



# Human Perspectives and Social Infrastructures: Prioritising People in GLAM Digitisation

DISCUSSION PAPER

MEGAN GOOCH 

REBECCA KAHN 

EDURNE KUGELER 

\*Author affiliations can be found in the back matter of this article

 ubiquity press

## ABSTRACT

Much discussion in current digital humanities research and funding is concerned with creating, using and maintaining technical and research infrastructures. These large-scale projects are often ambitious – designed to bring tools, data and researchers together to pool resources, work across silos, leverage regional competencies and avoid duplication of effort. Technologies such as the semantic web and linked data are deployed to collate diverse collections into cross-institutional platforms. National and supra-national bodies earmark funds to develop portals enabling access to aggregated digital heritage, and shared data spaces.

This consolidation is often driven by calls for broader access to materials, wider accountability from holding institutions to a range of different publics, and the need for reproducibility – particularly for publicly funded projects and institutions. This is paired with increased demand for more quantified metrics and approaches within heritage digitisation. The massive volume of material coming online every year is not always easy to find, maintain or use.

Based on our individual (and shared) experiences of working in an academic library, a national museum and a university digital humanities group, we argue that it is essential to consider the public as a network, as we would in terms of infrastructure. It is this network which should drive thinking on how digitisation is done, and how infrastructures are developed. We argue that modern information organisation standards are **not** necessarily human oriented, and that we must take consideration and care to ensure that social infrastructures are not overlooked in these contexts. The FAIR and CARE principles are not always easy to apply to massive collections of (often heterogenous) materials, and additional guidelines and best-practice approaches need to be developed, if we want to allow the users of digital collections to experience the transformative moments in collections.

## CORRESPONDING AUTHOR:

**Rebecca Kahn**

Department of History,  
University of Vienna,  
Vienna, Austria

[rebecca.kahn@univie.ac.at](mailto:rebecca.kahn@univie.ac.at)

## KEYWORDS:

museums; libraries;  
digital humanities; digital  
infrastructure; digital  
collections

## TO CITE THIS ARTICLE:

Gooch, M., Kahn, R., & Kugeler, E. (2025). Human Perspectives and Social Infrastructures: Prioritising People in GLAM Digitisation. *Journal of Open Humanities Data*, 11: 13, pp. 1–12. DOI: <https://doi.org/10.5334/johd.274>

## (1) CONTEXT AND MOTIVATION

### (1.1) INTRODUCTION

Digitisation has changed the humanities in the 21st century. When discussing<sup>1</sup> this, the focus often falls on universities and academic researchers. The establishment of the field of Digital Humanities is a case in point. Less focus is given, however, to Galleries, Archives, Libraries and Museums (GLAMs) – institutions which have facilitated academic research and made their own scholarly contributions. They are directly impacted by and implicated in the evolution towards digital methods. Namely, they are expected to digitise their collections as completely and quickly as possible, whilst adhering to (or sometimes even establishing) international best practice examples and standards of interoperability – such as the Collections as Data initiative in the US, which helped to define conceptual and practical approaches for publishing GLAM collections in computational formats (Padilla et al., 2019), and the global GLAM Labs community, which has published a checklist for institutions who wish to publish their collections as data (Candela, Gabriëls, et al., 2023).

Indeed, the idea of making collections available to many through digitisation and online publication is exciting (including to us, the authors). It holds the promise of reaching both specialised audiences and an ever-wider array of publics who will use the collections to build new knowledge and skills through activities like text searches and comparing metadata, and all from the comfort of one's home or office.

What is sometimes forgotten, in this relatable excitement, is the work and resources required for successful and useful digitisation. Digitising collections – even making digital inventories of them – requires skills and resources that GLAM organisations have had to build up in recent times (Towards a National Collection, 2024). They have had to find, fund, and often train people who are able to formulate requirements, handle technical infrastructure, work with the new tools, know international best practices. They have had to source, and sometimes develop, adequate software that suits their traditional inventorying practices as well as being compatible with the potential of the interoperable digital world. They have had to acquire technical infrastructure that allows them to host the digitised and digital material, as well as maintain and expand this infrastructure to fit evolving needs. And for this, they have had to develop new internal strategies and, not least, allocate new budgets (Chambers et al, 2022).

In addition to the resources required for digitising in the first place, digitisation often comes with the expectation that the materials will then forever remain accessible. Ensuring long-term accessibility is a complex and resource-intensive undertaking – digital materials will deteriorate beyond readability or usability if left alone for too long. The long-term care of digital materials usually falls to GLAM institutions (including university libraries and archives) who, as with digitisation, have had to develop capacity to do this.

All of this to say: digitisation, especially if it is done sustainably, requires specialist resources. It is therefore necessary, in our opinion, to make sure the balance between the exciting new possibilities and the effort it requires to make them a reality is kept. GLAM organisations must ensure digitisation is “worth it”.

### (1.2) CONTEXT

As a term, ‘digitisation’ does a great deal of heavy lifting in general discourse. It is a catch-all, used to describe a range of processes, from the mass scanning of hundreds of thousands of printed pages in newspaper collections, to creating high-resolution 3D imaging of the interior of archaeological sites (*Explore Kasematten Niedergrünwald-Olizy in 3D*, n.d.), infrared analysis of medieval documents (Barrett, 2023), photographing 3D museum objects (MNAHA, n.d. b) and enriching them with detailed metadata and the ingestion of petabytes of data into aggregated repositories (*Europeana*, n.d.). As in all things, it is important to be specific when discussing what kinds of digitisation we are concerned with, not only for the sake of clarity, but because different types of digitisation require different approaches from the institutions involved and the publics they serve.

---

<sup>1</sup> As part of the special issue on Amplifying GLAM Collections, this full length research paper does not describe a dataset, but discusses challenges, and limitations in the creation, collection, and management of data in Humanities research and the GLAM sector. Hence our departure from the usual JoHD data paper format.

