.
- Moulaison Sandy, H. and Freeland, C.** (2016). The Importance of Interoperability: Lessons from the Digital Public Library of America. *International Information & Library Review*, **48**(1): 45–50 doi:10.1080/10572317.2016.1146041.
- Nauta, J. G., Heuvel, W. van den, 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-](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).
- Nyhan, J. and Flinn, A.** (2016). *Computation and the Humanities: Towards an Oral History of Digital Humanities*. (Springer Series on Cultural Computing). Cham: Springer International Publishing doi:10.1007/978-3-319-20170-2. <https://link.springer.com/10.1007/978-3-319-20170-2> (accessed 15 December 2023).
- Nyhan, J., Vlachidis, A., Flinn, A., Pearlman, N., Carine, M. and Hill, J.** (2023). *Second Report - Sloane Lab: Looking Back to Build Future Shared Collections*. (Discovery Projects). London: Zenodo doi:10.5281/ZENODO.7995409. <https://zenodo.org/record/7995409> (accessed 9 June 2023).
- ODI** (2023). Shaping future services and promoting productivity with cutting-edge expertise <https://www.theodi.org/project/data-innovation-for-uk-research-and-development/> (accessed 5 June 2023).
- OpenGLAM Initiative** (2023). What <https://openglam.org/what/> (accessed 7 June 2023).
- Ortolja-Baird, A. and Nyhan, J.** (2022). Encoding the haunting of an object catalogue: on the potential of digital technologies to perpetuate or subvert the silence and bias of the early-modern archive. *Digital Scholarship in the Humanities*, **37**(3): 844–67 doi:10.1093/lc/fqab065.
- Ortolja-Baird, A., Pickering, V., Nyhan, J., Sloan, K. and Fleming, M.** (2019). Digital Humanities in the Memory Institution: The Challenges of Encoding Sir Hans Sloane’s Early Modern Catalogues of His Collections. *Open Library of Humanities*, **5**(1): 44 doi:10.16995/olh.409.
- Padfield, J., Bolland, C., Fitzgerald, N., McLaughlin, A., Robson, G. and Terras, M.** (2022). *Practical Applications of IIF as a Building Block towards a Digital National Collection*. Zenodo doi:10.5281/ZENODO.6884885. <https://zenodo.org/record/6884885> (accessed 9 June 2023).
- Padilla, T., Allen, L., Frost, H., Potvin, S., Russey Roke, E. and Varner, S.** (2019). Final Report --- Always Already Computational: Collections as Data. Zenodo doi:10.5281/zenodo.3152935. <https://zenodo.org/record/3152935> (accessed 22 March 2021).
- 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).
- Paltrinieri, C.** (2023). *Consolidation Report: Insights from Towards a National Collection Foundation Projects*. Zenodo doi:10.5281/ZENODO.7674816. <https://zenodo.org/record/7674816> (accessed 9 June 2023).
- Pandey, R. and Kumar, V.** (2020). Exploring the impediments to digitization and digital preservation of cultural heritage resources: A selective review. *Preservation, Digital Technology and Culture*, **49**(1): 26–37 doi:10.1515/pdte-2020-0006.
- Patton, M. Q.** (2015). *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*. Fourth edition (e-book). Thousand Oaks: SAGE Publications, Inc.
- Pavis, M. and Wallace, A.** (2019). Response to the 2018 Sarr-Savoy Report: Statement on Intellectual Property Rights and Open Access relevant to the digitization and restitution of African Cultural Heritage and associated materials. doi:10.5281/zenodo.2620596. <https://zenodo.org/record/2620596> (accessed 6 February 2020).
- Pawlicka-Deger, U.** (2021). Infrastructuring digital humanities: On relational infrastructure and global reconfiguration of the field. *Digital Scholarship in the Humanities*, **37**(2): 534–50 doi:10.1093/lc/fqab086.
- Pekel, J.** (2013). Case Study: Rijksmuseum releases 111.000 high quality images to the public domain <https://openglam.org/2013/02/27/case-study-rijksmuseum-releases-111-000-high-quality-images-to-the-public-domain/> (accessed 31 March 2023).
- Peters, R. and Kalshoven, L.** (2016). *What Rights Clearance Looks like for Cultural Heritage Organisations – 10 Case Studies*. (Europeana Factsheet)



