Deliverable Title: Deliverable D3.1 Report on Historical Data as Sources

Deliverable Date: 30/03/2018
Version: 1.0

Project Acronym:	K-PLEX
Project Title:	Knowledge Complexity
Funding Scheme:	H2020-ICT-2016-1
Grant Agreement number:	732340
Project Coordinator:	Dr. Jennifer Edmond (edmondj@tcd.ie)
Project Management Contact:	Michelle Doran (doranm1@tcd.ie)
Project Start Date:	01 January 2017
Project End Date:	31 March 2018
WP No.:	WP3
WP Leader:	KNAW
Authors and Contributors (Name and email address):	Nicola Horsley Mike Priddy (mike.priddy@dans.knaw.nl)
Dissemination Level:	PU
Nature of Deliverable:	R = report

Abstract:	What do cultural heritage institutions and their practitioners do and how is this changing in the big data era? Through a series of interviews with cultural heritage practitioners and an online survey, this report presents the investigations into the handling of knowledge complexity, changes in archival practice and data use, how data is shared, and also hidden, as part of the historical record. Furthermore, the barriers to the application of big data computational method to the historical record are considered, conclusions and recommendation are drawn.

Deliverable D3.1 - Report on Historical Data as Sources

Table of Contents

Table of Figures

Work Package Objectives

The objectives of this work package were:

- To evaluate the issues and challenges surrounding the aggregation of historical data as knowledge, and in particular for those institutions which are not active participants of large national or international aggregations.
- To further define a model of cultural heritage holdings as data (digital and otherwise) and investigate cultural and ethical barriers to big data approaches to historical and cultural sources, through interaction with cultural heritage institutions.
- To synthesise and communicate the findings as a white paper for policy/general audiences and a journal article.

This deliverable presents the results and findings of the investigation (through, literature review, interviews, and an online survey) of the issues, challenges, and barriers towards the possible application big data approaches in analysis of the historical record. Furthermore, the work presented here lays the groundwork towards a white paper.

Introductory Literature Review

In order to navigate an information environment experiencing a 'data deluge', we seek ways to reduce noise and enhance signal, most obviously through the use of metadata. Clearly this practice involves judgements of value to determine what is worthy of the mantle of 'signal' and what is labelled 'noise'. Archival science navigates the blurred contours of this landscape, which has always been shaped by cultural and temporal perceptions and the affordances of technology. The technologies that become part of standard practice in an archive then favour the creation of certain narratives over others. If data complexity is suppressed or left unaccounted for by those technologies, it will occupy a blind spot within the archive but if its description is too bound up with its complexity, its diverse potential uses will not be discovered. Either extreme represents a dilution of the richness of knowledge creation.

In discussing big data in relation to archives, K-PLEX is interested in approaches that support the potential for data to be re-used and re-analysed in conjunction with other data that may have been collected by unrelated researchers. Such research is facilitated through the use of descriptive metadata, appropriate preservation systems, informed institutional practice, and architecture for sharing across institutions to enable discovery by diffuse researchers. Mirroring wider society, academic research is currently in thrall to big data. Funding calls offer large grants to researchers who can apply the least datafied (Schafer and van Es, 2017) research interests to data-rich areas, 'consigning research questions for which it is difficult to generate big data to a funding desert and a marginal position within and outside the academy' (Kitchin, 2014a). Researchers taking on this challenge must then grapple with the socially constructed nature of datasets containing knowledge complexity that must nevertheless exemplify the gold standard of a five-star (re-)usability rating, a hallmark of epistemic authority that can only be achieved by containing some of that complexity in a black box (Latour, 1999). Such flattening of nuance is described as the defining characteristic of data engineering, which leads to what McPherson (2012) calls a lenticular view of knowledge.

To understand what such a turn really means for archival practice, it has been argued (Bolin and Schwarz, 2015; Kitchin, 2014b that we must clarify whether big data is genuinely being adopted as a heuristic by academic, governmental and associated actors, or if the 'myth' of 'Big Data' (boyd and Crawford, 2012) is merely a useful discourse for those whose interests are served by the promulgation of an evangelical 'dataism' (van Dijck, 2014). This phenomenon has parallels across society. For example, Williamson (2016) analyses how the Hour of Code and Year of Code initiatives saw 'a computational style of thinking' infiltrate schools in the US and UK, which he describes as a style of thinking that 'apprehends the world as a set of computable phenomena'. Williamson draws attention to a deficit of reflexivity amongst advocates of computational approaches to social problems, which obfuscates the 'worldviews, ideologies and assumptions' of the creators of systems for processing data, black-boxing the processes that delimit data use. Berry (2011) draws on Fuller (2010) in pointing out that the potential for new technologies to produce and reproduce inequalities in society is not simply a matter of a 'digital divide' but is significantly influenced by the commercial roots and market values of much of this *techno-solutionist* innovation.

Archivists are uniquely placed within this discourse, with everyday practices and systems for managing collections, and the confluence of traditions of working with cultural heritage holdings and adaptation to emerging technologies, all in their purview. As such, cultural heritage practitioners are more than a vital link in the chain through which historical data are maintained and transmitted. Engaging with practitioners' perspectives is fundamental to understanding the drivers behind data use and non-use and viewing the knowledge landscape from their position of archival thinking offers insight into how the computational turn is experienced in practice and how this may render new forms of research engagement with the historical record.

A shift towards big data approaches necessarily poses questions of how the contemporary landscape is characterised and what the custodianship of cultural heritage looks like at present moment of the computational turn, on the cusp of big data's installation as the dominant discourse across research disciplines. Perhaps because of the nature of the academic research lifecycle, academic publications have not kept up with the challenges K-PLEX is exploring. The most illuminating literature has come from major European digital research infrastructure projects federating cultural heritage data for use by researchers. These projects (including CENDARI[1], EHRI[2], DARIAH-EU[3], DASISH[4], PARTHENOS[5], ARIADNE[6] and HaS[7]) have all faced and to some degree addressed challenges associated with the sharing of cultural heritage data. The EHRI and CENDARI projects exposed threats to the sustainability of sharing knowledge from previously 'hidden archives', including a lack of consistency in the information that could be viewed across institutions and projects that, in one example, resulted in thousands of digitised sources that could not be displayed on the institution's website being hidden from potential users as well as from the institution itself

[1] Collaborative European Digital Archival Research Infrastructure http://www.cendari.eu
[2] European Holocaust Research Infrastructure https://www.ehri-project.eu
[3] Digital Research Infrastructure for the Arts and Humanities https://www.dariah.eu
[4] Data Service Infrastructure for the Social Sciences and Humanities http://dasish.eu
[5] Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies http://www.parthenos-project.eu
[6] http://www.ariadne-infrastructure.eu
[7] Humanities at Scale http://has.dariah.eu

(Vanden Daelen et al., 2015). The CENDARI and EHRI projects also found evidence of a need for archival workflows to be more transparent and reproducible as engaging with a research infrastructure currently involved ‘talking with at least three people: the one who could tell you “what” (was described and relevant for the portal), the one who could tell you “how” (software and standards or mapping) and the one who could tell you “yes” (the authority to give permission to integrate data from this archive into the portal) (Vanden Daelen et al., 2015: 8). With core elements of practice obscured, there was a danger that data could end up hidden between the cracks of the institution. This supports Star’s (2007) observations of the hiddenness of much of the ‘work, practice, and membership’ of socio-technical networks.

It has been suggested that, rather than seeking to maintain or arrive at a finished model, practitioners’ ideas of completeness may be more akin to ‘equilibrium in flow’ (von Bertalanffy, 1949) In which case, the historical record should be seen as a process, not a product. There are many lines of enquiry about the compatibility of archival thinking and practice with the computational turn and scholarly literature has only recently begun to turn its gaze in their direction. This early analysis has tended to privilege the most sensational hypotheses. For instance, Kitchin’s (2014a) reporting of the humanities’ parallels with an increasing marginalisation of deductive approaches in scientific fields is intriguing. While it would seem unlikely that humanists and social scientists would reject deductive methods in favour of purely inductive methods, the extent to which data-driven approaches are supported by archivists may be revealing. The practitioner view of the opportunities and challenges for broader use of data that big data approaches offer has been conspicuously absent from a discourse that largely represents them as passive actors, resistant to change Duderstadt et al. (2002; Edwards et. al, 2013).

Ribes and Jackson’s (2013) investigation of the workings of the data archive describes how ‘the work of sustaining massive repositories reveals only a thin slice in the long chain of coordinated action that stretches back directly to a multitude of local sites and operations through which data in their "raw" form get mined, minted, and produced. What remain at repositories are the distilled products of these field sites; behind these centers lie an even more occluded set of activities that produce those data themselves’. Extant research has not fully documented the extent to which existing metadata and practices across the sector already represent a big data approach to historical and cultural sources.

The myth of big data hinges on the occlusion of human intervention, which is the basis of claims that big data approaches render invisible or ‘remove’ ‘human bias’ (Kitchin, 2014a). Of course, bias is central to historical inquiry, and researchers’ power to recognise and expose it is key to their epistemic authority, so we might ask: if bias is hidden is a historical approach neutered? Rather than simply being a profession that is hostile to novel forms of knowledge creation (and perhaps there are some myths around preservation and conservation at work here), archivists may well have some considered reservations about the computational turn that have yet to be documented. When acting at the site of convergence of data practices as diverse as ethnography, with its concern for making researchers’ positionality explicit, and big data, which tests the boundaries of linking data collected for different purposes, surely some tension is to be expected. To date, however, there has been limited exploration of ways in which knowledge complexity might impact upon archival thinking and practice.

Indeed, previous research has suggested that archivists are constantly changing and adapting their practices and systems (Vanden Daelen et al., 2015; Borgman, 2015) and this will continue through and beyond the era of big data. Of course, there is a risk inherent in making any change to the way in which the historical record is passed on, that breaks in the chain may cause data to become hidden. Crucially, however, new practices must allow batons to be passed to future systems that might be better able to accommodate that data, as obsolescence is inevitable. This is not a new problem. Over seventy years ago, Broadfield (1946: 65-66) described how classification systems cannot last forever and called for declines in technology to be properly managed to preserve knowledge, arguing that: '[all] classifications in their existing forms are destined to become dust; sensitive adjustment should enable the classifier to consign them to dust himself [sic], instead of allowing the common enemy Time to do so'. Archival practice therefore never stands still, though it may change course, and an appreciation of knowledge complexity in archives can further understanding of why some paths are taken while others are left unexplored.

Weinberger's (2007) exploration of the 'new digital disorder', now ten years old, claimed that 'the real problem is that any map of knowledge assumes that knowledge has a geography, that it has a top-down view, that it has a shape'. Acknowledging this uncertainty, which was taken up by Kouw, Scharnhorst and van den Heuvel (2013), demands that, rather than attempting to simply trace the A-to-Z of an idealised research data life cycle, the black boxes that characterise knowledge creation processes must be engaged with. Theories that might usefully be applied to archival practices to help to explain how metadata and actions become obscured include Karup and Block's (2011) concept of quasi-actants. Karup and Block's work supplements Latour's (1999) vision of black boxes as they describe how these actors erase their traces so their work is not visible even when the black box is opened – in other words their *mediation* does not become *metadata*. This level of analysis could be illuminating if applied to research questions relating to practitioners' concerns about big data. For instance, previous research has not explored whether big data approaches might be seen as a fundamental shift or re-purposing of the archive, with practitioners feeling less like they are being nudged at the micro level and more like they are being 'enrolled' (Latour, 1999) by discourses of data science or commercialisation at the macro level. Latour and Callon described how translating the terms of a problem from the language of one discipline to another achieved *intéressement* when the translation is maintained and reinforced in order to complete the transfer of power from one set of actors to another (Star, 2007). Concepts such as Lave and Wenger's (1991) legitimate membership of communities of practice might be applied to describe experiences of translation. With some observers suggesting that academia is in the grip of an intellectual land-grab (Hess and Ostrom, 2003), knowledge practitioners may fear becoming a 'McArchive'.

Simply observing that many data scientists enjoy finding patterns in numbers and many historians are motivated by a passion for telling the stories of people who have suffered in being reduced to numbers does little to progress debate or practice in either field. It is nevertheless instructive to contrast a commitment to learning from extraordinary past events as a typical feature of an archive's mission statement to McPherson's (2012) analysis of the lenticular view of UNIX-style structures for coding, in which complexity is managed and controlled through the 'principles of information hiding' and the creation of discrete modules devoid of relation and context. McPherson highlights the benefits of such a modular

approach for coding, while warning that it also represents ‘a worldview in which a troublesome part might be discarded without disrupting the whole’. This approach threatens to engineer apophenia – ‘seeing patterns where none actually exist, simply because enormous quantities of data can offer connections that radiate in all directions’ (boyd and Crawford, 2012) – creating an information environment in which (potentially erroneous) macro-level patterns govern our view of knowledge creation. As well as reservations about offering up data to abuse or putting a balanced understanding of the past at risk, practitioners may have fundamental ethical fears about data linking that stem from their professional knowledge of the potential use of their collections that are yet to be researched. If practitioners are not convinced that privacy and research integrity can be maintained when datasets are linked or have concerns about ownership, control or access, research must now turn to rigorously interrogate these issues to ensure cultural heritage knowledge is used to the advantage, and for the advancement, of the world’s knowledge.

Methodology

The primary methods employed for this investigation were one-to-one interviews with Cultural Heritage Institution (CHI) practitioners’ and an online survey to both inform the interview methodology and to provide broader information about the community of cultural heritage practitioners.

The aims of the interviews and survey were as follows:

- To further understanding of (CHI) practitioners’ perspectives on data use and non-use in the context of a shift towards big data approaches.
- To explore the potential impact of CHIs’ policy and practice, especially where they are not active participants of large national or international aggregations, on the future of the historical record.
- To further define a model of cultural heritage holdings as data (digital and otherwise) and investigate cultural and ethical barriers to big data approaches to historical and cultural sources, through interaction with cultural heritage institutions.

The first aim listed above was added to the initially agreed aims of the work package in order to bring out the study’s use of the term ‘hidden’ to encapsulate any data that is at risk of not being used by researchers. ‘Hidden’ was found to be a somewhat problematic term because of its inherent *active* connotations that preclude passive neglect but the study continued to use the term where appropriate.

