Edmond, Jennifer, Nicola Horsley, Jörg Lehmann, and Mike Priddy. "Data incognita: How do data become hidden?." *The Trouble With Big Data: How Datafication Displaces Cultural Practices*. London,: Bloomsbury Academic, 2022. 89–104. *Bloomsbury Collections*. Web. 26 Apr. 2023. <<http://dx.doi.org/10.5040/9781350239654.ch-5>>.

Downloaded from Bloomsbury Collections, www.bloomsburycollections.com, 26 April 2023, 02:37 UTC.

Copyright © Jennifer Edmond, Nicola Horsley, Jörg Lehmann and Mike Priddy 2022. Released under a CC BY-NC-ND licence (<https://creativecommons.org/licenses/by-nc-nd/3.0/>). You may share this work for non-commercial purposes only, provided you give attribution to the copyright holder and the publisher, and provide a link to the Creative Commons licence.

Data incognita: How do data become hidden?

In the last chapter, we saw how general search engines such as Google return considerably more background noise for any specific search query than usable resultant web pages and how the long tail of the search algorithms hides research resources. This is a phenomenon not limited to cultural material, but where the existence of a class of professional trained and tasked with ensuring access to, and preservation of, these materials enables us to see clearly. We therefore now turn to another such manifestation of high (but potentially misplaced) societal trust in big data, that is the way in which analogue disconnected digital collections can sometimes become detached from their function as input into knowledge and identity formation processes. Cultural heritage institutions (CHIs) are moving to meet the new challenges of the digital world but there are yet more ways in which cultural heritage resources can be concealed from the researcher. To make their holdings digitally accessible and machine-readable CHIs must remove the aspects of hiddenness that reduce discoverability and accessibility for their users.

We use the term ‘hidden’ here not to imply active choices but to speak of the result: that data and cultural heritage resources are not visible to researchers who might otherwise use them. In asking why data are not used we are concerned with all factors that may lead to data becoming ‘hidden’ from the historical record. Such hiddenness will necessarily take many forms on a spectrum from inconsistent cataloguing practices, or a loss of institutional expertise, to the obvious forms of concealment when data has a privacy dimension, or being more obfuscated or ‘buried’ in a way that diminishes researchers’ chances of discovery. Cultural heritage practitioners are fully aware of many of these issues. These forms of ‘hiddenness’ exacerbate the discoverability challenges faced by researchers and particularly when search engines are the predominant means of discovery.

In this chapter, we will discuss the challenges caused by hidden data, as well as how data and research infrastructures can aid data discovery and reuse

for researchers, beyond the *de facto* Google method, lessons of wider use in addressing challenges of biases and misinformation.

Hidden by digital obscurity

Collections at the majority of CHIs (see Figure 3) are hidden from online discovery and an institution's complete holdings can only be revealed by other means.¹ There is apprehension that the utility of online catalogues and search engines could mask problems of discoverability where context is not clear:

I would argue that in some ways context is even more important [when researchers find holdings by drilling down directly from metadata] because the problem is I'm not necessarily confident that we have metadata records for everything we have. I'm certainly not confident that you can find everything we have using our systems and only a very small percentage of our collections in total are digital. And an even smaller percentage we can make available online for example and I'm concerned, as is, I think, the head of our curators, that that sort of skews the view of what we have.

(WP3 INT10)

One of the consequences of digital discoverability for users of this national library is a shrunk perception of available knowledge. It could be argued, however, that neither users nor practitioners have ever held a grand vision of the library's total holdings, because they are often built up over centuries in a number of amalgamated institutions. This is especially the case given the compounded obfuscation of incomplete metadata and inadequate systems, and thus these are not resources that have been lost *per se*. Hidden by obscurity is not necessarily an aspect of knowledge specifically being in, or missing from, the digital domain but is a challenge if search engines are the primary discovery mechanism.