Although it is useful to conceptualise GLAM as a sector comprising memory and heritage institutions with similar missions to curate, preserve and engage with traces of the past, this conception can often flatten some of the significant differences between the institutional types (Cannon, 2013; Logan & Liew, 2023). An earlier iteration of the University of Oxford's GLAM Digital Strategy aimed for the creation of a digital surrogate for every item in its collections. Something of a stretch target for the University's four museums, but not achievable for the Bodleian Libraries where there is estimated to be more than 13 million printed items collection, most of which with hundreds of pages per object. Archives and libraries will often have collections which vary by a factor of hundreds of thousands, to millions, from the numbers found in museums, galleries and heritage sites. The complexities of different object forms also necessitate different digitisation approaches. A complex museum object may require multiple images or 3D photogrammetry, an archival collection will need connections to be made in the metadata between different parts of a series or collection, and a series of periodicals in a library or multi-volume work within a bound book (bound-withs) require complex metadata to interpret and organise any resultant images.

At the most general level, it is also useful to be clear about the distinction between mass and critical digitisation (Dahlström et al., 2012, susan pui san lok, 2024). Mass digitisation is primarily automated, produces clones of the original, and is agnostic to the version or edition of the primary source. Mass projects aim to be exhaustive and quantitative, and focus on common, regular and foreseeable traits in a large number of documents. By contrast, critical digitisation processes recognise the distortion digitisation brings about, and are the result of selective analyses of their sources. Their focus is qualitative, with a concentration on what makes an item unique and prioritises depth over scale. Both approaches make a claim to completeness, which is, of course, impossible.

While an overview of the trends in recent years reveals a shift towards large-scale humanities infrastructure projects, built on the technical promises of big linked data technologies (Barbot et al., 2019) we contend that these infrastructures are, in fact, inseparable from the digital assets they purport to contain, or host. Aggregated infrastructures and individual digital collections have common objectives – the hosting, preservation and publication of digitised cultural heritage. Examples of national and international research projects and infrastructures in Europe and the UK include DARIAH (the Digital Research Infrastructure for the Arts and Humanities) CLARIN-ERIC (Common Language Resources and Technology Infrastructure – European Research Infrastructure Consortium), Europeana, EHRI (European Holocaust Research Infrastructure) and Towards a National Collection (*Clariah – Common Lab Research Infrastructure for the Arts and Humanities, n.d.; CLARIN ERIC, n.d.; Europeana, n.d.; Towards a National Collection | Collections United, n.d.*). These infrastructures have obvious practical, financial and social appeal to heritage institutions. They enable a pooling of skills and resources, the economies of scale make the budgets in GLAMs go further, and they are a way for organisations to show value at an (inter)national level by reaching larger publics. In addition, the conceptual benefits include the ability to connect collections, objects and thematic areas which have previously been scattered throughout different institutions. These large infrastructures aim to create one-stop shops for humanities scholarship, as well as for other audiences. However, they also present the risk that some people, whether they be the users of data, producers of data, or heritage professionals, get left behind in the larger ambitions.

We look at how large-scale infrastructure projects – Europeana in particular – relate and compare to the smaller-scale projects they are often composed of, like the ones that exist in our own individual GLAM institutions. Because of our knowledge of working on them we have chosen to focus on digitization projects from two of our home-institutions, the Bodleian Libraries and the National Museum of Archaeology, History and Art of Luxembourg (MNAHA). In this comparison, we consider Europeana as an attempt to gather and provide access to many smaller digitisation projects, via one portal. While the comparison will be framed by the overarching question of what makes digitisation “worth it”, this range of cases also allows us to examine what elements are enhanced and what elements are lost in projects of differing size.

## (2) DISCUSSION

### (2.1) DEPTH VS BREADTH OF CONTENT

As we have said in our introduction, the promise of digitisation is providing “universal access to all knowledge” — incidentally also the Internet Archive's stated mission (*Internet Archive,*

n.d.) — and via this access, lead to the creation of ever new knowledge. The limiting factors of people, time and technology make it necessary, however, to carefully question, and add nuance to, this utopic yet persistent idea.

The phrase digitisation itself suggests passive technical automation, and elides the work of the people that undertake the many tasks and workflows that create digital collections (Zeffiro, 2019). We think that looking at digitisation in concrete contexts and from a human centred-perspective will bring us closer to doing it in a way that is “worth it” for those that we want to benefit from it.

The idea of providing “universal access to all knowledge” is often translated into attempting to make the largest possible number of digitised items available in as short an amount of time as possible and as widely as possible (Gibson & Kahn, 2016). However, the very real constraints on resources in the GLAM sector means choices have to be made on what is sacrificed for the sake of volume. Europeana, for example, has a relatively low bar for metadata standards to be able to deposit (Isaac & Clayphan, 2013). The resulting collections on this pan-European portal show the breadth of collections across the continent, but little of the depth (Capurro & Plets, 2020). Similarly the new UK Museum Data Service has been created to offer a low bar to entry for small and medium-sized museums to store their metadata, but crucially, this is not intended as a repository for public consumption, but a service for museums to store, curate and exchange metadata (*Museum Data Service*, n.d.). The British Museum made a decision to publish nearly 2 million objects on its public-facing collections online platform by 2009, irrespective of whether there were full records or images (Ross & Terras, 2011; Szrajber, 2008) but as they appeared in the internal museum database, which led to disparities in data and metadata quality, notably the absence of images for many records (Griffiths, 2010).

While this might technically fulfill the “requirement” of giving access to collections, does it really? Incomplete or erroneous metadata can mean that, even though an item is on an online platform it still cannot be found by the people who are looking for it. And, if it is found, the person looking at the item might then not get the information they need to be able to interpret, and thus benefit from, it. One might also be disappointed by the lack of images if – as many of us are – one is accustomed to ‘click on it and see’. One approach to mitigating these tradeoffs and tensions is to consider the ‘slow digitisation’ approach suggested by Prescott and Hughes (2018), which advocates the use of technology to gradually “...excavate the complex layers...” of an object, rather than racing to provide speedy access at a superficial level.