In asking why data are not used we are concerned with all factors that may lead to data becoming ‘hidden’ from the historical record. Our use of ‘hidden’ is not to imply any active choices but speaks of the result: that data are not visible to researchers who might otherwise use them. Such ‘hiddenness’ will necessarily take many forms on a spectrum from being less conspicuously validated, for example by a missed opportunity for duplication in a specialist archive, to being more obfuscated or ‘buried’ in a way that diminishes researchers’

chances of discovery. K-PLEX[8] therefore seeks to apply theories that might help to explain how metadata and actions become obscured.

A working document of more refined research questions was then drafted. These questions were then assessed on how appropriate they were to be address by the survey or in interviews, based on the depth of answer anticipated and whether they represented an issue to be resolved in the survey, in order to inform interviews, or whether they spoke to a theme that may need to be explored more flexibly. Questions for the survey or interview schedule were then crafted from these research questions (see annex?), while retaining links between the two for reflection at the coding stage to facilitate analysis based on initial aims as well as themes arising from the data. The survey questions and interview schedule were further refined to ensure they were in productive communication with each other and likely to garner complementary data.

Survey

The survey was then designed by hand and later structured and formatted using Survey Monkey. There was a total of 54 questions and the length of the survey represented a risk to the recruitment and retention of respondents. The design of the survey was therefore refined to redress the balance between free text questions and those for which a response could be captured in a matter of seconds. The ordering of questions was also changed to reflect this balance in the perception of respondents and create the feeling that completing the survey would not require an undue amount of time and effort.

The survey was then piloted at KNAW-DANS. The researcher observed data archive practitioners completing the survey, timing the process, before asking for their feedback on the clarity of the questions. Feedback was gathered on question phrasing, layout, the appropriateness of certain formats, and their general reception as participants in the research. This feedback informed editing of the survey and the addition of a notice that it may take up to 30 minutes to complete.

The survey then went live on Survey Monkey. At its inception, and periodically throughout the four months it was live, the link to the survey, as well as further information for participants, was publicised through three main channels: relevant JISC email lists, Twitter and known contacts, including interviewees. There was also some snowballing via interviewees. Responses were reviewed and reflected on periodically using Survey Monkey's analytics tools and downloading interim data to ensure aims were being met. The survey achieved 77 complete responses from 18 different countries before it was closed as, although just under the 100-response target, it was felt that saturation had been achieved in the clustering of respondents' answers. Given the length of the survey, it was felt it achieved an impressive richness of response and the commitment of respondents to completing questions suggested that practitioners were grateful of the opportunity to have their voices heard.

[8] Knowledge Complexity https://kplex-project.eu

Interviews

Ten interviews were conducted with cultural heritage practitioners in archival institutions. The sample of interviewees was derived from cultural heritage practitioners known to have varying levels of involvement in data sharing infrastructures. Ten practitioners were contacted, all of whom agreed to take part in the research. Eight of these worked in conventional archives – including a university archive, a national archive, an international archive, one that was part of a Holocaust memory institution and one linked to a museum – one was a national library and one a digital repository within a university library. These practitioners were chosen to represent a range of practice. Some institutions were born out of the need to preserve records that were routinely generated. At some, the bulk of their collections were made up of material donated by the public, whereas others had considerable budgets for buying new items. Most importantly, interviewees were chosen from across Europe, representing practice in six different countries.

Half the interviews were conducted face-to-face and half via Skype video calls. There was no difference in the level of rapport the researcher felt was achieved between face-to-face and Skype interviews. Interviews lasted an average of one hour and were recorded using a digital audio recorder. The audio files were transcribed verbatim by a professional transcription service based in the UK.

Analysis

Preliminary descriptive analysis of survey responses using Survey Monkey's online tools fed into the refinement of interview questions. Survey data was later downloaded as a Microsoft Excel spreadsheet, cleaned to remove incomplete responses, including two respondents who did not identify as practitioners, and analysed using Apple Numbers. This analysis was found to complement the findings from the analysis of interview data.

Transcripts were coded inferentially using NVivo, informed by reflection on the themes for exploration from the research questions. The audio files were consulted where there were ambiguities or errors in the transcripts and the researcher's role in both conducting the interviews and analysing the data ensured the validity of the data in this report. After further refinement of coding, the importance of each of the themes arising from the data for addressing the study's research questions was addressed and this analysis was then written up for dissemination, using quotes anonymised with the convention of assigning numbers to interviewees.

Survey data used to validate and guide interview themes

The survey captured a range of respondents in terms of job roles, experience in the cultural heritage sector, and type of institution.

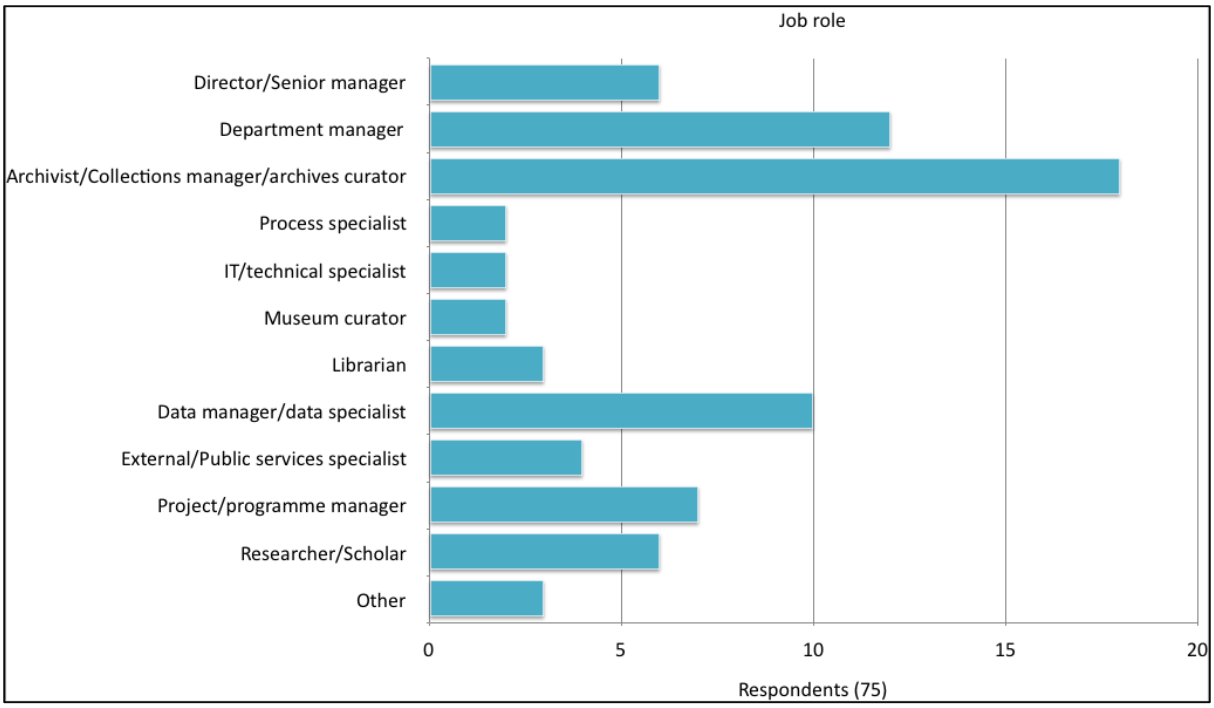

Figure 1: Job roles of survey respondents

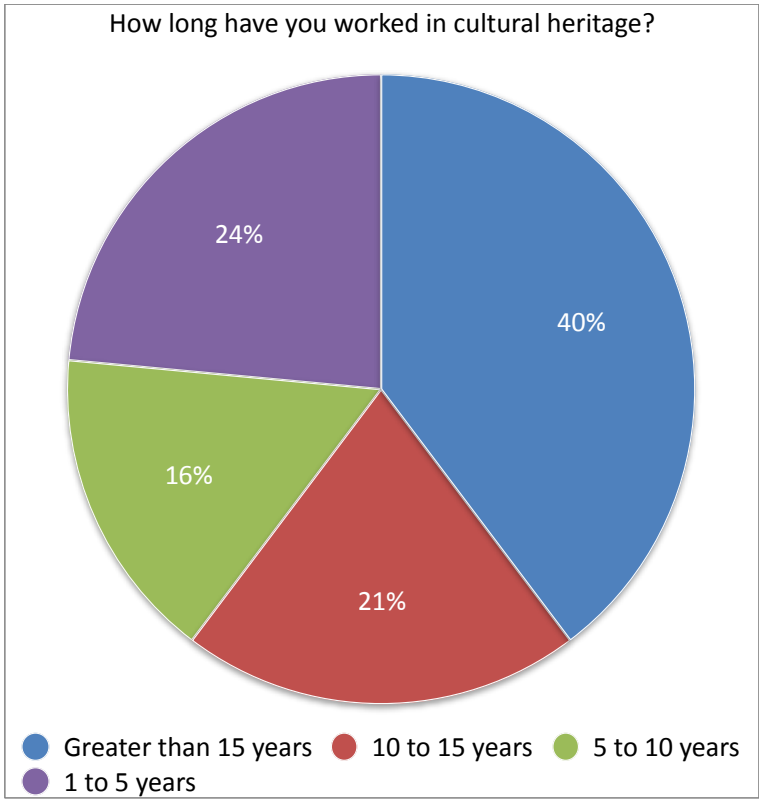

Figure 2: Experience of survey respondents

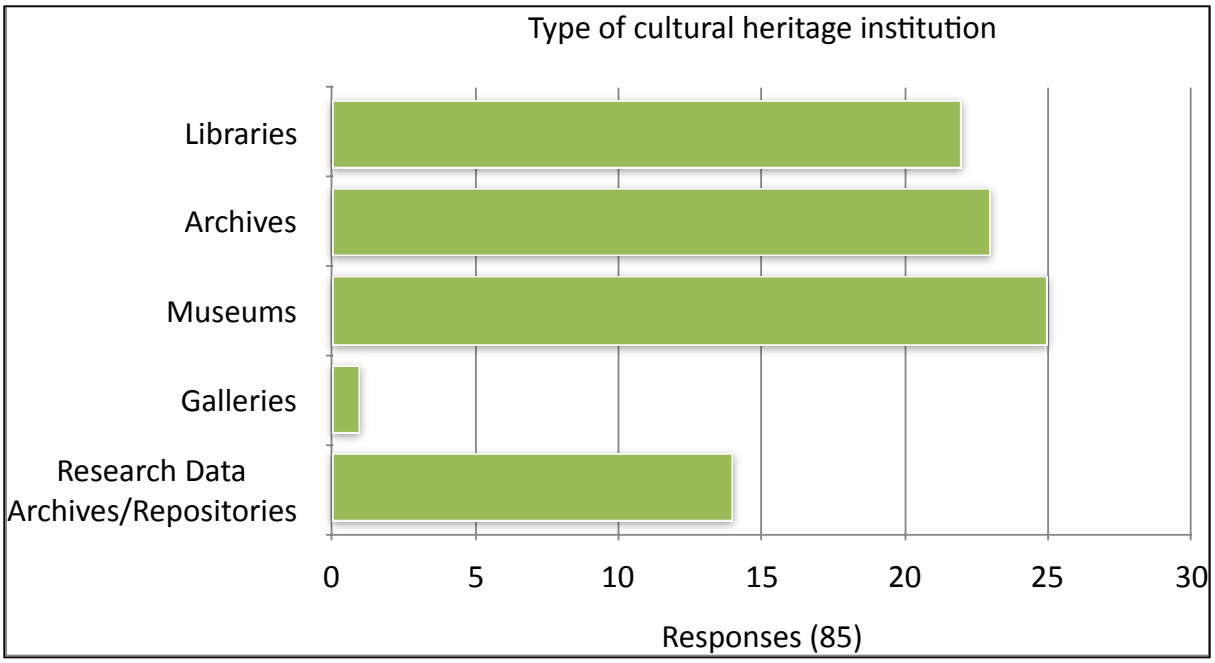

Figure 3: Types of cultural heritage institution of survey respondents

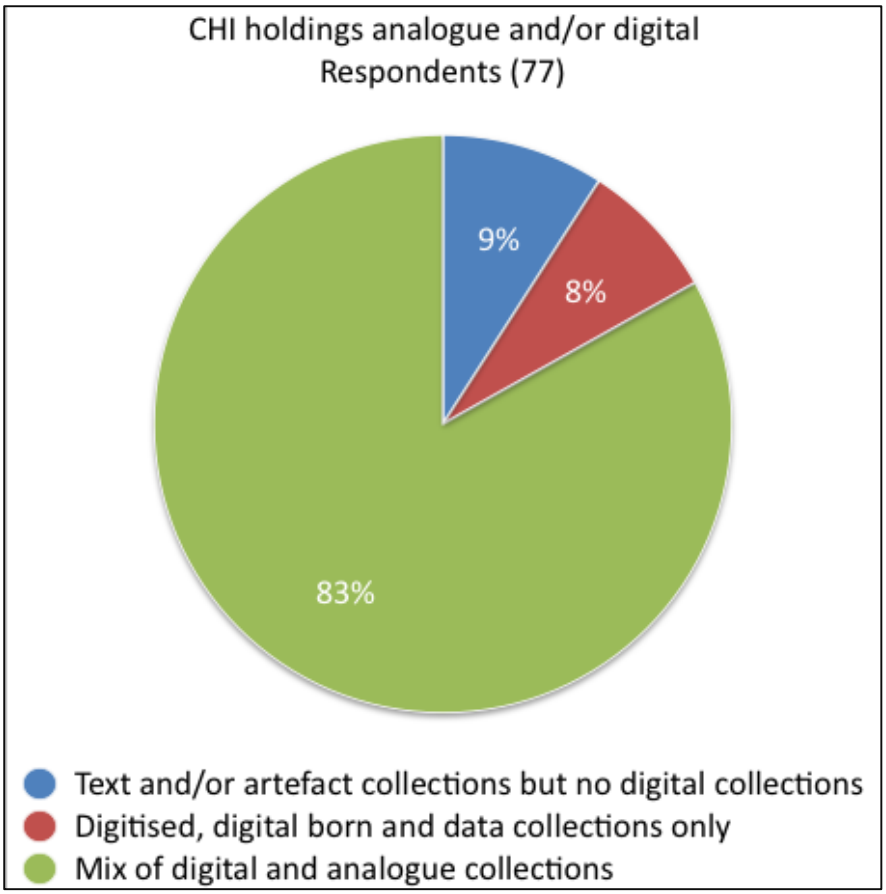

Figure 4: Types of holdings worked with by survey respondents

	Academic researchers	Student researchers	Businesses	Genealogists	School children	Other members of the public
highest % of users	14	4	0	4	0	5
significant % of users	14	10	3	11	8	19
large % of users	18	21	0	7	2	4
small % of users	8	14	11	8	10	15
very small % of users	3	7	23	16	24	12
0% of users	0	0	15	5	9	0