In practitioners' experience the work left to do to achieve digital openness means that, for much of the knowledge they hold, creating the possibility of online discovery was the ambition:

I think the most important part of our job now is ... to create collection descriptions and to put the collection descriptions online so that they are see-able, that they are visible, because we talked about hidden data and hidden collections, and they deserve to be seen. I think for the next [few] years, that will be the focus point ... That's our future.

(WP3 INT3)

Hidden by working practices

Occasionally knowledge is hidden by working practices in the institution, where for convenience the practitioners tend to use the digitized holdings:

But we mostly use digitised, so some documents are less used, not because they are less important, but just because they are not digitised. That's a problem. But we cannot digitise everything, it will take money and time. But, I think there's consequences on the research and on the use of the documents ... Of course, we try to digitise the most important, or the most requested document, but that's not really possible for everything.

(WP3 INT2)

Archivists as well as researchers are thus becoming more familiar with digital holdings as they eschew what may sometimes be troublesome non-digital material. Thus, a Matthew effect² is established where only a limited subset of the institutional holdings is available online and consulted on a regular basis.

You can say well it doesn't really matter if people don't realise, we have much more, but it limits use of our collections and it also skews research towards what's easily available, properly catalogued, easy to find and ideally available freely online because that's what researchers will go to because it's just the most convenient.

(WP3 INT10)

The first threat to the survival of knowledge in the cataloguing and metadata creation process is that some data is hidden and likely to remain so because it had not been captured initially. For example, a potential problem for those concerned with accurately representing the life cycle of cultural heritage data is the 'almost entirely hidden' nature of aspects of preservation, with metadata creation described as 'something that few people care about' (WP3 INT6). It is thus inevitable that an overwhelming amount of data, that does not become formally recorded as metadata, is stored as embodied knowledge with individual cultural heritage practitioners. The data is then hidden because – although it is possible to make it machine-accessible and searchable – this is not considered when it comes to describing and therefore the groundwork is not laid.

Hidden by inconsistent methods of description

Methods and fashions in resource description creation in institutions vary over time. These changes are most obvious with the introduction of digital recording of the catalogue inventory. Moreover, institutions may circumvent the standards introduced by software cataloguing applications to complement their historic practices, as this librarian describes:

And I think that it's the diversity of the collections and operating at that large scale and having lots of stuff that's been catalogued to different standards in the 18th century or the 19th century, even I mean over the 20th century cataloguing standards have changed. So, there are quite a few challenges. It's both combinations of the systems not suitable or the system in the way how we are currently forced to use it not being suitable.

(WP3 INT10)

There was even variation amongst colleagues in how they worked:

I do a few hundred of those inventories and [this work is hindered] when the same colleagues come with the same faults, so that's hidden. I think it's a good example. One of the points in the inventory is who did what? And now they are very tolerated, so it was the archivist [B], twenty years ago, made the first list. Archivist [A] adapted it. It was colleague [C] who typed the inventory, read it and so on and so on. (WP3 INT7)

It was this kind of individual labour of activating embodied knowledge that was felt to be the most hidden aspect of archival and object description as its value was largely neglected. Attaining uniformity across individual practices within an institution was an ongoing process for some and regulating this standardization was said to be time-consuming:

Now we try to use the software to make what we have even more standardised than now. To make that they can't do their own thing anymore. So, then we have to be the bad guys but it has to be done.

(WP3 INT7)

Even when a recognized international metadata standard has been adopted for the majority of holdings, it is not necessarily suitable for other collections.

I mean the issue that we have is that we have very large scale and lots of different stuff, in particular I mean monographs, books, journals are all pretty nicely

captured in standard library systems. Standard library systems don't do so well on archival material and therefore you then end up having separate systems.

(WP3 INT10)

In such a situation metadata normally collected for an appropriate standard is not collected, or the standard is incorrectly utilized, and thus, this leads to a paucity of information about a research resource. Therefore, it could be concluded that discoverability is reduced and the resource is nominally hidden from a search that would normally reveal it to the researcher.