- 
- [https://pro.europeana.eu/files/Europeana\\_Professional/IPR/160331rights\\_clearance\\_case\\_studies\\_public.pdf](https://pro.europeana.eu/files/Europeana_Professional/IPR/160331rights_clearance_case_studies_public.pdf) (accessed 20 May 2021).
- Petrus, A., Wildi, T. and Müller, S.** (2023). Preproject 'Swiss Virtual Natural History Collection'. *Database*, **2023**: 1–9 doi:10.1093/database/baad072.
- Poikola, A., Verdonck, B. and Joosten, R. (eds).** (2023). DSSC Glossary Data Spaces Support Centre (DSSC)  
<https://dssc.eu/space/Glossary/55443460/DSSC+Glossary+%7C+Version+1.0+%7C+March+2023?attachment=/rest/api/content/55443460/child/attachment/att110362680/download&type=application/pdf&filename=DSSC-Data-Spaces-Glossary-v1.0.pdf> (accessed 8 November 2023).
- Ponte, D. and Camussone, P. F.** (2013). Neither Heroes nor Chaos: The Victory of VHS against Betamax. *International Journal of Actor-Network Theory and Technological Innovation*, **5**(1): 40–54 doi:10.4018/jantti.2013010103.
- Poole, N.** (2015). Guest blog: Aggregation & the Culture Grid *Museum Computer Group*  
<https://museumscomputergroup.org.uk/culture-grid/> (accessed 27 February 2023).
- Preuss, U.** (2016). Sustainable Digitalization of Cultural Heritage - Report on Initiatives and Projects in Brandenburg, Germany. *Sustainability*, **8**(9): 891 doi:10.3390/su8090891.
- Pringle, E., Mavin, H., Greenhalgh, T., Dalal-Clayton, A., Rutherford, A., Bramwell, J., Blackford, K. and Balukiewicz, K.** (2022). *Provisional Semantics: Addressing the Challenges of Representing Multiple Perspectives within an Evolving Digitised National Collection*. Zenodo doi:10.5281/ZENODO.7081347. <https://zenodo.org/record/7081347> (accessed 9 June 2023).
- Punch, K.** (2014). *Introduction to Social Research: Quantitative & Qualitative Approaches*. Third edition. Los Angeles, California: SAGE.
- Rees, G., Gadd, S., Horgan, J., Hunt, A., Isaksen, L., Morris, V., Musson, A., Simon, R., Strachan, P. and Vitale, V.** (2022). *Locating a National Collection (LaNC)*. Zenodo doi:10.5281/ZENODO.7071654. <https://zenodo.org/record/7071654> (accessed 9 June 2023).
- Renshaw, C. and Liew, C. L.** (2021). Descriptive standards and collection management software for documentary heritage management: attitudes and experiences of information professionals. *Global Knowledge, Memory and Communication*, **70**(8/9): 697–713 doi:10.1108/GKMC-08-2020-0129.
- Rockwell, G.** (2010). As Transparent as Infrastructure: On the research of cyberinfrastructure in the humanities. In McGann, J. (ed), *Online Humanities Scholarship: The Shape of Things to Come*. Houston: Rice University, pp. 613–30 <http://cnx.org/content/m34315/1.2/> (accessed 16 March 2020).
- SAA** (2023). International Standard Archival Description (General) [ISAD(G)]  
<https://www2.archivists.org/groups/standards-committee/international-standard-archival-description-general-isadg> (accessed 9 November 2023).
- Saldaña, J.** (2016). *The Coding Manual for Qualitative Researchers*. Third edition. Los Angeles: SAGE.
- Sanderhoff, M.** (2014). This belongs to you: On openness and sharing at Statens Museum for Kunst. In Sanderhoff, M. (ed), *Sharing Is Caring: Openness and Sharing in the Cultural Heritage Sector*. Copenhagen: Statens Museum for Kunst, pp. 20–131.
- Sherratt, T.** (2015). On seams and edges: dreams of aggregation, access and discovery in a broken world. *ALIA Information Online*. Sydney: Zenodo doi:10.5281/ZENODO.3556475. <https://zenodo.org/record/3556475> (accessed 28 February 2023).
- Shirazi, R.** (2018). Reproducing the Academy: Librarians and the Question of Service in the Digital Humanities. In Sayers, J. (ed), *Making Things and Drawing Boundaries: Experiments in the Digital Humanities*. University of Minnesota Press, pp. 86–94 doi:10.5749/j.ctt1pwt6wq. <http://www.jstor.org/stable/10.5749/j.ctt1pwt6wq> (accessed 26 March 2022).