Figure 5: Survey respondents' descriptions of their institutions' user community

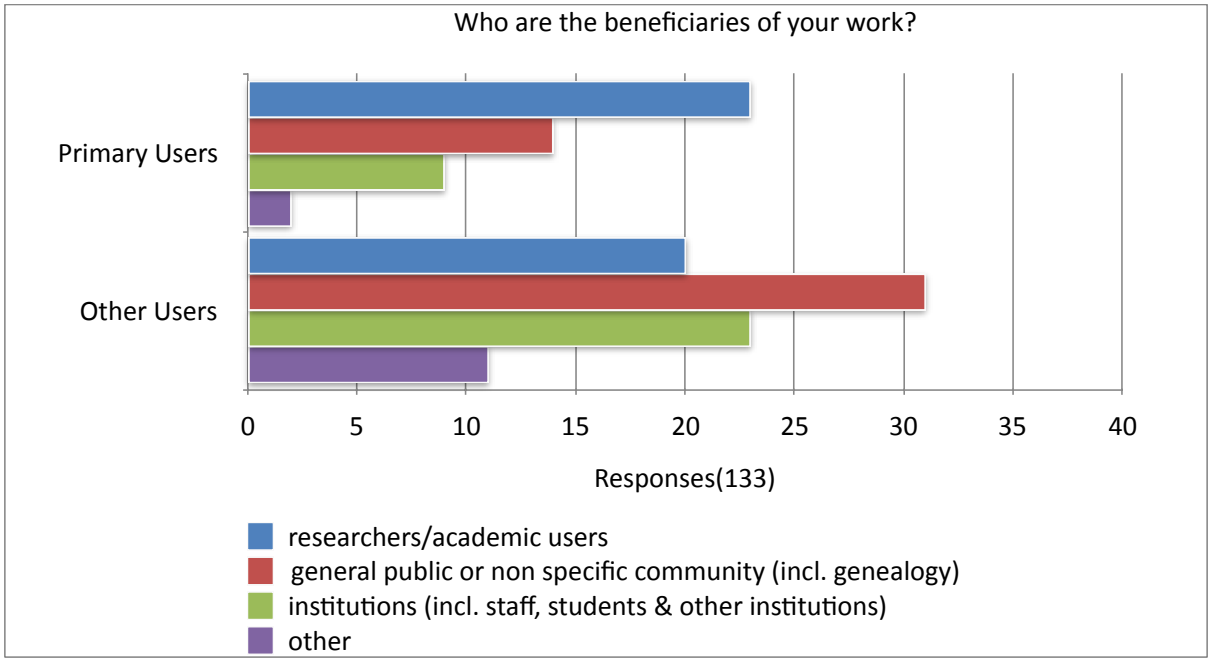

Figure 6: The beneficiaries of the work of survey respondents

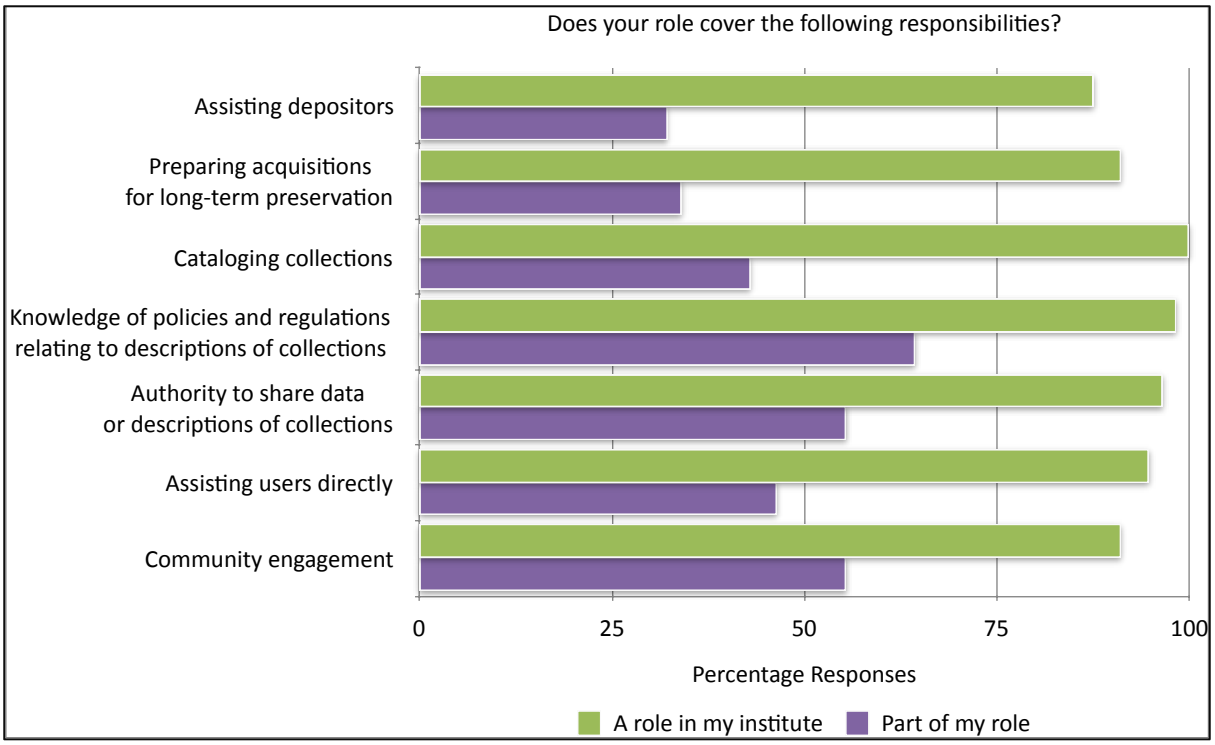

Figure 7: Responsibilities of survey respondents and their institutions

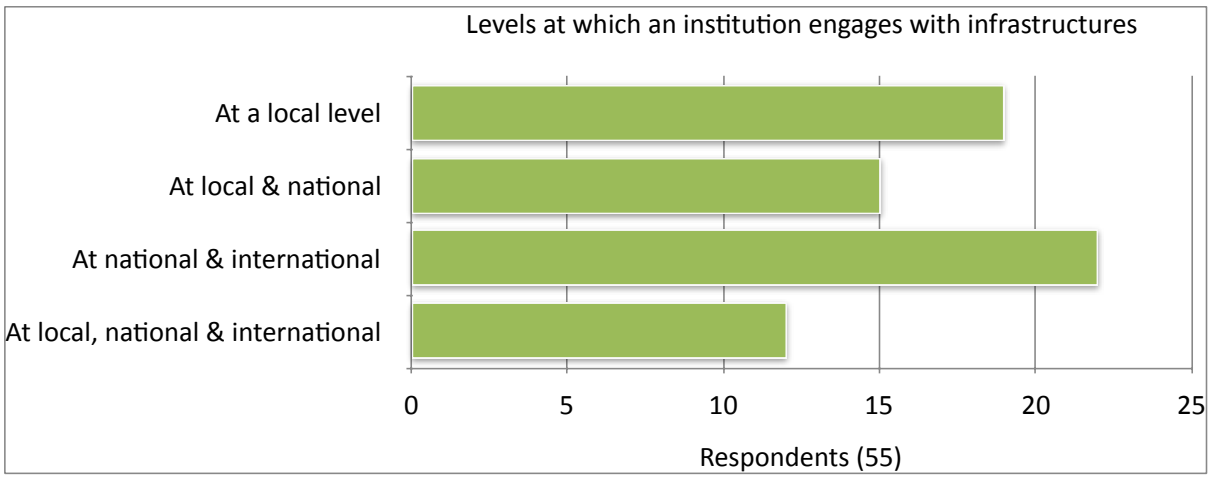

Figure 8: Engagement with data sharing infrastructures identified by survey respondents

This demographic data provided validation of the themes from the survey to be explored in greater depth through interviews. The major themes were:

Most influential changes to practice:

- Digitisation of analogue resources/material
- Acquisition of digital born resources
- Computerisation of cataloguing

Impact on collections:

- Visibility, accessibility and searchability of collections
- Preservation and growing storage

- More and changing work to catalogue, preserve & curate. Change of scale
- New workflows of knowledge production

Further responses gleaned from the open-ended questions in the survey also helped to inform the interviews and this report, were:

Elements of respondents' roles they felt beneficiaries would not be aware of:

- Most would not know that my role exists
- Preservation
- Working with volunteers
- Crowdsourcing
- Governance, housekeeping & maintenance
- The amount of data worked with
- Information given from sources originally not in our archives
- Collections management as a whole is largely invisible to most users
- Efforts made to make the archives available and searchable
- Work and money required to digitise collections

Does your institution provide information [metadata] about your collections to an external portal/aggregator? If so, why?

- Wider dissemination of information about our collections, tend to be collection specific
- We do this to a limited extent, but mainly the metadata is not clean enough
- Very specialised collection that links to an international site focused on this type of collection
- Increase accessibility
- To allow access to dispersed collections
- To enable a broad research on one topic in the collections of more than one institution
- To let people know it's there
- External partner's interest
- Improving access
- Data sharing, keeping up with the Joneses
- Visibility; satisfying the requirement of open access policy
- Being member of (inter)national consortia/infrastructures
- Metadata are not useful without some information.
- It is necessary to make known the history of the holocaust to the greatest number

Have you seen any changes in how collections are used since joining the aggregation project?

- People want more digital objects and not just metadata
- The amount of inquiries rises up.
- More frequent use; more visibility; more hits
- Requests have become more specialised, focused

- More users
- We have more foreign researchers

Were there any unforeseen challenges in participating in the aggregation project?
metadata not "good enough" for hackathon

- Technical and communications (possibly legal too)
- Yes, data mapping changes
- Problem of long-term sustainability
- Transmitting knowledge to others scientists and the public
- We had to make amendments to metadata and go through a validation
- There is a need for consultation and expertise
- Trouble in generating valid EAD
- Cleaning of data is more time consuming and difficult than expected

What are the three greatest challenges that prevent your institution from sharing more information?

- Proprietorial notions of the content
- Unwillingness from traditional curators to follow standards when cataloguing
- Staffing
- Limited resources - largely volunteer run
- Quality metadata
- Infrastructure is not yet set up
- Resources
- Legal barriers
- Technical
- Money for technical resources
- Legal obstacles
- Lack of cataloguing
- Fear of illegal copying
- Personal data security
- Costs: to add more descriptions we have to hire people to put it in the catalogue
- International differences in privacy regulations
- Perceived security
- Lack of staff on ground to do retro-conversion of catalogue data
- Copyright
- ICT network (insufficient wifi, lack of storage etc.)
- Presenting information in a digital format
- Institutional knowledge of collections
- Time; monetary concerns
- "Dirty" metadata which cannot be aggregated
- Technical Capabilities and current data format
- Dependency on collection income

- Data quality
- Messy data
- Hardware resources
- Extent of collections
- Interpretation of data
- Communication
- Poor choice of software for digital preservation; tends also to be very expensive for smaller institutions
- GDPR
- Too much of the collections are still only catalogued on analogue system which isn't easily searched
- Lack of trained human resources
- Need to refresh staff skills in a rapidly developing area - lack of good training courses
- Ensuring that both institutions have similar aims and methodologies

Introduction

Many sociological studies of big data in social science and humanities research environments have taken social media as their muse (an understandable response to a new wellspring of data self-identifying as *social*) (see Ruppert et al, 2015; van Dijk, 2014; Manovich, 2014, 2011). Less attention has been given to existing data that may be reconstituted by the big data era and re-contextualised by the novel uses and methods of discovering and analysing such material afforded by this new era.

Archival cultural heritage institutions are engines of knowledge (Robertson, 2017) that reflect 'realities as perceived by the 'archivers'' (Ketelaar, 2001). That is not to say that practitioners working in these institutions exercise complete control over the knowledge on which their users may draw. The big data era provides a case study of how archival processes may be changed by influences external to the institution. Understanding these influences is essential to an appreciation of knowledge complexity in the research process because, as Derrida (1996) explains, 'the mutation in technology changes not simply the archiving process, but what is archivable – that is, the content of what has to be archived is changed by the technology'. Technology has already made a considerable impact on cultural heritage institutions, as this study's survey of practitioners shows (see Figure 4).

The practice of working across different forms of cultural heritage sources that is the norm for this study's survey respondents will have implications for the representation of knowledge complexity. Practitioners' representation of the historical record requires them to be ''neutral' intermediaries between users and information', applying their 'theories, methods, models, and descriptions [which] are as presumptuous and controlling as scientists' construction and containment of nature' (Olson, 2001). According to Bowker and Star (1999), using these value-laden tools of their trade requires practitioners to make ethical judgements as '[e]ach standard and each category valorises some point of view and silences another. This is not inherently a bad thing - indeed it is inescapable. But it is an ethical choice, and as such it is dangerous - not bad, but dangerous'. Such danger is ongoing, as the performance of cultural

heritage (Crouch, 2010) is a never-ending process with values and goals but no pretensions of arriving at an end product in which truth is settled and knowledge finalised.

Conceptualising Data

K-PLEX's use of *data* is elucidated in the report of Work Package 2 and for the purpose of this study it can be understood to encompass all sources of knowledge held by cultural heritage institutions that may be used by researchers. This definition reflects institutional practice, guided by this research and encapsulated by an interviewee's assertion that: 'I think we don't really define the document in itself, we define more of the information in the document. It's not really the bearer that concerns us, it's the information that we try to collect' (I3).