Hidden by a loss or unavailability of expertise

As discussed in Chapter 4, the exchange of the conventional logic of the archive in CHIs for the new logic of the search engine is a sea change for cultural heritage practice. Archivists' knowledge of the deep connections of hierarchical context must either endure, while no longer being reflected in the search-dominated representation of knowledge and researchers' methods, or it will become entirely redundant. It may be that the archive environment will become a hybrid of deep hierarchical structures represented in the flat linear space of search results. Will researchers be able to make the cognitive leap from the flat to the hierarchical when presented with the search results list or will they simply ignore the spacelessness in their quest for an answer to the query typed? If the context of the holdings becomes hidden when users drill down in this way, archivists' crafted narratives, which attempt to give items 'a shape, a pattern, a closure – to end their inevitable openness, close off their referents,'³ will lie dormant. This may result in a lack of understanding of potential uses for items or may result in incorrect assertions and connections to other resources.

'Google lookalike' (WP3 INT7) searches would marginalize sources that with the aid of specialist intervention would help researchers to discover in the institution's holdings. The separation of researcher from the source of collections knowledge held embedded within the curators may then diminish the quality of research conducted. Even if the balance does not swing all the way from a staff of 'historically correct' practitioners to one of results-orientated clinicians, it was considered that the growing presence of colleagues working from a position of computational thinking⁴ presented challenges for conventional archivists.

This manager of research services at a major library recognized the challenges presented by the introduction of technology and the likely ensuing skills shortage:

And when people expect to get the sort of digital full text type image, we'll probably see rapid changes for which the sector isn't always quite prepared in terms of skills. We're not overly agile in terms of our structures and we'll probably see a lot of shifts where people who believe that their job is secure forever will suddenly find that they'll be out of a job because we might not need as many people doing cataloguing in the future.

(WP3 INT10)

Hidden by a lack of material resources

A lack of material resources was a ubiquitous barrier to knowledge use that manifested in myriad ways but was seen as a way of life, underlying assumptions that the development of practice and systems would inevitably reach a limit. Funding models are not dynamic enough to offer institutions capacity to fully progress their ambitions, from digitization projects that had to be restricted to 'the most important, or the most requested document', as it was 'not really possible for everything' (WP3 INT2), to the arrested development of tools, which made it 'not possible to do anything and so it's not sufficient, it's not enough' (WP3 INT4). Under-funding could also nip innovation in the bud, as this practitioner had found

the problem is that we are totally understaffed. We don't have enough people to really start an ambitious policy and ambitious management of our metadata, that's the problem ... we participate in research projects about crowdsourcing, but the problem is that we don't have money here in the institutions to develop crowdsourcing tools ... sometimes we have a real lack of information in the caption of the pictures [held in the archive], and crowdsourcing would be a really good solution, but we really don't have money to develop a specific tool. So that's frustrating sometimes because we have the will, we have the ideas here, but we don't have the way to concretize what we would like to do, and that's really the biggest problem.

(WP3 INT4)

Opportunities to reap significant returns were therefore missed for want of what may be a relatively negligible investment. This could hold institutions back from

their goals of sharing their collections, as complex work such as conforming to standards could not be easily resourced.

Over-stretching was thought to be a particular problem for smaller archives without a critical mass of staff or the institutional infrastructure to support developments in practice.

The persons and the money to cooperate with something like [an infrastructure project]. It's possible for us, but for smaller archives it might be problematic because they don't have the persons or even don't have a database where they describe their holdings. They might have, I don't know, a book.

(WP3 INT1)

Smaller institutions were therefore at risk of becoming marginalized as they drifted away from the orbit of the standards and technological developments used by better-resourced institutions.

Practitioners had been frustrated to find that adopting technical solutions could amount to exchanging one inflexible way of working for another if these tools could not be developed over time. In this case, this was part of a general lack of ongoing investment to embrace digital tools:

We don't have a structural budget to manage digital collections here, to manage digital environments, and so we must ask. We will receive funds for four years, for example, and so we cannot really have a long-term vision and long-term policy without a structural funding of the digitisation.