The slower, more critical approach can be found in digitisation projects which prioritise high quality both in their metadata records and images. One such example can be seen in the University of Oxford’s Digital Bodleian system, with over one million images (*Digital Bodleian*, n.d.). This approach, using IIF infrastructure is not unique to Oxford, and indeed IIF was the subject of a 2020 UK Arts and Humanities Research Council-funded project scoping its applicability for a national infrastructure (Padfield et al., 2022). Another example is the Luxembourg National Museum of Archaeology, History and Art’s (MNAHA) online collections platform. Each object published on it has extensive metadata in three languages (English, French and German) which is carefully checked prior to online publication, as well as catalogue-quality high resolution images that can be downloaded whenever copyright restrictions allow.

Being this thorough before publishing digitised objects of course means that it takes much longer to reach a similar quantity of publicly accessible items as a mass digitisation project would. It took almost five years for the MNAHA to publish its 10,000th object, of an estimated 300,000 in the collection total (Kugeler, 2024a, 2024b). This means that very human choices have to be made about what objects are prioritised for publication and the lack of transparent criteria when making these choices might lead to biases. This approach also requires a large investment in terms of personnel resources. In addition to people managing the platform, experts with different skills, such as handling of objects, photography, subject matter etc. are needed to ensure the high quality standards.

## (2.2) OPEN ACCESS VS LEGAL CONSTRAINTS

Beyond the issues of choosing between breadth and depth whilst digitising and publishing collections, there are other concerns and constraints. A push for open publication of digitised collections, making them free to view and (in many cases) download has coincided with the

push for open access and open scholarship. In national and publicly funded museums, there are political and policy drives to make collections available to the publics whose taxes fund them, and similar policies and funder mandates exist to make the results of publicly funded research publicly available. However, the issue of copyright, IP and personal data legislation can hinder these efforts. Recent collection items such as archival papers of living individuals, literary outputs still within copyright, and artistic or technological objects with extant intellectual property restrictions put very real limitations on publishing some collection items. The fear of transgressing these complex laws is also sometimes more limiting than the laws themselves (Wallace, 2022).

The recent rise of AI has raised some additional issues for GLAM institutions. AI and technology companies have traditionally used means such as the Common Crawl to harvest data on the internet (*Common Crawl – Open Repository of Web Crawl Data*, n.d.), as well as using their own proprietary methods. Any content openly and publicly available on the web, has already been harvested. Yet we are in a moment of reckoning with AI, with authors and creators calling publicly for legal recognition and financial compensation for the use of their intellectual property to train AI models (United States District Court & Southern District Of New York, 2023), and with governments and businesses around the world looking to harness AI business models for economic growth (Clifford, 2025). It is unclear how (or when) these issues will be settled, but it is clear that individuals within GLAMs need to be AI literate and aware of the issues involved in digitisation and publication of digital collections. Such awareness has made it possible for some GLAMs to take public positions on the use AI; the National Library of the Netherlands have taken technical precautions to restrict the use of their collections by commercial AI companies (National Library of the Netherlands, 2024), while the National Library of Scotland have taken an approach which they say will “prioritise both our people and collections” and committed to being transparent about how and when they use AI tools internally (National Library of Scotland, 2024).

Collections images have traditionally been a source of income for many GLAM institutions, so the push for openly available collections online has sat uneasily for many within their digital ecosystems (Wallace, 2022). This complexity has made use and reuse of digitised collections confusing for many researchers, and has added financial and legal barriers. Even organisations that offer free high resolution downloads of their collections, may also in the same place, or elsewhere on their website, insist on charging for non-commercial use of collections images. This is something many of those working within GLAMs are aware of, and working towards resolving. However, untangling long-standing institutional practices and assumptions, working cultures, budget targets and digital infrastructures is not something that can happen quickly for most institutions, much to the dismay of some researchers hoping for a quick change in GLAM image policies following the THJ vs Sheridan case in English law. The case concluded that faithful photographic reproductions of out-of-copyright objects do not accrue a new copyright (Secker, 2024).

It is clear from looking at the ways we digitise that there are human decisions at the heart of digitisation policies, that the type of collection or institution will have a huge impact on what gets digitised, and that complex external factors such as legislation and technological developments are only made more complex when translated into organisational structures and decisions.

### **(2.3) INSTITUTIONAL STRUCTURES**

One of the frustrations we frequently encounter from researchers through email enquiries, in person discussions and on social media is that GLAM organisations are slow to respond and adapt to changes. This pace of change, as seen with the UK Copyright case law above, is driven by many factors, not least the complexities apparent within large organisations. Staff sit within departments with sometimes competing targets or agendas, such as the need to balance public engagement activities with the equally important need for many organisations to generate commercial income. Even if teams are working towards the same organisational goals and strategies, the increased professionalisation of many GLAM roles such as collections management, resource description, interpretation and exhibition development, and public engagement mean that staff are often functionally separated by their departmental structures. In large and old buildings, or multi-site institutions such as the Bodleian Libraries, the staff are often physically separate as well.

This is acutely important to the process of making digital collections available, as teams and specialists such as curatorial, collections management or metadata specialists, conservation, imaging, copyright and legal, digital infrastructures and website or comms teams must work together in a long and complex series of workflows (McAulay, 2021, see also section 2.2 in *Towards Digital Collections*). Digitisation cannot happen in one department alone, and coordination between departments takes time. Elements of the process can be lengthy, such as complex or contested provenance research or copyright due diligence. This can seem achingly slow to researchers working on projects who are able to move relatively small numbers of staff at a rapid pace (Grosvenor, 2023).

Another key challenge of digitisation is the sheer costs involved and the never ending nature of the work. Even if an organisation accepts that it will not digitise all of its collections — something more likely for a library or archive due to the number of pages in their collections — the scale of the task can be demoralising. There is also a growing acceptance that in our organisations we are not undertaking a “once and done” task: imaging standards change over time, metadata records need updating or adapting to make them more interoperable, data will need to be migrated to the latest system, and once-acceptable descriptions will need to be reappraised in the light of social change.