What do Cultural Heritage Practitioners do and how is this Changing in the Big Data Era?

Creating Cultural Heritage

Handling Knowledge Complexity – Acquisition

Each cultural heritage institution is a jigsaw piece in that they are shaped by their own identity – they are what they keep and they keep what they are (Cook, 2011) – and together they render the enduring cultural knowledge upon which future researchers may draw. Each institution has its own particular remit and mission, which is reflected in its approach to acquisition.

Acquisition criteria were thematic and/or place-based, in that the archive served intellectual inquiry into a particular country, region or the specific institution itself (two collected resources of the universities of which they were part, and one had a special interest in the Nazi deportation camp whose site it occupies). Some had a public remit and others looked after private collections.

Two of our participants who were responsible for the preservation of data at universities talked about determining the significance of material. At both institutions, data considered of use for future research were retained in-house, but a distinction was made by one to encapsulate pure "cultural heritages" that should be specially preserved off-site. In both instances, selection was governed both by high-level policies and continually-developing local level practices that took a more nuanced approach to defining and assessing the value of research objects. The position of one participant's role as research data manager at the university's library meant acquisitions were informed by demand:

> We're introducing a new policy that we'll also be accepting systematic review search strategies as well, because there was a demand for that from our academics. Despite it being called a data repository, there's a wide range of resources that don't necessarily fit within that description. (I5)

This participant's view of what could be defined as research data was therefore flexible and guided by potential uses by researchers who were known to him. Having oversight of all documentation produced by her university placed another participant's task of identifying potentially significant resources alongside her legal responsibility to preserve some materials and destroy others:

> So, we go to all these 50 locations [within the university] and, with the owners of the information, we select the data which should be destroyed right away, should be destroyed after a certain period, or which is cultural heritage. And the material we say, okay, that's cultural heritage, we bring that to a central repository … We have a very global inventory on that material. And then, if we have all the cultural heritages together in one repository, we will make more detailed inventories, better descriptions, regulations on accessibility. And if that's finished, it will be brought to the City Archives. (I9)

On the surface, this three-way categorisation seems straightforward. This participant explained, however, that although compliance with the law was her starting point, there was a hidden complexity in the work of ascribing materials to be worthy of preservation:

> Well, we work with bulk, with everything. And there is legislation on which documents an organisation such as the X creates but also has to destroy and also has to keep as cultural heritage. So, this legislation helps us. And our main drive is to be compliant. …
> It should [be black and white what qualifies as cultural heritage and what does not] but the legislation is not always very clear because it was created a couple of years ago, and there is a grey area. And there is also a lot of new legislation coming up, but that's mainly on being transparent and destroying information. So that's not from the cultural heritage perspective, but it's important for my work because we have to give dates when information should be destroyed. Which is, obviously, not cultural heritage but it's part of my job to advise on that as well. (I9)

In this participant's account, her wider role and the legal framework that governs part of it are important. Her task of preserving cultural heritage is put into perspective by her responsibility to destroy other kinds of data. Moreover, she suggests that the legal framework does not keep pace with practice, resulting in a marginal category of data that has the potential to be used for cultural heritage research. This grey area brings archivists' discretion to the fore and daily debates about cultural significance are to be expected, as in another participant's example:

> … sometimes people think that these kinds of photos are not interesting, but they are because they can tell us something about the way people were living, which kind of houses they had, which kind of furniture they had, which kind of clothes they wore at this period (I4)

Of course, cultural heritage practitioners are adept at identifying the *value* of material, but they must work within structures that are steered by other concerns. Even in comparatively

well-resourced institutions, decisions have to be made about priorities, as this participant from a national library explained:

> … we have a still reasonably generous but constantly shrinking budget and … my team actually sort of sets the strategy for how that budget is used … And we have a content strategy that informs what we buy and now we're currently in the process of refreshing that. So far in the past we've bought everything we could on the chance someone might want to use it and now we're being a bit more focused and specific and also in relation to usage data … so we're looking at usage data and then we might decide this isn't used enough based on how much we pay and then we cancel it or try and renegotiate, whatever the approach is. (I10)

Practitioners who have previously a enjoyed certain freedom to exercise their expertise in selecting and nurturing cultural assets are therefore somewhat straitened. In prioritising particular kinds of knowledge over others, institutions expose their position in the cultural heritage landscape, and this participant goes on to qualify his narrative of constraint to note that the library also buys 'interesting older material' on an annual basis using a separate budget for 'areas that are of particular interest or documents which we feel are of national value and should really be held by a national library' (I10). This account highlights how definitions of cultural heritage objects and institutional identity and purpose are in dialogue with each other.

Handling Knowledge Complexity – Description

Archivists' role in promoting the use of the holdings they acquire begins with their description of them. This process is recognised as being shaped by practitioners' personal backgrounds and their institutional cultures, as well as the power dynamics, of geography, class and gender, which govern the construction of meaning more broadly. In Duff and Harris' (2002: 275) analysis, every 'representation, every model of description, is biased because it reflects a particular world-view and is constructed to meet specific purposes. No representation can be complete. The representer's value system, shaped by and expressing a configuration of the forces mentioned above, is the final arbitrator on the content of a representation'. At the level of practice, description was indeed felt to be a complex process, as this participant explained:

> For me it's very important to get the knowledge of our holdings to the public … I also think that it's very important to show people what we do and how we do it because you can only understand a holding when you know how it was that we did something with it. It might not be in its original structure because for example we had to throw out some things or found things that didn't belong there and placed it somewhere else or gave it to another institution. I personally like to do guided tours of the archive … and explain what we do or have some documents and let people work with them. Help them maybe explain something but mainly let them work and let them answer questions. (I1)

The description of items was acknowledged to be loaded with artefacts of its journey to the user, which become inscribed in this metadata and are inseparable from the data itself. This participant estimated that no more than ten percent of her archive's users were non-researchers and she was confident about users' capacity to apprehend the complexity of the description process. As well as being the bridge that must be crossed in order to discover data, description provides a signpost to appropriate uses. Being transparent about what was deemed not to "belong there", and why, alongside providing this direction, becomes part of a dialogue between data, practitioners and users – a dialogue that is perhaps a cornerstone of the research process itself and certainly not a step that might be skipped or circumvented. For a researcher seeking to make novel use of a resource then, appreciation of its potential comes from developing an understanding of this process, rather than seeking to extricate holdings from their descriptions. Where institutions catered to other types of user, they made different decisions about how best to represent their holdings, as exemplified by archives that provided data about victims of the holocaust to their descendants as well as researchers:

> The families have not got the same needs. The families want to find information about one relative. So, we have a special way to index documents. Here, in the documents, we have a list of people. We have to put all the names in the catalogue and to make biographies, and to collect all the data we have on each person. That's very particular, because the families want to find everything about their relative. (I2)

Practitioners therefore sought to maintain their freedom to use their discretion when deciding what form of description would be most useful for their users. Description was informed by the user at all institutions:

> I believe there will be a need to think about … different strategies to make stuff even more easy and accessible but which will have to be … outside the classical description [but rather] narrative-based introduction to collections … making people a bit curious … maybe layers upon documents with pop-up explanations … But even at best, we try our best, they will not be, I mean, completely usable by anybody [although the institution's specialism appeals to those who] will not be people who classically would use an archive … not only historians or genealogists but really casual users and … there will be a limit to what we can explain by our type of description. (I6)

Institutions must therefore prioritise what they are able to offer to potential users of different backgrounds and this would be informed by their sense of the capacities of those users to work with their material and metadata. There was also a sense of what valid engagement with an archive looks like:

> … they have to be as archive, from the point of view of the archival description. You don't have to make lower your level to meeting the request of the users. No. This is my personal point of view and until now it is my direction, my orientation. So, first as we aren't, I don't know… How can I say – a recreational centre? We are an archive and as archive we are to follow the rules of the archives and coming to that views we are to provide our

> information … maybe I present myself like a very rigid person, I know. But I think that every framework and every environment has its rule. (I8)

This was one manifestation of a drive to uphold the rigor of academic research that was common among participants, whose desire to expand use of their resources was not without qualification. This is not to say they would go as far as to demand particular *qualifications*, indeed it was expected that users' skills in navigating the archive would need to be complemented by archivists' expertise:

> And our descriptions here are very, very detailed. For example, I have a manuscript here and I'm describing it. I look through it. I look for every name. Except first names, if there is just a mention of an Anna or a Maria I don't write that down because no one will look for a Maria or Anna. But a last name, that's okay, that might be of use. And we look for all the names of cities, countries, regions and so on, or institutions. That takes quite a lot of time. But at the end, you know the whole thing. It helps us very much when we have to answer requests. Because on one side you can find all that data in our database, and on the other side, even if there are some technical [reasons a user] can't find it, we can find it. (I1)

This example is revealing because it suggests that some aspects of description are clearly routine, content a computer could read and conceivably produce, but it is the archivist's process of description that embeds in her the knowledge that allows her to "find it" when the channel from database to user fails. It is then inevitable that an overwhelming amount of data that does not become formally recorded as metadata is stored as embodied knowledge in cultural heritage practitioners themselves.

A significant amount of archivists' energy goes into creating descriptions; and in this case the collection level takes priority as a point of contact likely to pique researchers' interest:

> it's often the collection description that can trigger researchers to take a look. The more information the better, but it's also time-consuming. It takes me about… if it's a big collection, it can take up to two or three days to make a collection description. It depends. You have to balance time and effort in all aspects. … We try to put the context into the collection description. That's one of the necessities. The portal website indeed is built to offer digitised documents which are linked to their collection description. That is the only context that a researcher has when he consults. Therefore, the description is so important and has to be standardised so that you get the same information from every collection description. (I3)

Description therefore plays a fundamentally important role in researchers' discovery of inspirational material. The first threat to the survival of knowledge in this process was that some data was hidden and likely to remain so because it had not been captured initially. Some institutions held items whose provenance was unknown. This could sometimes be addressed through research into other collections. Practitioners therefore devoted some of their time to pursuing clues to their holdings' relation to those at other institutions but it was noted that "there will always be blank spaces" (I5) where an item's origins were never

recorded. This of course presents problems for research integrity so practitioners tried to be proactive with new acquisitions, especially newly created knowledge over which they had greater control. Where practitioners sought to instil good descriptive practice in depositors, however, they had found that their standards were not upheld:

> I spent many years battling with academics who just wanted to give me metadata about their paper when describing their data. Getting high-quality metadata about this, the data itself, is still a bit of a struggle, and much of the time I have to rewrite what they've written, because it wasn't… it's not very accurate or it's not sufficiently descriptive … in terms of the descriptions that they provide, many, many times the data aren't described accurately, it doesn't provide the specific information about the data itself. In terms of the content, the type of information they provide is often vague… (I5)

This account from a digital repository demonstrates that the description process can be complicated even when a smooth path from data creation to deposit might be assumed, which underlines the vital role of archival gatekeepers even where other procedures were more routine, as in the same participant's description of the process of making data available in a repository as:

> [not] particularly complicated, I have to admit. There's not an awful lot of manual processing involved. If I get a STATA[9] … a DTA[10] file, I'll … export it as a CSV[11] file, and export the Codebook as well … [and] make sure that it can be used in some other format besides DTA, and I generate the PREMIS[12] metadata and store it on the file system, because we don't have a system to manage that at the moment. But the principles are there in terms of preservation … And that is a notable difference from my [conventional] archival colleagues, who are all about the original order in terms of preserving the paper copies because you can't just reorganise a box of papers without losing something, whereas in the digital realm, it's a bit easier to provide different views upon things without making irreversible changes… (I5)

Practice in "the digital realm" might therefore be more fluid, and perhaps open to relatively risk-free experimentation without the threat of *losing* contextual information, but the need for proficiency in creating effective descriptions was felt across institution types to demonstrate the value of practitioners' overview of collections.

The kinds of users and *uses* that were facilitated were determined at the many levels at which decisions were taken, including how to describe an item, which could be more or less

[9] "Stata is a complete, integrated statistical software package that provides everything you need for data analysis, data management, and graphics" https://www.stata.com/

[10] "The Stata_dta format (with extension .dta) is a proprietary binary format designed for use as the native format for datasets with Stata, a system for statistics and data analysis.
"https://www.loc.gov/preservation/digital/formats/fdd/fdd000471.shtm

[11] Comma separated version

[12] "The PREMIS Data Dictionary for Preservation Metadata is the international standard for metadata to support the preservation of digital objects and ensure their long-term usability" http://www.loc.gov/standards/premis/

conducive to its potential for sharing. Potential connections could be aborted where an item was loaded with meanings considered too profound to justify a more simplistic description that would demonstrate what they had in common with other items of a similar form. Inevitably, decisions were made to represent items' previous use that will affect their future use:

> … we decided not to make a characteristic description of the pictures. That means that we only give the original caption of the photograph. So, we decided, for us, the original caption is part of the archive, because we have private archives, but we also have photographic archives from press agencies, and so this kind of photographs have captions. These are propaganda captions, in fact, because these pictures were taken during the war by specific agencies … that is why we decided to transcribe the original caption, but without giving any other information. And also, because we don't have time, we don't have staff to describe individually each picture, so we inventorised pictures with a thematic logic. (I4)

The reasons for which these photographs were taken were therefore allowed to eclipse basic descriptions of what they showed, which could have much in common with other types of photograph. Presenting the original context of the holdings was prioritised over making connections with wider collections. This decision was taken on the basis of assumptions about future usefulness to researchers but also as the most strategic use of resources. This may appear to be a matter of furnishing users who come across a resource with one set of suggested connections rather than another but such an 'anticipation of an authoritative disclosure of meaning' (Butler, 1999) has the power to elevate one meaning above the exploration of complex alternatives.

Such decisions were often guided by both the archivist's projections about potential uses and the availability of resources. Creating descriptions is therefore a pivotal process in creating cultural heritage in the mind of the reader – the users perceived to comprise the 'market' for this knowledge. As there is 'no representation without intention and interpretation' (Olson, 1994), the act of describing materials is in reflexive dialogue with practitioners' other daily tasks, notably their work providing assistance to researchers.