(WP3 INT4)

Insufficient investment was a significant threat to the institutions' ability to plan for, and respond strategically to, the digital era. In the absence of long-term security, many took up externally funded opportunities, such as research and data infrastructures, to share resources across the sector with enthusiasm.

Hidden by privacy

It was widely accepted that CHIs must work within the framework of laws protecting individuals, which must include taking an active role in identifying risks to individual privacy:

The only reservation I have is about the privacy issue because that can be a problem if some data in our database, for example, linked with other data

in other databases, make the opportunity to recognise someone who wasn't recognisable at the beginning in our own database. So ... we must be careful about that and think about that.

(WP3 INT4)

Knowledge sharing is often said to be in tension with the protection of data and this was perhaps most evident in relation to legal and ethical constraints on practice. One participant described being 'bound by certain regulations' as 'the most limiting element of all' (WP3 INT3) when asked how far her role as an archivist could promote the use of her institution's collections. Privacy legislation may have an impact on the hiddenness of certain materials as even the archival descriptive metadata may risk revealing personal data:

We're currently discussing if we can put only the collection description fully available online ... Unfortunately, every collection description has a biography of the person who donated it, and [there is national] privacy legislation that prevents you from publishing certain information. If the donor's still alive, in some cases that can be difficult. We are currently discussing if we can just publish it and see what happens. Maybe nothing happens. Usually, donors are very honoured to be acknowledged as the donor. We don't expect much trouble, but we have to take into account that there might be here or there someone who objects. It hasn't happened so far. We'll see.

(WP3 INT3)

Questions about privacy legislation and the publishing of metadata for discoverability therefore arise as the intentions of regulations and the consequences of compliance could diverge with shifts in how institutions interact with their users.

The dark side of discoverability

You don't know what kinds of tools are going to be out there that will allow you to draw conclusions that won't have been recognised previously, and it does put people off making data available, even anonymised data, because people are often making this available openly.

(WP3 INT5)

Tools, such as those used in big data knowledge extraction techniques, may make incorrect assertions even if these are not driven by the intent of the researcher or the CHI. If this concern is widely held amongst practitioners, limiting access may become the *de facto* standard practice for any contemporary materials and data. This would be customary privacy practice for data about the living, which is normally guaranteed in law,⁵ but may also need to extend to the parents and grandparents of living persons. Examples are bystanders⁶ and those accused, but not found guilty, of crimes, and their close descendants who may not know of the accusations. Moreover, how individuals are categorized when involved in such events, and how their actions are evaluated and interpreted, changes over time, adding yet another layer of complexity to cultural heritage resources.

Digital discoverability of certain archival materials may increase the vulnerability of holdings to ‘bad use’, as one practitioner put it, such as Holocaust denial but the answer was thought to be found in greater sharing: with the exposure of ‘vivid’ narratives the key to undermining the twisting of historical sources to undesirable political ends (WP3 INT2). The risks of the digital environment and the legal duty for CHIs that hold contemporary history

What are the three greatest challenges that prevent your institution from sharing more information?
Responses (79)

Figure 4 The greatest challenges that prevent survey respondents’ institutions from sharing more information.

materials should not be underestimated, especially as they present additional staff-learning and financial pressures for the institutions at a time when they are already under pressure to do more with less. As can be seen in Figure 4, legal, privacy, ethical and intellectual property rights are practitioners’ biggest concerns in sharing more information, second only to the need for more staffing. Adhering to existing legal and ethical principles cannot guarantee outcomes in an unknown landscape of data linking and sharing, which unsurprisingly may lead to data being held back and hidden out of fear of its potential uses.

Strengthening the institution’s digital infrastructure and staff capabilities is seen as essential in assisting researchers who may not discover what they are looking for in the online catalogue (or finding aid), as well as safely making use of more sensitive materials.

It should be both, partly in simply improving our structure and I mean the more stuff you make available in digital form, either through download or APIs or other approaches, will allow more computational research in our collections. ... So that has a sort of personal component but we need to develop both the infrastructure and also our ability to support that type of research.