In the light of a slow and never-ending task it is important for GLAM organisations to ask what is digitised and why (*Roadmap for Putting Collections Online – MNAHA Collections, n.d.*), and institutions are asking these questions. At the Bodleian Libraries, the *We Are Our History* research project aimed to look at both collections and audiences, including reviewing its digitisation pipeline (*We Are Our History, n.d.*). The Bodleian is certainly not alone in reevaluating the biases inherent in its collections, and seeking to rebalance funding and resources to make discoverable the breadth of its collections to a broader range of audiences. In the UK, the National Trust’s *Colonialism and Historic Slavery* report drew much negative press from its publication (*Interim Report on the Connections between Colonialism and Properties Now in the Care of the National Trust Including Links with Historic Slavery, 2020*), but the resultant furore has not had a negative effect on visitor numbers (*12% day tripper boost to National Trust sites reveals new visitor trend, 2024*). Increasingly, GLAM organisations are not just digitising areas of their collections, but engaging source communities in dialogue, co-curation or restitution conversations as part of this process, and as a way to reducing barriers between people and collections and object that hold resonance for them (MNAHA, n.d. a).

Most museums have less than a small percentage of their permanent collection on display at any one time – The MNAHA, for instance, can only show about 2,2% of their collection on 7400 m<sup>2</sup> (Kugeler, 2024a). This proportion would be even smaller for museums with much larger collections – they are often accused of ‘hiding’ collections, and it is an important function of digital collections to showcase more objects than can be physically on display at any one time, and to engage audiences and readers beyond those who can visit a particular site. The Enlightenment model of the curator as both caretaker, guardian and fount of knowledge has been evolving for some time (Hooper-Greenhill, 1992; MacGregor & Williams, 2005), and whilst curators who ferociously guard ‘their’ collections may still exist, the expected models and training for these roles has changed. Curators and other GLAM professionals (for it is not just a curator who cares for a collection anymore) are encouraged or mandated to work with and for their audiences, to provide services for the people they hold collections in trust for.

The scaled digitisation of GLAM collections has accelerated this process. Digitisation efforts require standards across organisations, which can be challenging for institutions with heterogeneous collections. But this process serves to eradicate departmental quirks of practice, just as international collaborative digital infrastructures such as Europeana serve to minimise unique local practices. A collection cannot be idiosyncratic if its owners want it to be interoperable.

Centering human involvement when looking at GLAM digitisation shows the value people bring to these processes. But of course, human involvement does not stop at publication. Usually the mission of GLAM institutions is to serve their varied audiences and publics, it is for them that digitisation projects are taken on.

To look at this other side of the medal, public reception and engagement with digitised documents, we look at Europeana as an example of a large-scale project aimed at a large-scale audience.

### (3) IMPLICATIONS

#### (3.1) NETWORKS OF PUBLICS

All cultural heritage institutions serve a variety of audiences and publics – some distinct and others overlapping (Cox & Tilton, 2019). In order to serve these networks, with their complementary (and sometimes conflicting) expectations and requirements, institutions (and the people within them) are obliged to consider a range of sensitivities and needs. In recent years, in many institutions across the world, this has taken the form of a process of decolonisation, which has resulted in a range of institutional responses. These range from including new narratives in exhibitions, to revising terminology in documentation with the help of indigenous source communities, reorganisation of collections, removal of certain objects from display and in some institutions, wholesale repatriation of objects to their communities of origin (Onciul, 2015; Iskander, 2016; Buchczyk 2023; Broekhoven, 2024; Peers, 2017). These processes are far from easy, often fraught and increasingly play out in the public eye, which adds another layer of scrutiny (and sometimes outright hostility) on and towards heritage professionals who have to meet the demands from their publics while also juggling internal tensions such as budgets and differing skill levels, as well as maintaining the professional standards of preservation and documentary authenticity. These tensions are not merely abstract. At the most extreme end of the spectrum was the case of the 2020 audit of existing, published, peer-reviewed academic research on National Trust properties, commissioned by the Trust and conducted by Professor Corinne Fowler at the University of Leicester with National Trust staff. The report found that many properties and the historical figures associated with them had ties to various kinds of colonial activities (National Trust, 2020). The report was co-opted into the culture wars and Fowler received threats and hate mail.<sup>2</sup>

These challenges are no different in the digital space, and if anything are increased several-fold (Candela, Pereda, et al., 2023). Most obviously, the audiences for digital collections and data are potentially greater, by many orders of magnitude, than those who would be able to visit an institution or use their holdings in person. A number of multi-site organisations have touted their website as their additional site, for example, the website was the ‘sixth palace’ when one author (Gooch) worked at Historic Royal Palaces (2010–2019). However, what we have often observed in our professional practices, is that the technological responses and solutions which have to be finessed to meet the demands of digital networks of publics are often more complex to execute, and sometimes result in less nuanced results than it might be possible to implement in the physical spaces of a GLAM collection.

A good illustration of this is the ways in which access to sensitive materials is controlled or restricted in digital collections, particularly in the aggregated collections we have previously discussed. For example, research conducted on access to images and records of human remains in Europeana showed that, while the overall volume of this material is a tiny fraction of the number of the objects in Europeana, it is still possible to access images and records of objects which would be unlikely to be displayed in museum galleries in the 2020s, such as *toi moko*, the preserved and mummified heads of Maori individuals, which were originally used as war trophies in traditional Maori societies (Kahn & Simon, 2023). Because aggregation of data in Europeana is a bulk action and depends on regular updates on the repository end of the equation (i.e., by the providing institution), it is still possible to access the records and images of these objects in Europeana, and then to travel backwards, via the links provided in the metadata to locate the objects and their records in the relevant collections online. It is not trivial to remove data from a database retrospectively, particularly if we are considering a few items in an aggregation of many hundreds of thousands. Wholesale removal is also undesirable, from an archival and preservation perspective (Conway, 2015). There is a world of difference between managing digital materials which raise ethical questions, and simply erasing the digital traces left by curatorial processes, decisions and activities. This case raises several questions, including asking who is responsible for maintaining the data that links into Europeana, whose task it should be to maintain quality checks on these kinds of materials and more broadly, how applicable the guidelines for ethical management of human remains in museums are to the digital iterations of these institutions.