Handling Knowledge Complexity – Helping researchers with research problems and methods

Another fundamental part of cultural heritage practitioners' practice is helping researchers with research problems and methods. Each of the participants' institutions provided help for researchers as part of their daily operations. The delivery of assistance varied from a specialist department serving the needs of researchers and research organisations using a national library to small archives where there might be only one archivist on hand to respond to queries. A common query that took up much of practitioners' time was finding specific items for researchers who were not able to find them using the online catalogue. In addition, a deeper involvement in the research process was often necessary:

> We have someone here today who is working on a doctoral thesis on converted Jews. In this case specifically, many different collections can serve

> his research question, it depends on what emphasis he wants to lay … if people come in with a specific research question, we can point them in the right direction and work on a specific source. It all depends on the research question they have. Most things we have are available, but the difficulties we have now are that … most of our collections are hidden in the sense that the collection description is not yet online ... It's more of a dialogue. What happens often is researchers specifically reach out to us because they know that we're specialised in [country]. Once they email, we usually start a dialogue and try to find out what their research question is, because a research question in itself might not be specific enough to determine which collection would be most appropriate. Usually, we ask, how do you want to approach it, what are your sub-questions, is there a specific angle you're trying to work on? We try to give, I guess, custom-made service to guide researchers one-on-one in their research, and once they come here, it's very labour-intensive, but we don't get that many questions. It's more of a dialogue. (I3)

This example shows the depth of interaction between practitioners and researchers, with practitioners central to the discovery of resources but also shaping the research process itself as they guide researchers through aspects of knowledge complexity. Another participant agreed that researchers' involvement with her institution was more complex than accessing an interface and described her role in training people to use an archive:

> For example, one teacher calls me because he had the idea … that all his students could be able to use the interviews as archive documents to write history, so to analyse the … context in which the interview is made and so. So, he asked me to give the opportunity to the students to consult many interviews conserved here at X. And we took a lot of time to explain to the students the context in which these interviews were realised … So, we gave a course about the way to analyse an interview, also about the way we conserve interviews here in the centre. Yes, that can be this kind of support that we give to universities to a campaign to go with the students and to help them in their work. (I4)

This was one of the ways in which cultural heritage institutions were active in nurturing effective research through fostering familiarity with archival processes and training in historical analysis. This example highlights practitioners' feelings that – as well as adopting the philosophy of opening access to materials – it was crucial to support the development of skills that allowed researchers to interpret the historical record themselves. At other institutions, practitioners were more reactive due to strained resources:

> We try to [engage with researchers] but we have a limited staff. So, in the local depot or here [in the national archive], you can ask questions. "I need that", "can you help me with that?", and so we talk with the persons and we try to help. You can look into that source, can be helpful, or that source can be helpful … When you have a problem with our digital archives, you'll have to fill in a form and maybe one half-hour per day I respond to these questions, but the responses are very standardised because we can't go in. We don't

> have the staff to go into all the difficulties of their questions, so you give general directions. (I7)

Researchers' experience of this archive was therefore less personalised as a result of the high ratio of users to practitioners and although they might be engaged in a dialogue with an archivist it was likely to be of a more instrumental nature, with a requirement on the researcher to stimulate and steer the process. This participant was somewhat frustrated by the implications of these constraints for the quality of research informed by his collections if he was not able to aid researchers who might otherwise perform "Google lookalike" (I7) searches that would marginalise sources that specialist intervention would help them to discover. Another participant suggested that more focused support for researchers was a rarity in archives but that her institution was atypical in its employment of dedicated historians who provided support for researchers:

> [researchers] need scientific support from the historians, not from the archivist … About maybe the interpretation, about more information that you don't find in the documents, for example. But it's another kind of service that you provide to the user, different and separated to the normal activities of an archive … But this depends on the nature of our institution because we are, yes, we are an archive … but we are also an institution for historical research, we promote historical researches. And for this reason, we receive this, kind of, request of support on the historical research. (I8)

For this participant then, providing in-depth support for researchers was not a performance of her institutional identity, instead it was quite plausible that an archive should exist that did not support historical research in any overt way. Neither our interviews nor our survey data have revealed practitioners whose practice is impervious to or unsupportive of researchers' needs, however. At a national library, a manager of research services outlined engagement with researchers' problems across the institution:

> I mean there are curatorial staff, there are … colleagues from the reference team who work in reading rooms, so reference and reader services run the reading rooms together sort of. I also have some staff working in the reading rooms and we try to sort of tap into the feedback that they get … I have a team of subject librarians who … go out to conferences, they work with other research organisations, they all work a little bit in the reading rooms to sort of keep an eye on what's happening in that space. And we've just a few weeks ago hired a [service insight manager]. That's a colleague whose role effectively is to collect all the knowledge that we have internally, to look at all the sort of external studies and research that is done and then identify the gaps in there, and then commission new research. Either stuff that we do internally or that we might pay someone else to do to make sure we have as accurate a picture of research as we need and concerns and problems. (I10)

This institution was actively engaged in furthering its understanding of researchers' needs. Working to capture experiences of the research process was then core to the library's mission and was overseen by the specialist research services department, whose primary task was described as ensuring the continuing *relevance* of the institution. For other cultural

heritage institutions, of course, relevance may not be so closely bound to the use of their holdings for research.

Handling Knowledge Complexity – Serving ‘non-research’ users

Researchers were the primary audience of many institutions. At a university archive “at least 99.9%” (I5) of users were said to be researchers. Where institutions included multiple types of cultural heritage collections, different user groups were catered to:

> … we have different audience depending on the subject of the materials, on the information that we have published and so on. So, researchers are, for sure, eighty percent of our audience if we are talking about archival materials but we have another kind of audience for other information that we are able to provide. This audience is made by relatives of victims [of the Holocaust] or family members and so on. (I8)

At a national archive, the vast majority of users were genealogists and only around twenty percent were researchers. When asked what proportion of their collections were used by researchers, however, participants tended to state very high numbers, or as one archivist claimed: “our holdings are a hundred percent for researchers” (I1). Of course, the broader the perception of what constitutes a researcher, the more loosely sketched the delivery of services to that audience:

> … we often talk about the research audience, and that includes people who start their own businesses, so we have a business and IP centre that has had satellites across the country in other libraries. So, we get people from the business community who basically come here to find out about patents, how to set up their own company and how to develop business models and the like. There's traditional researchers from universities, freelancers, there are amateur historians, say family historians, and others. There's a very broad mix of people. (I10)

All participants were keenly aware of different publics they served and could confidently estimate what proportion of their total users each group represented. This awareness was seen as key to their ability to provide services appropriately:

> … we must make a difference between researchers working inside the institutions, and academic researchers working in universities or in other research centres and coming here to consult our archives. We know that seventy percent of our external users are academic researchers, but in academic researchers, I also count academic students who do research in the context of their studies. Yes, that’s seventy percent of the users … But the problem is that it's really difficult when we have, for example, web analytics data, it's impossible to see if the person using the computer is a researcher or is another person, another profile, and so we cannot say precisely which kind of collection of holdings are consulted by the researchers. (I3)

Where a user's identity or affiliation became obscured, it presented a challenge to tailoring content. This was a fundamental feature of online access:

> We do a lot of sort of cultural engagement and learning, and there is for sure material on there that isn't used by researchers. But we can't exactly tell because if someone just browses the website and finds some material that's aimed at schools but might actually be relevant for their own research, either because they research schools or it just happens to have a digitised image of a manuscript that they're interested in. And we could if you're not logged in or anything, we couldn't necessarily tell whether you are a researcher or what user you are. We're trying to get better data on that, but I mean the short answer is in principle everything could be used by researchers. In practice, it certainly isn't because not everything we have in our collections [has been] even once looked at by humans since it was added to the collections. (I10)

Providing a superabundance of accessible cultural heritage could therefore compromise institutions' ability to remain in dialogue with researchers' needs as the nature of their engagement becomes lost from view amid that of all other users. Offline, practitioners who identified a mix of user types tended to be quite considered in their approaches, as in the case of catering to both researchers and families of Holocaust victims. This case amounted to privileging access to certain knowledge for a particular user group, with the intended consequence of controlling the researcher user group's access as it was explained that, legally, the archive "cannot make the information available on the website, but we can put it in the reading room" (I2). The opportunity to drill down to this personal level would therefore not be as obvious to a researcher looking online in the first instance. This differentiation was also rationalised by the sense that researchers and non-researchers had very different needs and expectations of how knowledge should be presented. For users looking into their family records, it was common to "always do the standard search. For them, we always consult the same archival series to give them a certain package" (I3). This practice was actually a manifestation of the prioritisation of *research* as further work would be "too labour-intensive" and a non-research user would therefore only qualify for additional searching of other collections if they presented themselves in person to pursue the kind of dialogue usually reserved for researchers.

Elsewhere, utility for non-researcher users steered practice:

> … we have ties with the universities so we want to talk with them to help them with studies, complex studies too, but on the other side there is the reality, the genealogists, and this means we have to make the decision for example for digitisation. What do we do? And because eighty percent of our visitors are genealogists, so the first priority was were there church records and civil registration [to be digitised]? There is also a project … it's very similar and it's ideal for research. But it's not the first project. The first project is we work for the genealogists. (I7)

At this institution, the critical mass of non-researchers leads the allocation of constrained resources, resulting in greater accessibility of data of genealogical interest and the relegation

of potential research knowledge. *Working for* that user group as their core business went against archivists' motivation to serve researchers. Furthermore, it might be speculated that such pragmatism might stymie the in-depth engagement of other institutions were they to face similarly limited resources as the labour-intensive *dialogue* that is their preferred method of support might be held up against the number of *packages* that might be produced with the same expenditure. At present, it seems that the majority user group holds sway but it is conceivable that the threshold for resource allocation is susceptible to shift in the event of drastic budget changes.

The work of cultural heritage practitioners might be best described as a conversation between acquisition, description and serving users. Each of these processes is informed by the others and each leaves its trace on the material with which practitioners' work. Data from survey respondents suggests that those who benefit from practitioners' work are unaware of the intricacies of conforming to policies and regulations, cataloguing collections and issues of sharing in particular (see Figure 9).

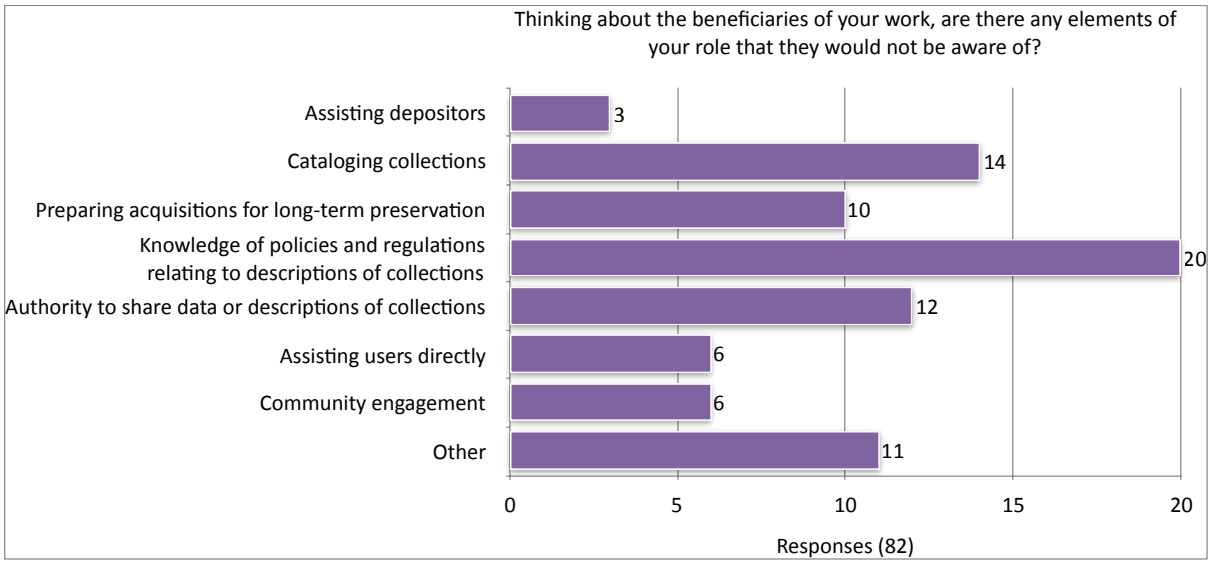

Figure 9: Elements of practitioners' roles survey respondents identified as unknown to their beneficiaries

Changing Practice

How is Archival Data Use Changing? – Getting a grip on what data is used and how

24% of survey respondents said that their institution did not monitor the percentage of its collections that was used and 20% did not know (see Figure 10). Survey data also suggested that digital use was more easily monitored than that which depended on physical access. In general, there was a perception that much cultural heritage knowledge was not used (see Figure 11).