- 
- Siebinga, S., Manghi, P., Mieldijk, M. and Van Der Werf, T.** (2012). *HOPE: Mission, Technical Vision and High-Level Design of the Architecture*.  
[http://www.peoplesheritage.eu/pdf/D2\\_1\\_Grant250549\\_HOPE\\_V2-0.pdf](http://www.peoplesheritage.eu/pdf/D2_1_Grant250549_HOPE_V2-0.pdf) (accessed 29 November 2021).
- Silva, G. M. da, Glória, A. C., Salgueiro, Â. S., Almeida, B., Monteiro, D., Freitas, M. R. de and Freire, N.** (2022). ROSSIO Infrastructure: A Digital Humanities Platform to Explore the Portuguese Cultural Heritage. *Information*, **13**(2): 50 doi:10.3390/info13020050.
- Sloan, K. and Nyhan, J.** (2021). Enlightenment architectures: the reconstruction of Sir Hans Sloane's cabinets of 'Miscellanies'. *Journal of the History of Collections*, **33**(2): 199–218 doi:10.1093/jhc/fhaa034.
- Star, S. L. and Ruhleder, K.** (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, **7**(1): 111–34.
- Star, S. L. and Strauss, A.** (1999). Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. *Computer Supported Cooperative Work (CSCW)*, **8**(1–2): 9–30 doi:10.1023/A:1008651105359.
- Stevenson, J.** (2019). Data Processing and Sustainability with a Large-Scale Aggregator: the UK Archives Hub. (Ed.) Depoortere, R., Gheldof, T., Styven, D. & Van Der Eycken, J. , **Trust and Understanding: the value of metadata in a digitally joined-up world**: 93–103.
- Stobo, V.** (2016). Risky Business: Copyright and Making Collections Available Online. In Wallace, A. and Deazley, R. (eds), *Display at Your Own Risk: An Experimental Exhibition of Digital Cultural Heritage*. Glasgow: CREATE, pp. 281–87 <https://displayatyourownrisk.org/wp-content/uploads/2016/04/Display-At-Your-Own-Risk-Publication.pdf> (accessed 2 June 2021).
- TaNC** (2023). About us <https://www.nationalcollection.org.uk/about> (accessed 7 December 2023).
- Terras, M.** (2011). The Rise of Digitization. In Rikowski, R. (ed), *Digitisation Perspectives*, vol. 46. (Educational Futures Rethinking Theory and Practice). Rotterdam: SensePublishers, pp. 3–20 doi:10.1007/978-94-6091-299-3\_1. [https://link.springer.com/chapter/10.1007/978-94-6091-299-3\\_1#citeas](https://link.springer.com/chapter/10.1007/978-94-6091-299-3_1#citeas) (accessed 5 July 2024).
- Terras, M., Coleman, S., Drost, S., Elsdén, C., Helgason, I., Lechelt, S., Osborne, N., et al.** (2021). The value of mass-digitised cultural heritage content in creative contexts. *Big Data & Society*, **8**(1): 1–15 doi:10.1177/20539517211006165.
- Thomer, A. K. and Rayburn, A. J.** (2023). 'A Patchwork of Data Systems': Quilting as an Analytic Lens and Stabilizing Practice for Knowledge Infrastructures. *Science, Technology, & Human Values*, **20**(10): 1–30 doi:10.1177/01622439231175535.
- Thylstrup, N. B.** (2018). *The Politics of Mass Digitization*. Cambridge, MA: The MIT Press.
- Timms, K.** (2009). New Partnerships for Old Sibling Rivals: The Development of Integrated Access Systems for the Holdings of Archives, Libraries, and Museums. *Archivaria*, **68**: 67–95.
- Tzouganatou, A.** (2021). On Complexity of GLAMs' Digital Ecosystem: APIs as Change Makers for Opening up Knowledge. In Rauterberg, M. (ed), *Culture and Computing. Design Thinking and Cultural Computing*, vol. 12795. (Lecture Notes in Computer Science). Cham: Springer, pp. 348–59 doi:10.1007/978-3-030-77431-8\_22. [https://link.springer.com/10.1007/978-3-030-77431-8\\_22](https://link.springer.com/10.1007/978-3-030-77431-8_22) (accessed 13 June 2023).
- UKRI** (2021). Artificial intelligence supports culture and heritage exploration <https://www.ukri.org/news/artificial-intelligence-supports-culture-and-heritage-exploration/> (accessed 5 June 2023).
- Van de Sompel, H. and Nelson, M. L.** (2015). Reminiscing About 15 Years of Interoperability Efforts. *D-Lib Magazine*, **21**(11/12) doi:10.1045/november2015-vandesompel. <http://www.dlib.org/dlib/november15/vandesompel/11vandesompel.html> (accessed 21 March 2022).