---

<sup>2</sup> The case was reported widely in the British media, both at the time of publication and in subsequent articles, as well as discussed in the British parliament. For more detail see the following articles: “*I’ve been unfairly targeted, says academic at heart of National Trust ‘woke’ row* (Doward, 2020) and “*I’m not afraid of anybody now’: the woman who revealed links between National Trust houses and slavery – and was vilified* (Adams, 2024) and Hansard HC vol 696 cols 775–6.

However, we would like to provide an additional reading of this example. As Bettiva and Stainforth point out, the term “Europeana” is a rather simplistic shorthand for a complex relational network of people (Bettiva & Stainforth, 2017), technologies, and perhaps most significantly, technological representations of analogue things (including both data and objects), sometimes referred to as a social machine (Shadbolt et al., 2019). These networks defy simple definitions – a quick look at the list of institutions which provide materials to Europeana reveals sources in New Zealand, Australia, Canada and the USA, not to mention post-Brexit United Kingdom and Ukraine – two nations where the politics of being in or out of Europe are particularly fraught at the moment. To our reading, this is a good illustration of the arguments presented by several scholars, that being part of Europeana represents an ideological (and therefore entirely human-driven) alignment with Europe and a European identity (Capurro & Plets, 2020; Thylstrup, 2019). At the end of the day, it is human beings who make decisions at the policy and implementation levels about how and where data is to be shared, and while their imperatives may not always align, we cannot lose sight of this human element.

The Europeana network includes large national institutions with correspondingly impressive budgets and sophisticated digitisation programmes alongside collections from much smaller, regional and specialist institutions, who lack the funds, skills and digital platforms which Europeana can provide. Capurro and Severo (2023) have highlighted the role that Europeana plays as a social hub gathering a network of national and local actors in the European cultural environment. Using these perspectives as a starting point, we argue that, just as a museum provides different levels of access to their collection for different users (gallery-level texts for casual visitors, in-depth documentation for researchers, preservation data for storage managers and object restorers, technical data for digital asset managers) so too should digitisation be considered as a process that produces a variety of forms and data outputs, or, put another way: one catalogue cannot rule them all. This is a challenge for GLAMs with heterogeneous collections – libraries have integrated library systems, museums have collections management systems, archives have archive management systems. This often results in staff having to shoehorn archives into integrated library systems, or books into a collections management system.

This brings us back to the dichotomy and the tensions between mass and critical digitisation. In the era of big data, when pressure is on institutions to make as much data available as openly as possible, there is an increasing risk that the blunt mass approach becomes predominant. In the cultural heritage context, this could be extremely problematic, not only in terms of content but also the ways in which data is presented. Since the FAIR data principles were defined in 2016 (Wilkinson et al., 2016), they have become a de facto standard for how to present cultural heritage data in-line with e-Science principles, and are promoted as a way of making heritage data as findable, accessible, interoperable and re-usable as possible, by the machines that harvest and aggregate data across the web. But, from a curatorial perspective it may be preferable that not all data be accessible, even when technically, it is relatively easy to implement. It is the human element in this equation which should dictate whether items and their digital surrogates should be accessible, rather than an automated ingestion system (de Oliveira et al., 2021; Dutta et al., 2021). Digital surrogates of complex and sensitive materials have to be accommodated in systems which can make allowance for the layers of catalogue and human knowledge which exists among specialist staff, but are much harder to replicate mechanically.

Some mechanisms, such as the CARE principles of Indigenous Data Sovereignty are increasingly gaining traction as a guideline for making sensitive data available at the computational level, while protecting the rights and interests of marginalised groups (Carroll et al., 2020). These principles are broadly applicable across domains, but are beginning to be explored, conceptually and practically, in GLAM settings, most significantly at the Smithsonian, where digital asset management policies which identify sensitive materials, and outline best practice for restricting access to them, have been adopted (Carroll et al., 2020; Cocq, 2022; Fredricksson, 2022; Watson, 2021).<sup>3</sup> Similarly, the Datasheets for Cultural Heritage Datasets initiative (Alkemade et al., 2023) is gaining broader use as a mechanism for providing additional documentation for cultural heritage datasets in terms of detailing possible biases and providing provenance data. While these interventions have a great deal of promise, they are still far from being de facto

---

3 See also Cocq (2022) for a discussion of the implementation of CARE principles in metadata ontologies; Fredricksson (2022) for a study of their implementation in digital libraries and Watson, (2021) for a discussion of the principles in the context of cataloguing. See also Conisbee, 2025 (this volume).

## ACKNOWLEDGEMENTS

Thanks to all the GLAM staff who have worked on any aspect of any dataset or element of digitisation mentioned in this paper, and indeed, to those who we haven't mentioned. You are the digital collections infrastructure.

## FUNDING STATEMENT

Parts of this research were funded through the the European Union's Horizon 2020 Marie Skłodowska-Curie/REWIRE programme (Grant Agreement No 847639).

## COMPETING INTERESTS

The authors have no competing interests to declare.