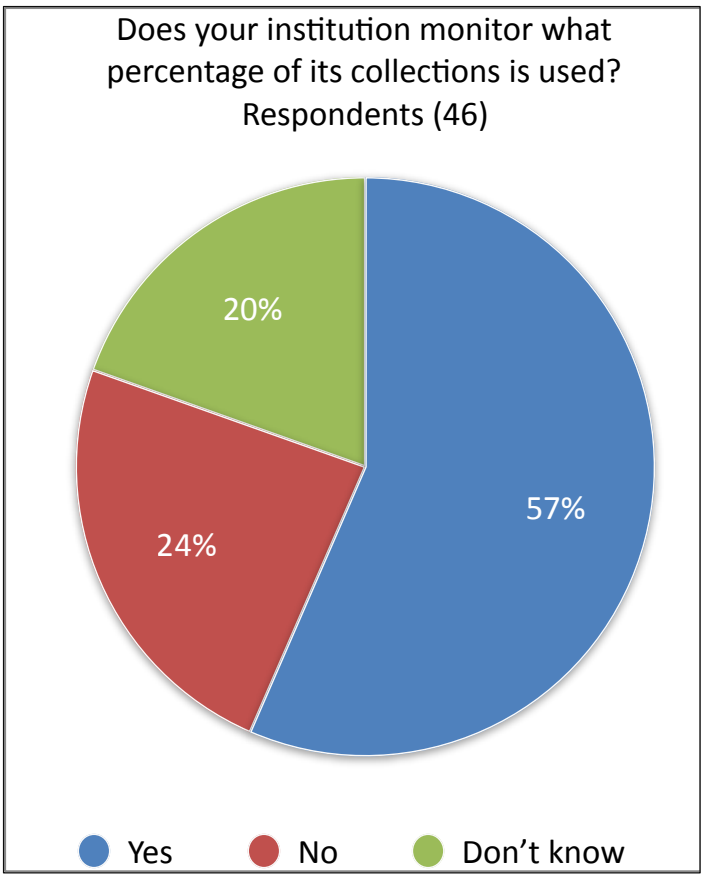

Figure 10: Survey respondents' institutional monitoring of holdings usage

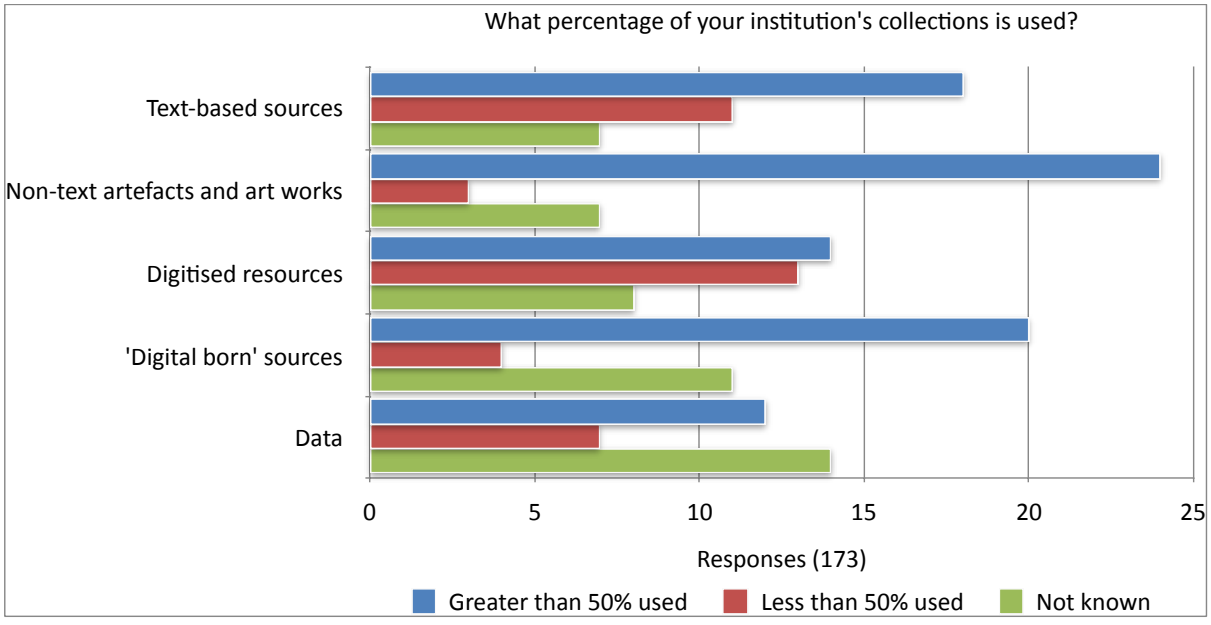

Figure 11: Survey respondents' perception of the percentage of their institutions' collections that are used, by type of holdings

The fundamental utility of their material was an ongoing pragmatic concern for cultural heritage practitioners but demonstrating their value was not always compatible with academic principles and methodologies, as this participant described:

> … for each [internal research] project we try to imagine the project in function of our collections. We have to argue to the [national] government to receive funds. We have to argue that we use collections of the institution, so we must use collections. The aim is to valorise our collection, but for a historian that's not possible to research only on a specific collection in an institution, so he needs to go to other institutions to consult other archives to have the opportunity to compare the information and to be really objective. So, it's not possible to limit the work of researchers to only our collections, but our collections are taken as beginning of a project, to think the project, really. (I4)

Valorising collections is therefore complicated when knowledge is contextualised as part of cultural heritage at large and practitioners did not seek to promote the value of their holdings at the expense of others'. There was no appetite for competition amongst peers, not least because: "if government cuts happen they tend to hit all of us" (I10). Instead, institutions made efforts to monitor the scale of use of their resources:

> … we keep a record of who is visiting the archives and … the material that they request, and then have internal reports that are produced about what kind of material is used and how popular it is … (I5)

For many, this level of monitoring was an ambition as, in practice there was "no real monitoring" (I7) or they were "trying to do it, but we start this kind of monitoring only very, very recently and not always we are able … to document the reuse of our material" (I8). Even where institutions had resources to put into keeping track of use, it was not an exact science:

> Interviewer: *Do you monitor the proportion of your holdings that get used and reused?*
> I10: Well yes, but a proportion is a bit difficult to say because we don't know what the total size of our collections is. That's partly to do with the difficulty of metadata that's been generated over hundreds of years, not always at standards that allow us to exactly tell what it is and also because sometimes it's difficult to define what sort of… If you have say twenty manuscripts that are bound into a volume is the usage that you look after for the whole volume or is it per manuscript? And depending on what you look at, what your metadata's like, you get a very different response … What we do monitor is visits to the website, the demand of items consulted online and the amount of items we deliver to our reading rooms. On that we have very precise data.

Understandings of the use made of holdings, even in proportional terms, were often evasive and the more complex the item – and the greater the potential for disparate uses – the more difficult it would be to grasp what use was made of it. The tendency of data to beget data meant that digital holdings leant themselves to being monitored more easily but producing this metadata was not in itself necessarily a progression from that obtained in analogue:

> It's difficult to say because we have people coming in our reading room, and so we have the number of demands for a year in the reading room … We only have the reference of the fonds of the archives, but it's difficult to say if the

> person was interested in one page or in many pages in one box, so it's a sort of monitoring but not so precise … We also have the number of demands of reproductions. If we send a digital reproduction of some documents, in this case, we have the number of the reproduction, of the demands of reproduction, and we also have, through the web analytics, we can see how many pictures, for example, are consulted online … But it's difficult to know with the web analytics tools which pictures exactly, so we can have a general average, a general number, but because pictures are in a database, we could have it. But it would take a lot of time to analyse all the data that we have with web analytics tools to know exactly which picture, which kinds of archives, which kinds also of keywords are introduced in the database. (I4)

Taking a step back from the core business of responding to user requests to take stock of how collections are used represented a use of resources that was not frequently prioritised. Instead, institutions sought to generate greater understanding through discrete projects, as one participant revealed:

> … we made a specific project during two years … and it was a project specifically interested in the way that people are consulting our database … but we know that it's absolutely not possible to make this kind of analysis every day in our institution, so that was really immediate. That's interesting now because it's a tool to give us information to develop the digitalisation policy, for instance, or the conservation policy in our institution for the following years. But it's totally impossible to continue this kind of recallment of data, and that's a pity, it would be interesting to know, for example, if we change something in the interface of our database, so to respond … to the complaints of the users, but it would be very interesting to know if the way they use the database is changing after we changed some things in the database. But now it's impossible to do that because we don't have the staff to do it, and the researchers who worked on the project are not working in the institution anymore, and so we lost the competencies in in fact. We learnt a lot about the needs of the users, about the way that they work with our documents and our database, but we learnt also a lot about web analytics, generally, about web analytics tools, but we don't have time and we don't have the skills here to really continue this work. So that's a pity and that's the problem, it's related to the way that research projects are funded because there is limited funding. (I4)

Only being able to resource this level of engagement with users' needs as a goal-driven, time-bound endeavour therefore provided institutions with a tantalising close-up of how their collections were being worked with but which was destined to recede from view once the snapshot was complete and this activity disappeared along with the skills and actors it required. Ultimately, learning about users' experience of accessing and using holdings came back to archivists' personal, day-to-day support:

> It's hard to [monitor use] because not all the documents are in the catalogue. And, they are less easy to find for the researchers, and so on. I have no way

> to tell that. [Discovery is made possible because] Every day, we have two people in the reading room who are helping users to find collections. (I2)

Even where researchers had sought to go it alone in their discovery process, the result was that the potential for system-generated metadata was exchanged for an enriched understanding of researchers' needs embodied in the archivists. Such barriers to discovery provide gatekeepers with an opportunity to stay abreast of patterns of use as a matter of routine, as was the case where there was international demand for physical resources:

> … we get requests from all over the world, it's very difficult for people to come here and consult the documents. We still have a research service. If people come here to consult, they often come for the big series. I mean, e.g., immigration files, we have 20,000 of those digitally, they come a lot for those. I think only ten percent to fifteen percent is consulted on a regular basis by external people. (I3)

For practitioners, staying close to both their material and trends in how it is used would have obvious benefits. Depending on the structure of an institution, however, knowledge about use could also be splintered across departments:

> It's a bit complicated because, I mean, at the time we have some collections online and we have some information concerning the usage of… I mean, how many people use it and how many people search. But, I mean, these are very small and unique collections which are not … our core collections and we also have a reading room and a … research department which untypically for many archives does… I mean we do not support the users in-house. That is done by research at the separate reading room. Which means again we have less, let's say day-to-day feel for what people are using. (I6)

This reflection from the head of archive descriptions, who described his work as leading to a 'gut feeling what is interesting and what is not' (I6), shows how knowledge about the use of collections depends on an institutional structure that is conducive to the flow of this information. Having a close link to the researcher's domain had fostered some transformative practice. This participant's role and department (a university digital repository) were created for the purpose of supporting new avenues of research:

> Because it's… a lot of it relates to research, that [researchers are] looking to find out what kind of research has taken place, but we also have personal papers of many of the academics that have worked here over the years where they've kept notes that they haven't previously made available. That is a useful research resource as well. And some of … this information is often used as a basis for, say, papers, books, other publications. (I5)

This was one of the ways in which institutions had been proactive in their approach to data use, identifying what could be of use to researchers and making those resources available for the first time.

How is Archival Data Use Changing? – The changing uses of archival data

Uses of archival data were found to be changing as a result of the increased visibility of descriptive metadata, research artefacts themselves and/or their underlying data. Practitioners had noticed that researchers were approaching them with more refined questions:

> … they are even coming to the reading room with [resource identifiers] to see documents. So, it's a new approach, because ten years ago, people would have to come to the reading room and ask if we have some information on a particular subject, and so on. And now, they come with a list of documents that they want to see, and they also see more documents because they are making more researches, looking and asking our archivists about documents. But, they ultimately, have some documents [they want] to see. (I2)

This represented a change from researchers asking a "general question like, do you have something about Holocaust?" (I8) and seeking guidance from archivists as to what the possibilities for narrowing their research questions might be. Other researchers who had already refined their inquiries had been awaiting such a development, as they had previously sought data that they knew existed but had been denied access as it was not available in an appropriate format:

> Because before, they asked for the database, for example, or a part of the database. Of course, we can't provide our database but probably now the kind of research for which they ask for the data [for] can be done directly on the web. So, I don't remember recently this kind of request from researchers. Sometimes we got, we have got requests like: can you provide a list of… I don't know, people arrested in this specific city or so on, so you have to go in your database and provide a list. Okay. But now, okay, this is a kind of query that they can do directly on the web so. Okay, it depends on the capacity of the people of querying data, okay, but I think things are a little changed from this point of view. (I8)

Ending the frustrations of researchers who had been able to discover but not access data was therefore an important step as well as promoting discoverability for other users. It was noted, however, that only those who were skilled in wielding these new tools would use the archive in this way. The individual dispositions of researchers were also a concern when institutions were promoting the use of their holdings beyond the context of their collections. This was seen as increasing the vulnerability of holdings to "bad use" such as Holocaust denial but the answer was also thought to be found in greater sharing, with the *exposure* of "vivid" (I2) narrative the key to undermining the twisting of historical sources to undesirable political ends.

There was general agreement that the changing use of collections developed in tandem with new skills and vocabularies through which imaginaries of knowledge might be articulated. This was seen to be deconstructing the research process:

> There has always been an approach in the academic world that researchers always… are willing to share their data, but they always think of the paper as the primary output by which people will find it. They just expect them to read the paper and get in touch with them to ask for the data … whereas as a result of having a repository of various types, not just our own but things like Figshare[13] and Zenodo[14] and others, there's a much greater emphasis on the data as an output in its own right and to be used as a basis for research that doesn't relate to the subject field. (I5)

In this way, the new competencies and tools at researchers' disposal were breaking down conventions of communication of knowledge and overriding barriers to engagement such as academic disciplines. Such fluidity was not necessarily a commitment of actors at every stage of the process, however:

> We get a lot of students who are just interested in the data so that they can analyse it in a research project, or academics who are just focussing upon a specific subject area and want to see those resources, or not just the data but also use… they don't really use the code, but they reuse the interview guides that we have or the questionnaires for similar research, and they're interested in those, what would traditionally be seen as supporting material as a basis for global work. From that perspective, yes, it's broadened the interest in these kinds of resources. Whether the metadata itself has actually improved access to things, I don't know. It's a slow process. I spent many years battling with academics who just wanted to give me metadata about their paper when describing their data. Getting high-quality metadata about this, the data itself, is still a bit of a struggle, and much of the time I have to rewrite what they've written, because it wasn't… it's not very accurate or it's not sufficiently descriptive. (I5)

Practitioners were therefore tasked with facilitating a flow of knowledge without any bottlenecks caused by discrepancies between the expectations and practices of users and those of others acting on the data. This facilitation was dependent on a culture of decision making and practice conducive to breaking down barriers to knowledge flow.