- 
- Van Hooland, S. and Verborgh, R.** (2014). *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata*. London: Facet Publishing.
- Van Strien, D., Beelen, K., Ardanuy, M., Hosseini, K., McGillivray, B. and Colavizza, G.** (2020). Assessing the Impact of OCR Quality on Downstream NLP Tasks. *Proceedings of the 12th International Conference on Agents and Artificial Intelligence*. Valletta, Malta: SCITEPRESS - Science and Technology Publications, pp. 484–96 doi:10.5220/0009169004840496. <http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0009169004840496> (accessed 30 June 2022).
- Wallace, A.** (2020a). Words Mean Things (A Glossary). *Open GLAM* doi:10.21428/74d826b1.51566976. <https://openglam.pubpub.org/pub/the-glossary> (accessed 27 October 2021).
- Wallace, A.** (2020b). Copyright. *Critical Open GLAM: Towards [Appropriate] Open Access for Cultural Heritage* doi:10.21428/74d826b1.556f5733. <https://openglam.pubpub.org/pub/background-copyright> (accessed 3 November 2020).
- Wallace, A.** (2020c). Introduction. *Critical Open GLAM: Towards [Appropriate] Open Access for Cultural Heritage* doi:10.21428/74d826b1.be9df175. <https://openglam.pubpub.org/pub/introduction-to-critical-open-glam> (accessed 29 October 2020).
- Wallace, A.** (2021). Decolonization and Indigenization. *Critical Open GLAM: Towards [Appropriate] Open Access for Cultural Heritage* <https://openglam.pubpub.org/pub/decolonization> (accessed 5 January 2023).
- Wallace, A.** (2022). *A Culture of Copyright: A Scoping Study on Open Access to Digital Cultural Heritage Collections in the UK*. Zenodo doi:10.5281/ZENODO.6242611. <https://zenodo.org/record/6242611> (accessed 23 March 2022).
- Wernimont, J.** (2021). Listening, Care, and Collections as Data. *Journal of Critical Digital Librarianship*, **1**(1): 23–42 doi:10.31390/jcdl.1.1.04.
- Winters, J., Stack, J., Dutia, K., Unwin, J., Lewis, R., Palmer, R. and Wolff, A.** (2022). *Heritage Connector: A Towards a National Collection Foundation Project Final Report*. Zenodo doi:10.5281/ZENODO.6022678. <https://zenodo.org/record/6022678> (accessed 9 June 2023).
- Zaagsma, G.** (2023). Digital History and the Politics of Digitization. *Digital Scholarship in the Humanities*, **38**(2): 830–51 doi:10.1093/llc/fqac050.
- Ziegler, S. L.** (2020). Open Data in Cultural Heritage Institutions: Can We Be Better Than Data Brokers?. *Digital Humanities Quarterly*, **14**(2) <http://www.digitalhumanities.org/dhq/vol/14/2/000462/000462.html> (accessed 11 September 2020).
- Zoller, G. and DeMarsh, K.** (2013). For the Record: Museum Cataloging from a Library and Information Science Perspective. *Art Documentation: Journal of the Art Libraries Society of North America*, **32**(1): 54–70 doi:10.1086/669989.

---

## Appendix A: Interview questionnaire collections-holding organisations

### Scope

The purpose of this interview is to learn from your experience, needs, and aspirations regarding aggregated heritage collections. In the context of this interview, we mean with aggregator a digital service that unifies heritage collections and makes them available online. An aggregator may unify from different heritage organisations catalogue descriptions only, or catalogue descriptions and digitised objects. Example aggregators include Europeana, ArtUK, or the UK Archives Hub.