(WP3 INT10)

Reservations expressed about the big data era brought to light another crucial risk limiting openness in sharing access to resources. Data might be required to

Figure 5 The extent to which survey respondents felt engaged in a public duty to share data.

be hidden as a consequence of certain types of sharing, linking of resources and the potential for automated assertions which may be erroneous or disclosive.

Most practitioners identify a public duty to share data, with which they feel engaged (see Figure 5). Aggregation of information from different CHIs was thought to be relevant to institutions' particular operation and goals and there is a high level of involvement with, and interest in, research infrastructure and aggregation projects.

Discovery through cultural heritage institutional involvement in (European) data and research infrastructures

Data and research infrastructures⁷ are also addressing the challenges of resource discovery for their designated research communities. Discoverability and interoperability between datasets were the primary stimulators for the establishment of many research infrastructures in the humanities and social sciences. However, they have many more functions in supporting communities of research data users and knowledge-holding institutions. Many of the research infrastructures aggregate heterogeneous data with the aim of consistent standardization and making it available, not only via their own indexing algorithms, but also potentially making it available to other aggregators. It is the unique understanding of the designated communities served that brings additional value to the research infrastructure and resource discovery through, for example, the addition of (multilingual) controlled vocabularies for advanced search techniques. Furthermore, they provide services and development support to metadata/data sharing CHIs in the form of tools for metadata improvements to meet community standards. For many of the smaller or less well known of the CHIs the inclusion of information about them in the infrastructure portals facilitates improved international exposure and enquiries by researchers, an advantage this archivist felt was inherent in being part of a larger research infrastructure that respects the nature of cultural heritage knowledge:

Because, we have to expose the documents to encourage local studies. If people can't see us, we can't exist, so it's very important. And, the [X] portal also corresponds to the view of the [institution]. It's a scientific portal, it's very human behind the portal ... So, it's the view of the [institution] ... because, we cannot be only for the local researcher ... And, since we are on the portal, we have a lot of foreign researchers, in fact. So, it's very important for us.

(WP3 INT2)

As an external force, research infrastructures acting upon institutional practice were said to be a positive agent for change in institutions:

In 2010, we had a system where there was one [Microsoft] Access database with item descriptions. Every collection that came in was split into items, and then you had an item description. That was it, nothing else was done with it. Then [infrastructure project] came in and we started making collection descriptions. All collection description items received a unique and fixed identifier. We didn't have that before. We went through this process with [infrastructure project], and in 2014, we had a new work method. We learnt from working with [infrastructure project] what the gain is from working with external partners.

(WP3 INT3)

Shaking up institutional practice from the outside could therefore achieve significant change in a relatively short time. Moreover, the benefits of sharing via a research infrastructure were also seen as reflecting the changing nature of cultural heritage into an increasingly networked domain:

So our holdings are known better because we have some information that completes the holdings of the other institutions.

(WP3 INT1)

Practitioners' enthusiasm about archival holdings fuelled their commitment to sharing their 'hidden treasures', which they saw as precious but also 'common knowledge', in the sense that such knowledge should be a commons (WP3 INT2). Thus, their involvement in and commitment to research infrastructures could be seen to be beyond the mere financial. This fundamental ideology of sharing knowledge was at the heart of practitioners' desire to participate in infrastructure projects, for example, as:

You can't stay in your own cocoon to do your own things ... you have to give your data to as much people as possible ... You can reach new people. There is a possibility for new research. Also, for the institution it has a second benefit. You have to be visible. You can increase your visibility.

(WP3 INT7)

There was give and take in this perception, with a generous spirit and European camaraderie driving openness and an expectation of a return on this investment in expanding an institution's audience in a way that it could not achieve alone. This was buoyed by observed successes of infrastructure projects:

No Archive can work like Google. We don't have the manpower or the finances of Google. But to present data in a platform like [X research infrastructure] has on one hand a chance that people are asking for your holdings. On the other side it's a big, big danger that they are only looking for that information and don't realise that we might have more.