How is Archival Data Use Changing? – Decision making about alternative ways of organising material and their impact

Most participants had experienced changes to the organisation of holdings. Some practitioners were frustrated by the pace of change within their institutions:

> Yes, the material organisation of the collections, it's still the old one. It hasn't changed. Only now with the digital, we have a digital repository, so that's a change. It's also a difficulty because it's rather new. The colleagues of other services, they don't know it. Our search engine, it also didn't change, but yes,

[13] "figshare is a repository where users can make all of their research outputs available in a citable, shareable and discoverable manner" https://figshare.com
[14] "…a catch-all repository for EC funded research." https://zenodo.org

> you have to look into new technology, new ways of developing new methodology. Even for me it's difficult to understand our own search engine because sometimes I know there's a result and I can't find it. (I7)

Even where organisational change was instituted then, there was a risk that it might not bring about desired changes in practice as new systems did not necessarily become embedded or were not accompanied by the appropriate tools. If improvements to discoverability were overlooked despite being a widely acknowledged need amongst practitioners, there might be grave consequences for users and the future utility of holdings. This problem was thought to be a matter of decisions being made without a full appreciation of their effects:

> I think the people who take the decision must know the context of their decision and the impact of it and the consequences, and I think that's one of the problems. I think they see we can do that, we have quick wins, we can spare money, we can do that, but they don't look at all the consequences. (I7)

There were therefore fears that adoption of new technologies and practices was not governed by long-term strategy and could be in thrall to passing trends. Everyday decision-making was mostly done at a local level, however, as this Head of archive descriptions explained:

> It's really something that I'm doing quite instinctively and the decision, I mean, as I've experienced, it's understanding that it's needed. I mean by talking to other people and understanding and it's not just like some strange archival urge I have. I mean trying to figure out what I'm doing and [tracing that] back to structural changes [that] are needed and make sense. And that is by talking to other people in-house, like the research department or the archival administration unit … Then selling it to my boss and hoping that what I do makes sense … I'm not a very theoretical person like that. I … talk to a lot of people, try to sell my point of view, test it, see what people think and then just do it … (I6)

At this institution, a strategic overview combined with input from a range of colleagues still had to be *sold* to the higher level of management. Elsewhere, it was suggested that this level of management was unlikely to counter with an alternative vision:

> Interviewer: And you say bigger changes would have to be approved higher up, but would they be coming from higher up?
> I8: Yes, they must be approved because when you talk about big changes you mean also investment. Okay. So, if you need to invest more money or find money for sure you have to involve high levels.
> Interviewer: But does it usually come from that level, it's usually coming from lower down you suggested?
> I8: … Yes, yes. No. The [higher managerial] level don't have…
> Interviewer: Any ideas…?
> I8: No. (I8)

Practitioners therefore saw themselves as the source of creative thinking about how to move practice forward, although the degree to which they felt empowered to act on their ideas varied. One development that had been embraced, at least at some level, across the sector, was the move from analogue to digital. This meant that even where the organisation of material had not really changed in recent memory, the digital revolution had mandated that institutions revisit their fundamental practices. For example, one participant cited the development of an online catalogue as the reason all her institution's collections now had proper descriptive metadata:

> Because, we have had an online catalogue since ten years ago. Before that, you had to go to the paper card file. That's a huge change, of course. We are more efficient to respond to the request, and that's a huge change. (I2)

A societal-level change could therefore reshuffle institutions' priorities, necessitating work they had not previously found time for when the alternative was to slip into irrelevance or obscurity. Another, related, external force acting on institutional practice was said to be the work of infrastructure projects:

> 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. (I3)

Shaking up institutional practice from the outside could therefore achieve significant change in a relatively short time. External influences were cited by many participants as the catalyst for adopting greater standardisation, as this account illustrates:

> Before we introduced the international standard archival descriptions, we used a customised metadata schema. When the museum was created in 1994, the archival description was not meant as an archival description. The staff members were collecting objects - photos, documents - to make the permanent exhibition. They created a metadata schema especially focusing on what is the topic of this photo or this document, where could we put it in the museum, what is the keyword so we can quickly find it to put it in the layout of certain panels? They developed a metadata schema which was accustomed to this permanent exhibition, and when the exhibition opened and more and more private persons came and donated, they continued to use this specific metadata schema, which was standardised for all items but it was not according to the international standards. When we made a transition in 2012, we did a mapping to see which fields were in the old schema and we tried to combine fields to fit the international standard, and that actually worked. The collection, the item-level descriptions, we had about 20,000 of them, are being transferred into international standards. We take out items

> that are donated by the same person and then we write a collection description, because that doesn't exist yet, and then we put the collection description with the digital images and the item descriptions on the portal website … We try, because there are very rich item descriptions, sometimes with biographies … it's very rich description, so we try to save them by mapping and transferring them. (I3)

Preserving the context of a resource was therefore a core concern of archivists, although the knowledge they had to work with might be understood as not so much a continuous narrative but more *archaeological* (Foucault, 1969) in that their understanding of what had influenced the representation of items at different points in time was critical to interpreting how it might be used in relation to other research resources.

How is Archival Data Use Changing? – The Importance of context of items in descriptions of archival holdings

Constituting context requires practitioners to craft a narrative that makes sense of the journey and potential trajectory of cultural heritage knowledge, without over-steering away from more ambitious destinations. Duff and Harris (2002: 276) describe this delicate process as 'working with context, continually locating it, constructing it, figuring and refiguring it. Context, in principle, is infinite. The describer selects certain layers for inclusion, and decides which of those to foreground. In this process, there is analysis, listing, reproduction, and so on, but its primary medium is narrative'. Documenting the context of knowledge was agreed to be a vital function of archives' role in supporting the appropriate use of their holdings:

> … the context where a document comes from is very important for interpreting it. And also, if you just draw out some document from this and there and so on you don't have a wide view of the whole, you just have this singular document and it can be misleading … That's also why we try to make out the provenance of our holdings or write something about it. We not only describe it but we also write about where it comes from, how it came to us, and how it came into existence. Was it a collection of a researcher or is it the documents of an organisation? It's something completely different because a researcher of course only has those documents that he needed. But an organisation also has the personal documents of people that worked there and not only the documents for one project or something smaller. (I1)

Metadata was therefore seen as one of the vehicles for conveying context to the user but using systems that were fit-for-purpose and enabling the user to travel between perspectives also played a role:

> I think generally it's useful if in the way you sort of structure your information you can capture relationships between objects and even things that aren't objects … Neither our systems nor our metadata is sort of properly set up to [provide this context]. So, in that sense I see less value in just describing an item on its own and more value in trying to in the way collections are put

> together, made available, to sort of build some of those links or make them more visible (I10)

A finely-crafted representation of context was seen as providing the foundations of sound research but researchers' experience and habits of exploring resources was also seen to be significant:

> … the [collection] description is so important and has to be standardised so that you get the same information from every collection description. We've discussed building a hierarchy in the portal website to show researchers where a specific document is physically, but … it's not that important anymore. Once you know the context of a collection and you know in which folder the document is, that's often all the context they need. I have almost never had a person here who wanted to follow the hierarchy in the collection completely, top down or bottom up. Context, I have a feeling, is becoming less important in the archival world, I mean context based on hierarchy. (I3)

Most participants thought context would continue to be relevant but it was also widely acknowledged that researchers had an increasingly narrow perception of context. This led many archivists to express caution about the "quick wins" of the keyword search:

> I never use the search engine when I want to do a research. Because in the end you have this, and you see things, or maybe I can use that. And when you go the straight way, and even when the search is still methodological, then you miss things, because you don't look for other sources, which can be important. You have to get to know your material for good research, and we can aid people with that, [advising them] to look into these sources and these sources, but [even] I learn things. So, I think context is that important … and one of the main problems is that new historians, the new school who use the digital tools, they don't have the feeling anymore with the context, the methodology of searching. They want quick wins. And you can't do decent research I think with quick wins. (I7)

The *feeling* of *getting to know material* was therefore endangered by keyword searches' bypassing of context, which also undermined the process through which archivists deepened their relationships with collections. In this instance, a sense of context was vital for developing an understanding of connections that might be missing or yet to be made, as one archivist reflected on the challenges of adapting to digital systems and research methods:

> … there is lots of stuff that isn't really accessed and I think finding a good way to making it clear how the … one isolated bit that you see here relates to the collections in total I would argue is almost more important now because in the past people sort of knew there would just be thousands of boxes of stuff and they had an understanding if they only look into two that there're still more than 900 that they haven't looked at. They don't necessarily have that same understanding by landing on the page that has some content. So as far as my personal view is [concerned, context is] even more important. (I10)

As well as the practically-reasoned argument that “when you go with the direct way, in the current state of the search engines, you miss the information” (I7), archivists were committed to representing context as a core tenet of archival practice. Contextual metadata was generally felt to be of profound importance to the use of archives, as this participant described:

> I think the context is the value, the context is the value of a single document or a collection of everything. Without context, you don’t have value. Documents don’t speak on their own, they speak if they are in a context, so… And in the digital, in the digital the context is much more important and relevant than in a physical archive. Because in the digital environment you are a lot in front of picture, if you don’t have the context that explain the value of the picture. That picture is mute I would say. You can appreciate the aesthetic, it is a beauty, very nice picture but you cannot appreciate the intrinsic meaning and value in general. (I8)

While the fundamental importance of context was agreed upon, there was still disagreement about what this meant for the future of knowledge amid research methods that circumvented an appreciation of hierarchy, even within an institution, as in this case:

> … the big question and really a debate between the archivists. We have archivists here, really traditional, who think that that the context is the only thing that users need. And so they must have access to the archives with the logic of an archivist and with the logic of the … hierarchy … And so, others say, no, people are researching with key words, and they're searching for specific documents and they don’t need to have more information about the context and they don’t understand the way that archivists are organising archive documents, so that it's not necessary to keep the context. We think here that a good way to work is between both options, so to develop a tool which gives the opportunity to make a search with key word, and maybe to make a more specific search. For example, if you enter key words, but you have thousands of results, you have the opportunity to reduce by choosing more specific key words, and with the opportunity to see these documents in the context of the archives. (I4)

It was implicit even in debating alternative approaches to knowledge discovery that researchers needed to be conscious of their choice of approach and skilled in carrying it out. In practice, even where a dual approach was adopted, the balance was seen to have swung in favour of the key word search:

> Because for historians, for example, it's really important, but the fact is that the problem is that now even researchers, even historians are making their search with key words and not with the logic of archives. And so the aim is to give the opportunity to make the research by key words, also by the logic of archives if necessary with the context, but first of all with the key words because even researchers are doing their search now by key words. But when you find a document, then you can see the documents in its context, in its archival context, that’s the aim and the that’s the idea that we see for the

> new tool that the [national] Archives is developing for maybe in two or three years. We are working in this logic. (I4)

The exchange of the conventional *logic* of the archive for the new logic of the key word search is a sea change for archival practice, as archivists' knowledge of the deep connections of hierarchical context must either endure while no longer being reflected in the representation of knowledge or researchers' methods or become entirely redundant.

If the context of 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' (Duff and Harris, 2002: 276), will lie dormant. This may result in a lack of understanding of potential uses for items or, in the logic of openness, it may pave the way for new connections to innumerable unexplored referents.

How is Archival Data Use Changing? – Moving from analogue ways of working to digital systems

Cultural heritage institutions today deal with sources held in a mix of digital and analogue forms (see Figure 4). The ways in which they handle access requests from their users therefore reflects this, with a high degree of automation, yet human intervention is still essential (see Figure 12).

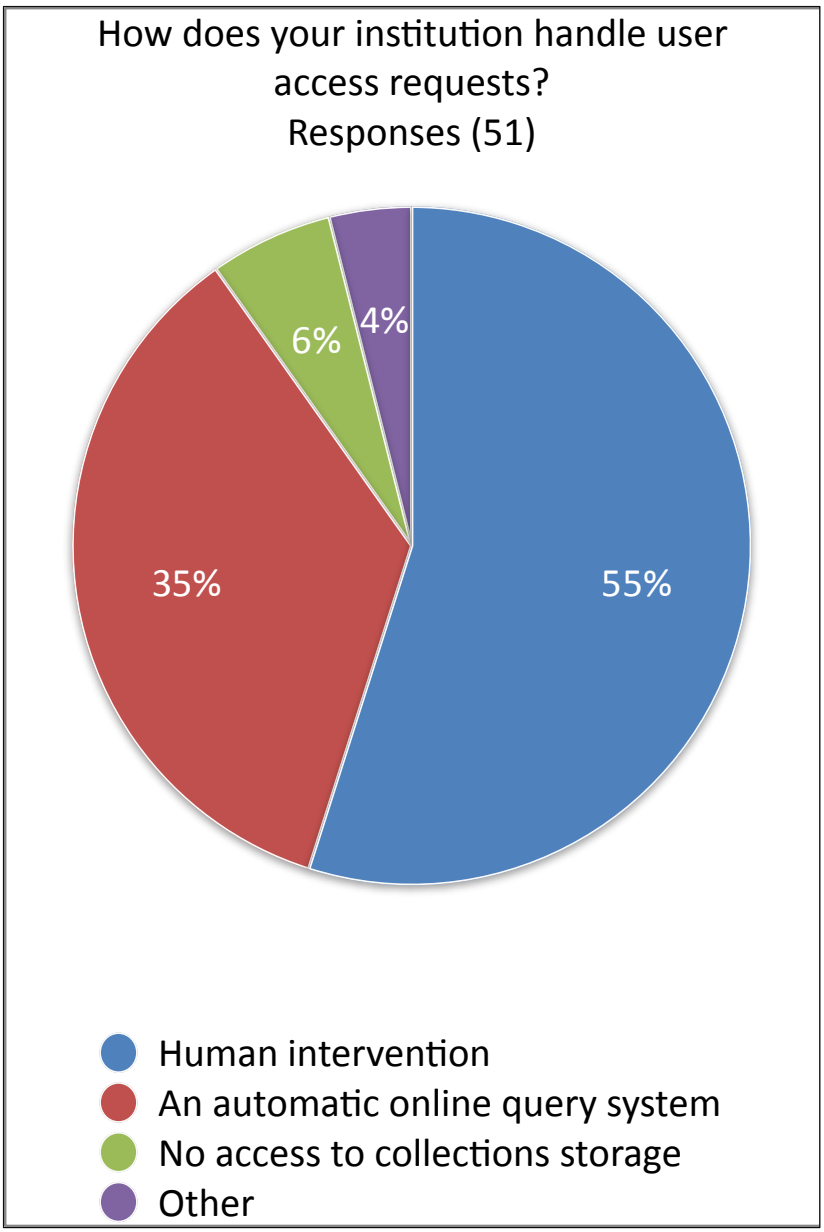

Figure 12: How survey respondents' institutions handle user access requests

Most institutions were in the process of digitising their holdings, with quite some way to go before all items and data were digitised and made available through the channels of sharing they were beginning to use. A number of reasons were given for the slow pace of this process, including the delayed commitment of colleagues at higher management level. Whereas archivists "were more open to the databases solution, because it will be easier to respond" (I2) to users' requests, those higher up in an institution often took longer to see the benefits of going digital.