## AUTHOR CONTRIBUTIONS


Author: Megan Gooch: Writing – Original Draft, Writing – Review & Editing


Rebecca Kahn: Writing – Original Draft, Writing – Review & Editing

Edurne Kugeler: Writing – Original Draft, Writing – Review & Editing

## AUTHOR AFFILIATIONS

**Megan Gooch**  [orcid.org/0000-0002-3190-0509](https://orcid.org/0000-0002-3190-0509)  
Bodleian Libraries, University of Oxford, Oxford, United Kingdom

**Rebecca Kahn**  [orcid.org/0000-0002-8837-5813](https://orcid.org/0000-0002-8837-5813)  
Department of History, University of Vienna, Vienna, Austria

**Edurne Kugeler**  [orcid.org/0009-0002-8670-9928](https://orcid.org/0009-0002-8670-9928)  
Digitization and Digital Collections Department, Musée national d'archéologie, d'histoire et d'art Luxembourg, Luxembourg

## REFERENCES

- Adams, T. (2024, April 21), 'I'm not afraid of anybody now': the woman who revealed links between National Trust Houses and slavery – and was vilified. *The Observer*. <https://www.theguardian.com/world/2024/apr/21/corinne-fowler-national-trust-colonialism-our-island-stories>
- Alkemade, H., Claeysens, S., Colavizza, G., Freire, N., Lehmann, J., Neudecker, C., Osti, G., & Strien, D. van. (2023). *Datasheets for Digital Cultural Heritage Datasets*, 9(17). <https://doi.org/10.5334/johd.124>
- Barbot, L., Moranville, Y., Fischer, F., Petitfils, C., Ďurčo, M., Illmayer, K., Parkoła, T., Wieder, P., & Karampatakis, S. (2019). *SSHOC D7.1 System Specification—SSH Open Marketplace*. <https://zenodo.org/records/3547649>
- Barrett, J. (2023, March 31). ARCHiOx, part 1: Finding stories in the margins. *The Bodleian Conveyor*. <https://blogs.bodleian.ox.ac.uk/theconveyor/archiox-part-1-finding-stories-in-the-margins/>.
- Bettvia, R. S., & Stainforth, E. (2017). All and Each: A Socio-Technical Review of the Europeana Project. *Digital Humanities Quarterly*, 11(3). <https://www.digitalhumanities.org/dhq/vol/11/3/000318/000318.html>
- Broekhoven, L. N. K. V. (2024). Unearthing Colonial Complicities in Maasai Collections at the Pitt Rivers Museum, Oxford. In *Pragmatic Imagination and the New Museum Anthropology* (pp. 135–160). Routledge. <https://doi.org/10.4324/9781003123354-9>
- Buchczyk, M. (2023). Diversifying the Collections at the Museum of European Cultures. In Macdonald, S (Ed.), *Doing Diversity in Museums and Heritage: A Berlin Ethnography* (pp. 193–210). <https://doi.org/10.14361/9783839464090>
- Candela, G., Gabriëls, N., Chambers, S., Pham, T.-A., Ames, S., Fitzgerald, N., Hofmann, K., Harbo, V., Potter, A., Ferriter, M., Manchester, E., Irollo, A., Keer, E. V., Mahey, M., Holownia, O., & Dobrev, M. (2023). A Checklist to Publish Collections as Data in GLAM Institutions. *Global Knowledge, Memory and Communication*. <https://doi.org/10.1108/GKMC-06-2023-0195>