(WP3 INT1)

Encouraging the use of a CHI's holdings could be a benefit of research infrastructures, but an institution holding back from sharing fully, whether by failing to digitally publish its finding aid or not sharing all of its metadata via the research infrastructure, may experience a Matthew effect in terms of what holdings get used.

Research and data infrastructure projects are seen as a good way to both 'see the importance of standards and norms' and 'have a larger view about our field and other scientific fields' (WP3 INT2). Despite the observed benefits of involvement in research infrastructures, however, there are challenges to overcome for CHIs looking to become involved, as this practitioner observed when considering the cultural and linguistic differences between the communities of practice involved:

The first time that I understood what [project partners] were saying, because I could understand the words and the meaning of the words, it was really a victory. And I really thought, okay, wow, there's a completely different world out there, and if we do what these people say, we could actually share and get collections from other institutes digitally and put their descriptions into our system directly without me having to put in all the metadata myself. That was like an epiphany.

(WP3 INT3)

Collaborating with research infrastructures and other metadata aggregators offers the benefit of an 'outside opinion' (WP3 INT6) and was seen by practitioners as useful in terms of positioning their institutions within the sector. Additionally, being 'forced to evolve' provided 'technical advantages' (WP3 INT7), such as the adoption of metadata standards, which allowed them to conform to widespread practices and not be left 'behind' (WP3 INT4). Adapting to sharing on this level could transform practice and practitioners expressed openness to changing their practice when they were confident of the benefits of sharing. There is no doubt, however, that this is challenging for some who are less capable of adjusting their practice.

The future should not be hidden

Hiddenness is part of the story of the acquisition, preservation, cataloguing, description and dissemination of cultural heritage artefacts and is still a multifaceted problem even with the support of research infrastructures. The additional data generated throughout the process is subject to choices made at the time by the practitioners and the degree to which hiddenness can occur is subject to many influences. The recognition and inscription of values and meanings in cultural heritage are in never-ending dialogue and practitioners make important contributions to this process, working to ensure that the materials in their care can continue to be used to create new knowledge. The use and non-use of cultural heritage resources necessarily re-inscribe the meanings given to them and the value they appear to have for future use. Certain materials will be more prominent or more hidden at any given instance in their consultation, depending upon the route of access. Every such instance, whether the motivation is the faithful construction of metadata, a seemingly passive fact-finding exercise, or analytical interpretation for research purposes, is an activation and a re-contextualization.⁸

Measures required for eliminating hiddenness, and thus improving discoverability, cannot be equal for all data, materials, practice, institutions and therefore researchers. Sometimes the technologies adopted are not fit for purpose, as evidenced by practitioners unable to find data using their institution's own search engines. Generic search engines such as Google cannot know or infer what is hidden or why this might be so. Even with the superabundance of resultant links it could be still assumed that there is a Matthew effect present each time a researcher creates a query, especially with the search engine's willingness to provide the question via autocomplete. An objective of research infrastructures covering cultural heritage is to improve discoverability and reuse of data and research materials, sometimes with a role to support both researchers and CHIs.

As identified by practitioners, presenting data in a research infrastructure gives an institution increased exposure, as more people ask for specific items, not realizing that there are other items available. This is perhaps caused by the institution not fully digitally publishing all records of their holdings but infrastructures must understand that cultural heritage data and metadata are messy, being the result of human activity, rather than that of machine or sensor output. Metadata aggregating infrastructures cannot just insist on perfect records that meet their exacting standards: they must aid CHIs with mapping,

transforming and correcting their metadata, ideally in an automated or semi-automated fashion.