### 1. Introductory questions

1.1 Please tell me briefly about your role in your organisation?

1.2 Could you please speak about your organisation's strategy for making collections available online, and if relevant, linking them to aggregators?

1.3 Could you please speak about your experience in working with projects that aggregate digital collections or catalogues?

1.4 What are your organisation's aspirations to aggregate digital collections? For instance, to:

- Connect collections thematically?
- Connect collections within your own institution?
- Connect collections with other organisations?

1.5 Do you engage users in the co-development of your digital collection, and if so, what has been your experience been in this regard?

### 2. Rationale for using aggregator services

2.1 In what ways has your digitisation programmes informed the appetite in aggregating digital collections?

2.2. In what ways has the engagement, digital or face-to-face, with your audiences informed the appetite in aggregating digital collections?

2.3 In what ways have existing aggregating partnerships, you are part of, fulfilled your strategic objectives?

2.4 Which factors would you say make you trust a collections aggregator?

### 3. Nature of collaboration through aggregator services

3.1 Who is or would in your organisation be responsible for the aggregation work?

3.2 What kind of reporting (like statistics on use) are you getting from your aggregator, or would you like to receive from an aggregator?

3.3 Could you please tell me whether the partnerships you have been involved in through aggregation projects led to other forms of collaborations with other organisations or the aggregator too?

---

3.4 How do you perceive the potential that aggregated collections could facilitate additional forms of collaborations with other organisations within the same aggregator?

3.5 From your perspective should an aggregator be responsible to provide services like for instance:

- Support for collaborating with other content providers?
- Training and capacity building?
- Technical services your organisation may not be able to host on its own?

3.6 What support would the sector need to provide in legal, technical, or organisational terms to help your organisation in participating in aggregation projects?

3.7 Who funded your aggregation, or who would need to fund your participation in aggregation projects in future?

#### **4. Technical capabilities and interoperability**

4.1 In which formats are you able to export your metadata (catalogue descriptions)?

4.2 To which extent are you able to automate the process of metadata export, and if there are any, what are the challenges?

4.3 What kind of technical investments would be needed for facilitating aggregation in your organisation?

4.4 To what extent would you be willing to invest in relevant technologies. For instance:

- The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)?
- The International Image Interoperability Framework (IIIF)?
- Linked Open Data technologies (LOD)?
- Others?

#### **5. Use of aggregated collections and technical enhancement**

5.1 Which kind of end-users do you have, or would you have, in mind when submitting collections to aggregators?

5.2 How do these users differ from users that access your collection through your own catalogue?

5.3 Could you please tell me whether you have noticed a change of user behaviour due to submitting collections to aggregators?

#### **6. Licensing policies**

6.1 Do you use online licensing schemes for your metadata and digital images, and if so, which licenses do you use when making your collections available online?

6.2 Open Access to cultural heritage is defined as “[...] a policy or practice that allows reuse and redistribution of materials for any purpose, including commercial”<sup>25</sup> Could you please tell me whether your organisation has any licensing agreements with third parties that may impede Open Access?

---

6.3 Are there any other reasons or concerns that may impede your organisation's ability to participate in Open Access frameworks?

6.4 What is your organisation's appetite for risk between 1 and 3 in making collections available online where you cannot identify copyright holders?

1 zero appetite for risk = only make accessible images where rightsholders have provided consent.

2 medium appetite = only make accessible images where rightsholders have provided consent OR works are identified as orphaned through due diligence.

3 High appetite for risk = make images accessible regardless of due diligence process. Risk managed by insurance policy and reliance on cease and desist and take down messages.

## **7. Future aggregator services and visibility**

7.1 What would in your opinion be an ideal funding model for creating and participating aggregated collections?

7.2 How do you, or would you, measure your return of investment in contributing to aggregated collections? For instance, in terms of:

- Visits to your physical collection
- Visits to your digital collection
- Attribution
- Additional revenue

7.3 Based on the information you have at hand could you please tell me what your understanding of Towards a National Collection (TaNC) is? (There is no right or wrong answer)

7.4 Could you please say something more about what do you perceive the relevance of TaNC to be for your organisation, or the sector?