- Candela, G., Pereda, J., Sáez, D., Escobar, P., Sánchez, A., Torres, A. V., Palacios, A. A., McDonough, K., & Murrieta-Flores, P. (2023). An Ontological Approach for Unlocking the Colonial Archive. *J. Comput. Cult. Herit.*, 16(4), 74:1–74:18. <https://doi.org/10.1145/3594727>
- Cannon, B. (2013). The Canadian Disease: The Ethics of Library, Archives, and Museum Convergence. *Journal of Information Ethics*, 22(2), 66–89. <https://doi.org/10.3172/JIE.22.2.66>
- Capurro, C., & Plets, G. (2020). Europeana, EDM, and the Europeanisation of Cultural Heritage Institutions. *Digital Culture & Society*, 6(2), 163–190. <https://doi.org/10.14361/dcs-2020-0209>
- Capurro, C., & Severo, M. (2023). Mapping European Digital Heritage Politics: An Empirical Study of Europeana as a Web-based Network. *Heritage & Society*, 1–21. <https://doi.org/10.1080/2159032X.2023.2266801>
- Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Rodrigo, S., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE Principles for Indigenous Data Governance | Data Science Journal. *Data Science Journal*, 19(43), 1–12. <https://doi.org/10.5334/dsj-2020-043>
- Chambers, S., Candela, G., Vodopivec, I., Cuper, M., Parkoła, T., Antonacopoulos, A., Atanassova, R., Kaukonen, M., Gabriels, N., Depuydt, K., de Does, J., & Romein, C. A. (2022). Sharing and Sustaining Digitisation Knowledge: an IMPACT White Paper [https://docs.google.com/document/d/1LJw4K-2l3EpBg\\_wWKYmd140aSv6Lsjo2390hDiHFJhQ/edit?tab=t.0#heading=h.eygc20lttc6m](https://docs.google.com/document/d/1LJw4K-2l3EpBg_wWKYmd140aSv6Lsjo2390hDiHFJhQ/edit?tab=t.0#heading=h.eygc20lttc6m) (last accessed 30 January 2025).
- Clariah—Common Lab Research Infrastructure for the Arts and Humanities. (n.d.). Retrieved from <https://www.clariah.nl/> (Last accessed: 2 November 2024).
- CLARIN ERIC. (n.d.). Retrieved from <https://www.clarin.eu/> (Last accessed: 2 November 2024).
- Clifford, M. (2025). *AI Opportunities Action Plan* [assets.publishing.service.gov.uk/media/67851771f0528401055d2329/ai\\_opportunities\\_action\\_plan.pdf](https://assets.publishing.service.gov.uk/media/67851771f0528401055d2329/ai_opportunities_action_plan.pdf) (2025) Retrieved 14 January 2025.
- Cocq, C. (2022). Revisiting the digital humanities through the lens of Indigenous studies—Or how to question the cultural blindness of our technologies and practices. *Journal of the Association for Information Science and Technology*, 73(2), 333–344. <https://doi.org/10.1002/asi.24564>
- Common Crawl—Open Repository of Web Crawl Data. (n.d.). Retrieved 7 February 2024, from <https://commoncrawl.org/>
- Conisbee, C. (2025). Research Data Management and Crowdsourcing Personal Histories. *Journal of Open Humanities Data*, 11(9). <https://doi.org/10.5334/johd.265>
- Conway, P. (2015). Digital transformations and the archival nature of surrogates. *Archival Science*, 15(1), 51–69. <https://doi.org/10.1007/s10502-014-9219-z>
- Cox, J., & Tilton, L. (2019). The digital public humanities: Giving new arguments and new ways to argue. *Review of Communication*, 19(2), 127–146. <https://doi.org/10.1080/15358593.2019.1598569>
- Dahlström, M., Hansson, J., & Kjellman, U. (2012). ‘As We May Digitize’—Institutions and Documents Reconfigured. *LIBER Quarterly: The Journal of the Association of European Research Libraries*, 21(3–4). <https://doi.org/10.18352/lq.8036>
- de Oliveira, T. M., Marques, F. P. J., Veloso Leão, A., de Albuquerque, A., Prado, J. L. A., Grohmann, R., Clinio, A., Cogo, D., & Guazina, L. S. (2021). Towards an Inclusive Agenda of Open Science for Communication Research: A Latin American approach. *Journal of Communication*, 71(5), 785–802. <https://doi.org/10.1093/joc/jqab025>
- Digital Bodleian. (n.d.). Retrieved from <https://digital.bodleian.ox.ac.uk/> (Last accessed: 4 November 2024).
- Doward, J. (2020, December 20). I’ve been unfairly targeted, says academic at heart of National Trust ‘woke’ row. *The Observer*. <https://www.theguardian.com/uk-news/2020/dec/20/ive-been-unfairly-targeted-says-academic-at-heart-of-national-trust-woke-row>
- Dutta, M., Ramasubramanian, S., Barrett, M., Elers, C., Sarwatay, D., Raghunath, P., Kaur, S., Dutta, D., Jayan, P., Rahman, M., Tallam, E., Roy, S., Falnikar, A., Johnson, G. M., Mandal, I., Dutta, U., Basnyat, I., Soriano, C., Pavarala, V., ... Zapata, D. (2021). Decolonizing Open Science: Southern Interventions. *Journal of Communication*, 71(5), 803–826. <https://doi.org/10.1093/joc/jqab027>
- Europeana. (n.d.). Retrieved 2 November 2024, from <https://www.europeana.eu/en>
- Explore Kasematten Niedergrünwald-Olizy in 3D. (n.d.). Matterport. Retrieved 2 November 2024, from <https://my.matterport.com/show/?m=KFRATurWNoP>
- Fredricksson, M. (2022). Balancing community rights and national interests in international protection of traditional knowledge: A study of India’s Traditional Knowledge Digital Library. *Third World Quarterly*, 43(2), 352–370. <https://doi.org/10.1080/01436597.2021.2019009>
- Gibson, L., & Kahn, R. (2016). Digital Museums in the 21st Century: Global Microphones or Universal Mufflers? *Museological Review*, 20, 39–51.
- Griffiths, A. (2010). Collections Online: The Experience of the British Museum. *Master Drawings*, 48(3), 356–367. <http://www.jstor.org/stable/25767237>