Without greater engagement with external infrastructures and other actors, not only will data become hidden but so will the institution: practitioners identified a lack of material resources that prevents the organization from standing out in the crowded field of spaceless search results where all appear equal. This does not mean institutions must indulge in the wholesale digitization of all objects in their catalogue or perish but they must ensure their catalogue is based on the current international standards for the collection type and is machine-actionable.⁹ Complex knowledge is at risk of being marginalized, not least because of valid and pressing concerns of inadequate resources across CHIs. Reliance upon funding from research projects with their own agenda may influence the choices of materials being made digitally available, while holding other material back risks the possibility of skewing the public perception and researcher expectations of the institution's holdings. Recognizing all of the institution's existing hiddenness challenges resulting from both active and passive practices, and having a strategy towards discoverability for holdings, is essential to maintaining relevance in the digital environment.

The options for digital discoverability of cultural heritage resources appear to be constantly in flux with new institutional, national, transnational, domain-specific and pan-domain data and research infrastructures. The wealth of choices for researchers may appear bewildering; which to choose and why is unclear. Should the researcher go to a general cultural heritage aggregator such as Europeana,¹⁰ a discipline portal,¹¹ an institution's website search engine, or Google for any given question? As mentioned in the last chapter, there is little risk to any individual choice, beyond a loss of time, but access to too much information in too many directions may scramble the researcher's knowledge-gathering strategies. As the research discovery services available mature and start to integrate additional benefits, which act as nucleators for the coalescence of communities, the choice of discovery services will become clear to a research domain. The development of research and metadata aggregating infrastructures may ensure a representation of archival knowledge endures and has a dynamic relationship with researchers' methods. However, for the moment it is not surprising that online resource discovery is mediated predominantly by Google despite its failings for cultural research.

Living, breathing cultural practices require the space to re-interpret artefacts of the past through the lens of contemporary practice. Furthermore, the reuse, remixing and sharing of experimental data require a space that facilitates

scholarly discourse and possible collaboration. These spaces cannot be the ‘fake news’ and conspiracy theory-filled environments of public social media platforms, nor the open public web, as the risk of discovery, misunderstanding, misinterpretation and misuse without context may be too great. This becomes critical when data have an impact upon individuals, groups or communities. Where emerging cultural practices of remix and recombination become very entangled with technologies that follow the software culture ethos and reject context, the potential for out-of-context misuse may have consequences well beyond the scholarly domain, as we will discover in Chapter 6.

Notes

- 1 Either by enquiring with the institution, or via a publish and printed holdings catalogue/finding aid, or a printed research guide.
- 2 Merton, Robert K., ‘The Matthew Effect in Science’, *Science*, 159/3810 (1968), 56–63.
- 3 Duff, Wendy M. and Verne Harris, ‘Stories and Names: Archival Description as Narrating Records and Constructing Meanings’, *Archival Science*, 2/3 (2002), 276.
- 4 Williamson, Ben, ‘Political Computational Thinking: Policy Networks, Digital Governance and “Learning to Code”’, *Critical Policy Studies*, 10/1 (2016), 39–58.
- 5 For example, in Europe the General Data Privacy Regulation has been enacted in EU states: Intersoft Consulting, ‘General Data Protection Regulation (GDPR) – Official Legal Text’, *General Data Protection Regulation (GDPR)* (2018), <https://gdpr-info.eu/>.
- 6 Ehrenreich, Robert M. and Tim Cole, ‘The Perpetrator-Bystander-Victim Constellation: Rethinking Genocidal Relationships’, *Human Organization*, 64/3 (2005), 213–24.
- 7 These can be local, national or international collaborations of organizations which together provide access to valuable research resources for designated research communities.
- 8 Ketelaar, Eric, ‘Tacit Narratives: The Meanings of Archives’, *Archival Science*, 1 (2001).
- 9 It should be possible for an algorithm to process the metadata without failure, facilitating advanced discovery and analysis.
- 10 ‘Europeana’, <https://pro.europeana.eu>.
- 11 For example, ‘Advanced Research Infrastructure for Archaeological Dataset Networking in Europe’, <https://ariadne-infrastructure.eu/>; ‘Collaborative European Archive Infrastructure’, <http://www.cendari.eu/>; ‘European Holocaust Research Infrastructure’, <https://www.ehri-project.eu>.