7.5 In your opinion, is there a need, or wish, for a 'national collection'?

## **8. Closing question**

8.1 Is there anything we have not covered yet that you would like to talk about, in relation to the areas we discussed?

---

## Appendix B: Interview questionnaire aggregators

### Scope

The purpose of this interview is to learn from your experience and aspirations in creating aggregated heritage collections. In the context of this interview, we mean with aggregator a digital service that unifies heritage collections and makes them available online. An aggregator may unify from different heritage organisations catalogue descriptions only, or catalogue descriptions and digitised objects. Example aggregators include Europeana, ArtUK, or the UK Archives Hub.

### 1. Introductory questions

1.1 Please tell me briefly about your role at your aggregator service?

1.2 What would you say are the main opportunities for creating aggregated collections?

1.3 What would you say are the main challenges for creating aggregated collections?

### 2. Rationale for creating aggregator services

2.1 What are your aggregator's aspirations to create aggregated collections?

2.2 Could you please speak about how it came historically to the creation of your aggregator service? For instance:

- Digitisation funding programmes
- Institutional collaboration and networks
- Updating an existing catalogue or portal

2.3 What would you say informs the appetite of your partner organisations to (content providers) participate in the aggregator service?

### 3. Nature of collaboration within aggregator services

3.1 What is the governance structure of your aggregator service and what is the extent of content providers' involvement in decision making?

3.2 Could you please speak about which aspects of the aggregation work is facilitated by you, and which work needs to be done by the content providers?

3.3 What kind of reporting (like statistics on use) can you give to content providers?

3.4 Could you please tell me whether your aggregation work has led to other forms of collaborations among your content providers, and if so, which are these other forms of collaborations?

3.5 In your experience, what is the potential for aggregated collections to facilitate additional forms of collaboration among content providers?

3.6 From your perspective, should an aggregator be responsible to provide services like for instance:

- Support for collaboration among content providers?
- Training and capacity building?
- Technical services that content providers may not be able to host on their own?

---

3.7 What support would the sector need to provide in legal, technical, or organisational terms to help content providers in participating in aggregation projects?

#### **4. Technical approaches and challenges**

4.1 In which formats do content providers need to export their metadata?

4.2 To which extent are you able to automate the aggregation process, and if there are any, what are the challenges?

4.3 Could you please tell me about the main challenges you experience in mapping the metadata of content providers to your data model, and how you approach these challenges?

4.4 Which kind of technologies do you perceive as the most relevant content providers need to invest in for participating in aggregation projects?

- The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)?
- The International Image Interoperability Framework (IIIF)?
- Linked Open Data technologies (LOD)?
- Others?

#### **5. Use of aggregated collections**

5.1 What kind of end-users do you have in mind for your aggregation service?

5.2 How do these users differ from users that access collections from the content providers' catalogue directly?

5.3 Which tools and methods are from your experience necessary to make aggregated collections data useful to users?

#### **6. Licensing policies**

6.1 Which kind of licensing policy does your aggregator service has for metadata and digital images?

6.2 Open Access to cultural heritage is defined as “[...] a policy or practice that allows reuse and redistribution of materials for any purpose, including commercial.”<sup>25</sup> Could you tell me from your experience when working with content providers, whether Open Access frameworks provide a challenge, and if so, how do you as an aggregator mitigate these?

6.3 What is your aggregator's appetite for risk in making collections available online where the content provider is not able to identify copyright holders (between 1 – no appetite- and 3 – big appetite)?

1 zero appetite for risk = only make accessible images where rightsholders have provided explicit consent.

2 medium appetite = only make accessible images where rightsholders have provided explicit consent OR works that are identified as orphaned through due diligence.

3 High appetite for risk = make images accessible regardless of due diligence process. Risk managed through reliance on cease and desist and take down messages.



---

## **7. Funding and visibility of content providers**

7.1 Who funds your aggregator service and how do you ensure its sustainability?

7.2 What would in your opinion be an ideal funding model for creating aggregated collections?

7.3 How do you ensure that content providers remain visible as individual organisations within the aggregate, and are able to report on the impact of their contributions?

## **8. Closing question**

8.1 Is there anything we have not covered yet that you would like to talk about, in relation to the areas we discussed?