- Grosvenor, B. (2023, December 29). Court of Appeal ruling will prevent UK museums from charging reproduction fees—At last. *The Art Newspaper – International Art News and Events*. <https://www.theartnewspaper.com/2023/12/29/court-of-appeal-ruling-will-prevent-uk-museums-from-charging-reproduction-fees-at-last>
- Hooper-Greenhill, E. (1992). *Museums and the shaping of knowledge* (1st ed.). Routledge. <https://doi.org/10.4324/9780203415825>
- Internet Archive*. (n.d.). <https://Archive.Org/About/>
- Isaac, A., & Clayphan, R. (2013). *EDM\_Primer\_130714.pdf*. [https://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/EDM\\_Documentation/EDM\\_Primer\\_130714.pdf](https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf)
- Iskander, D. (2016). Collecting with Communities: Promoting Indigenous Voices in Museum Spaces. *Journal of Museum Ethnography*, 29, 113–130.
- Kahn, R., & Simon, R. (2023). Skulls, skin and names: The ethics of managing heritage collections data online. In C. Palladino & G. Bodard (Eds.), *Can't Touch This*. Ubiquity Press. <https://doi.org/10.5334/bcv>
- Kugeler, E. (2024a). Gems from the depot: Showcasing objects from our vast collection. *MuseoMag*, 3. <https://www.nationalmusee.lu/volumes/files/MUSEOMAG-2024-3.pdf>
- Kugeler, E. (2024b, December). Die Hüter Der Datenbank: Vom Schatten Zum Wort. *MuseoMag*, 4, 30–33.
- Logan, M. A., & Liew, C. L. (2023). GLAM Convergence Revisited: An Examination of User Perception and Experience. *Journal of Library Administration*, 63(8), 1014–1043. <https://doi.org/10.1080/01930826.2023.2281340>
- MacGregor, N., & Williams, J. (2005). The encyclopaedic museum: Enlightenment ideals, contemporary realities. *Public Archaeology*. <https://doi.org/10.1179/pua.2005.4.1.57>
- McAulay, E. (2021). Always Be Migrating. *The Code4 Lib Journal*, 50. <https://journal.code4lib.org/articles/15568>
- MNAHA. (n.d. a). *Looted Objects: The Albert Spring Collection*—s. Retrieved 4 November 2024, from <https://collections.mnaha.lu/stories/spring/>
- MNAHA. (n.d. b). *Homepage—MNAHA Collections*. Retrieved 2 November 2024, from <https://collections.mnaha.lu/index/>
- Museum Data Service*. (n.d.). Museum Data Service. Retrieved 2 November 2024, from <https://museumdata.uk/>
- National Library of Scotland. (2024). AI Statement <https://data.nls.uk/projects/ai-statement/> Retrieved 15 January 2025.
- National Library of the Netherlands. (2024). Statement on commercial generative AI <https://www.kb.nl/en/ai-statement> Retrieved 18 December 2024.
- National Trust. (2020). *Interim Report on the Connections between Colonialism and Properties now in the Care of the National Trust, Including Links with Historic Slavery*. Retrieved from <https://nt.global.ssl.fastly.net/binaries/content/assets/website/national/pdf/colonialism-and-historic-slavery-report.pdf> (Last accessed: 18 December 2024).
- Onciul, B. (2015). *Museums, Heritage and Indigenous Voice: Decolonizing Engagement*. Routledge. <https://doi.org/10.4324/9781315770246>
- Padfield, J., Bolland, C., Fitzgerald, N., McLaughlin, A., Robson, G., & Terras, M. (2022). *Practical applications of IIIF as a building block towards a digital National Collection*. Zenodo. <https://doi.org/10.5281/zenodo.6884885>
- Padilla, T., Allen, L., Frost, H., Potvin, S., Russey Roke, E., & Varner, S. (2019). *Final Report – Always Already Computational: Collections as Data*. DOI: <https://doi.org/10.5281/zenodo.3152935>
- Peers, L. (2017). The magic of bureaucracy: Repatriation as ceremony. *Museum Worlds*, 5(1). <https://ora.ox.ac.uk/objects/uuid:1310f980-d311-499d-86c0-fd10f1f2d3d5>
- Prescott, A., & Hughes, L. (2018). Why Do We Digitize? The Case for Slow Digitization. *Archive Journal*. Retrieved from <http://www.archivejournal.net/?p=7962> (Last accessed 15 January 2025).
- pui san lok, susan (2024). *Museum × Machine × Me*. Exhibition and research report. University of the Arts, London [https://www.arts.ac.uk/\\_data/assets/pdf\\_file/0020/453602/MxMxM-UALTate-2024.pdf](https://www.arts.ac.uk/_data/assets/pdf_file/0020/453602/MxMxM-UALTate-2024.pdf)
- Roadmap for putting collections online—MNAHA Collections*. (n.d.). Retrieved 4 November 2024, from <https://collections.mnaha.lu/roadmap/>
- Ross, C., & Terras, M. (2011). Scholarly Information-Seeking Behaviour in the British Museum Online Collection | [museumsandtheweb.com](http://museumsandtheweb.com). In J. Trant & D. Bearman (Eds.), *Museums and the Web 2011: Proceedings*. Archives & Museum Informatics. [https://www.museumsandtheweb.com/mw2011/papers/scholarly\\_information\\_seeking\\_behaviour\\_in\\_the](https://www.museumsandtheweb.com/mw2011/papers/scholarly_information_seeking_behaviour_in_the)
- Secker, J. (2024). THJ v Sheridan: A new era for copyright in UK cultural institutions? *Copyright Literacy*. <https://copyrightliteracy.org/2024/01/12/thj-v-sheridan-a-new-era-for-copyright-in-uk-cultural-institutions/>

- Shadbolt, N., O'Hara, K., De Roure, D., & Hall, W. (2019). *The Theory and Practice of Social Machines*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-10889-2>
- Szrajber, T. (2008). Public access to collection databases: The British Museum Collection Online (COL): A case study. In *Annual Conference of CIDOC*. Athens.
- Thylstrup, N. B. (2019). *The politics of mass digitization*. MIT Press. <https://doi.org/10.7551/mitpress/11404.001.0001>
- Towards a National Collection | Collections United*. (n.d.). Retrieved from <https://www.nationalcollection.org.uk/> (Last accessed: 18 October 2022).
- United States District Court & Southern District Of New York. (2023, December 27). *The New York Times Company, Plaintiff, vs Microsoft Corporation, OpenAI, Inc., OpenAI LP, OpenAI GP, LLC, OpenAI, LLC, OpenAI Opco LLC, OpenAI Global LLC, OAI Corporation, LLC, and OpenAI Holdings, LLC, Defendants*. [https://nytco-assets.nytimes.com/2023/12/NYT\\_Complaint\\_Dec2023.pdf](https://nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf)
- Wallace, A. (2022). *A culture of copyright: A scoping study on open access to digital cultural heritage collections in the UK* (Towards a National Collection) [Commissioned report]. AHRC. <https://doi.org/10.2139/ssrn.4323683>
- Watson, B. M. (2021). Advancing Equitable Cataloging. *NASKO*, 1–28. <https://doi.org/10.7152/nasko.v8i1.15887>
- We Are Our History*. (n.d.). Retrieved from <https://www.bodleian.ox.ac.uk/about/libraries/our-work/we-are-our-history> (Last accessed: 29 October 2024).
- Welcome to Towards Digital Collections: Towards Digital Collections – Resources for Galleries, Libraries, Archives & Museums. <https://canvas.instructure.com/courses/9719434/wiki> (Last accessed: 14 January 2025).
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), Article 1. <https://doi.org/10.1038/sdata.2016.18>
- Zeffiro, A. (2019). Digitizing labor in the Google Books Project: Gloved fingertips and severed hands. In *Humans at Work in the Digital Age*. Routledge. <https://doi.org/10.4324/9780429244599-10>

Gooch et al.  
*Journal of Open  
 Humanities Data*  
 DOI: 10.5334/johd.274

#### TO CITE THIS ARTICLE:

Gooch, M., Kahn, R., & Kugeler, E. (2025). Human Perspectives and Social Infrastructures: Prioritising People in GLAM Digitisation. *Journal of Open Humanities Data*, 11: 13, pp. 1–12. DOI: <https://doi.org/10.5334/johd.274>

**Submitted:** 04 November 2024

**Accepted:** 22 January 2025

**Published:** 19 February 2025

#### COPYRIGHT:

© 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

*Journal of Open Humanities Data* is a peer-reviewed open access journal published by Ubiquity Press.