> I think that the head of the [memory institution] has agreed with the databases only when we made the monument … At the entrance of the [memory institution], you have the [monument]. And, we need all the databases to make such a monument, so it became okay. But, before that, I'm not sure it was very clear for even the head of the [memory institution]. (I2)

This shift was therefore not always as straightforward as might be assumed but the value of digitisation was seen to align with many institutions' priorities:

> … digitisation is very important to us. We don't have that many original documents in our climatised room, but we tried, since the [original museum] was created in 1994, to digitally bring together all these collections, because they were very dispersed throughout [Country X] but also in [Y] and [Z]. That's really the focus point: if you ask something about the [specialism], we can give you an answer based on sources from very diverse collections which are physically all over the world. That's really what we try to do (I3)

> Digitisation had also opened up opportunities for acquisition as donors were more likely to offer a digital copy of items than originals. As part of an institution's mission, digitisation was seen as a process, not a goal, and offering data through a dedicated web portal tool searchable by key word, category and other terms was now regarded as preferable to the dead-end of "just giving digital copies by putting them with the metadata and the collection descriptions in a data management system" (I3).

Infrastructure for discoverability across institutions was therefore part of the progression to digital practices. Even at a digital repository, digitisation was not seen to be complete as:

> … one of the limitations of the repository is it is just a final store for data, it doesn't allow you to interact with it in any way. From that perspective, just making it available for download is sufficient for the moment, but I think we'd look at improving that over time. (I5)

In this example, an institution that would appear to be pioneering archival practice had made less progress in improving the discoverability of its data than some more conventional institutions. The creativity that had allowed complex knowledge to be opened up to digital methods was thought to be the source of expanding openness still further:

> [Digitisation is] not the big issue it was ten years ago. Nobody is, like, proud of his huge digitisation project … the question is now what to do with this material and I strongly believe that, first of all, the tendency will be towards openness, I hope. And a lot will be available online … And it's very searchable but it can be much more. (I6)

Another participant pointed out that collections that were purely digital offered new possibilities through their fluid nature:

> [There] is a notable difference from my archival colleagues, who are all about the original order in terms of preserving the paper copies because you can't just reorganise a box of papers without losing something, whereas in the digital realm, it's a bit easier to provide different views upon things without making irreversible changes; not always but… (I5)

> Experimentation that would involve a prohibitive investment of resources if applied to physical artefacts was therefore possible with digital holdings. Research methods drawing on digitisation were also not static and participants saw a need to respond to evolution within the digital revolution, with one reflecting on training through which she had learnt that the use of advanced search methods was declining, after having invested in this functionality for her institution's web portal. Another archivist felt that "the difference and the real evolution will come from the quality", in that, once the "fashion" for launching digital tools had lost its novelty, only the best ideas would attract attention and investment (I8).

A more holistic approach was therefore advocated once institutions had settled into digital practices and started to think more strategically about how they might serve their specific needs and aims more exactly. In developing support for computational research methods, this was seen to encompass both technical and personal dimensions:

> 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. But … for doing this onsite it needs a personal component where we help people to get access to certain collections, something we'll need to police because while they might be able to work on the premises we have to make sure that they don't take a copy and walk out of the premises. But we don't want them to feel sort of too much that they're boxed in; rather help them as much as possible to get as much out of our collections as we can. 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. (I10)

It was agreed that supporting computational research methods meant taking operational measures that were both technical and personal and that both elements were needed to "police" appropriate use of holdings, as well as helping researchers realise their potential. The biggest impact of this digital turn on archival practice was that assisting researchers who could not find what they were looking for in the online catalogue had become a large part of the job of many archivists.

How is Archival Data Use Changing? – The Consequences of Digital Discoverability

Digital methods of communication played an important role in improving the discoverability of knowledge. Communicating with researchers using institutions' own websites and those of infrastructure portals was a significant channel for the regular dissemination of information (see Figure 13). However, only 39% of respondents to the survey considered that they had a significant amount of their holdings described online with only 9% stating that 100% of information describing their collections is available online (see Figure 14). Digital ways of working had yet to result in routinisation overtaking the application of practitioners' specialist skills as the main way in which new practices were adopted but routinised techniques were employed at a significant level (see Figure 15).

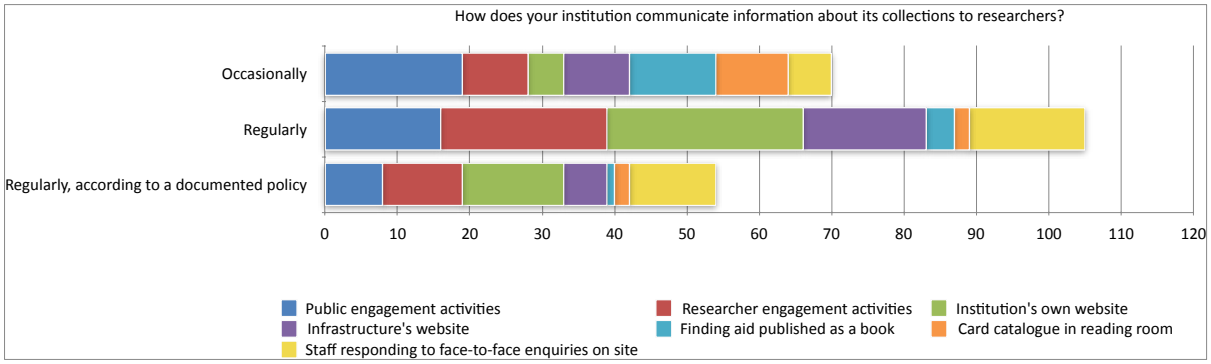

Figure 13: Methods of communicating information about collections to researchers reported by survey respondents

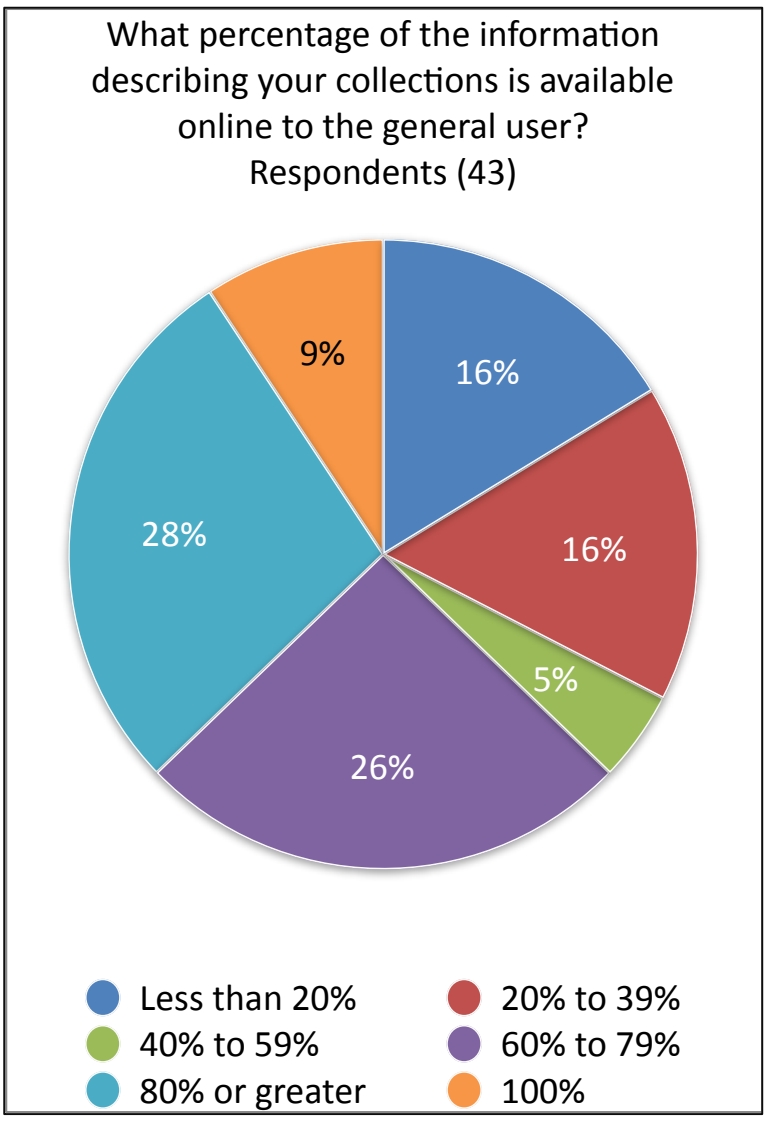

Figure 14: Survey respondents' percentage of the information (metadata) describing their collections is available online to the general user

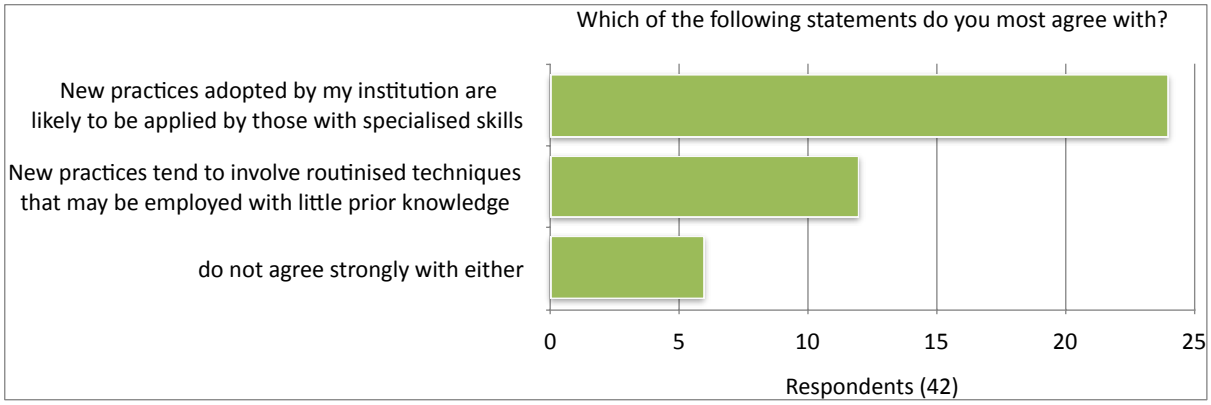

Figure 15: Survey respondents' perceptions of the significance of specialised skills versus routinisation in adopting new practices

As a consequence of going digital, some archivists felt they had a changed relationship with their collections:

> … because I think we have a different relation with descriptions. In a way, we use more digitised collections than the collections that are not digitised yet. Because, it is easier to show and there are no problems with the original documents. I think it is a problem, in fact. Because, digitised documents are more used than the other documents. It's not a question of… In French, we say *le forme est le fou*. How to explain it? What's inside the documents is interesting in both cases; digitised and not digitised. 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. (I2)

Far from developing practice being stymied by practitioners trapped in established habits of consulting physical materials, the ease of digital working has taken hold. Archivists as well as researchers are thus becoming more familiar with digital holdings as they eschew the troublesome non-digital. Digitising *everything* was widely regarded as *impossible* so the incorporation of complex knowledge into future research using digital methods may be precarious, requiring the double hurdles at the level of both archival gatekeeper and researcher to be overcome. Consequences for users were, however, largely framed in terms of the benefits of discoverability and independence that came with digital dissemination:

> But generally, it made a big difference with our database on the internet. People have different requests, more concise. They don't ask, do you have anything about this? They ask, can I have access to this and that? … They will search on their own and we have less work on this part (I1)

Researchers going further on their own was seen as a more efficient use of resources and one that has changed their interactions with archivists substantially as they have more refined questions and ideas about what they want to use by the time they visit the reading room:

> So, it's a new approach, because ten years ago, people would have to come to the reading room and ask if we have some information on a particular subject, and so on. And now, they come with a list of documents that they want to see, and they also see more documents because they are making more researches, looking and asking our archivists about documents. But, they ultimately, have some documents [they want] to see. (I2)

While the research process still involved a dialogue between practitioners and researchers, online catalogues had shifted the balance of control towards the user. For smaller institutions that had not previously enjoyed exposure to a wide audience, digitisation had expanded the proportion of material used:

> Because before people ask only for the same documents and the same collections. Thanks to the digital library and the publication of the inventories we recorded an interest for other collections that before [had] no requests, not requested or very rarely requested. So, the digital library is a very good … vehicle to make known all that we hold. (I8)

Digital *vehicles* have therefore served as a *pull* to draw users in to a range of sources. They could also be used to *push* researchers out of the routine of conventional archival inquiry, as making information available on the web, for example:

> is a way to have less work for us. So, you don't have fifty people by day that ask for information or anyway if they write to you asking for information you say: go to the website … they can have information directly on the web. So, it is a benefit for us and it is equally a benefit for the users. (I8)

Encouraging reluctant researchers to incorporate digital expediency at some level of their process reduced inefficient use of practitioners' time in a manner comparable to customer service roles that have become more specialised to take care of enquiries that cannot be resolved by customers' interaction with company websites. Some participants were apprehensive, however, that this utility could mask problems of discoverability where context was 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. (I10)

One of the consequences of digital discoverability for users of this national library is a shrunken 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, especially given the compounded obfuscation of incomplete metadata and inadequate systems, and so this is

not something that has been *lost* per se. Accepting that argument highlights the undeniable knowledge *gains* on the other side of the scales, as endorsed by this participant:

> I think I notice that people are much more interested in history, especially in their own history and the history of their family and the history of their local surroundings. It's not only something for old people or the typical teacher that doesn't work anymore but now needs a project. It's a big, big topic. I think that it can be a very good chance even for small institutions … to be present and to be noted. Not only by presenting metadata or information on the internet, but also [in combination with other forms of public engagement]. (I1)

Thinking beyond the advantages of digital discoverability for researchers that were their primary audience, practitioners had witnessed a shift in the potential of materials they had grown used to seeing used in a certain way. This outcome reflects the utopian ideals of opening access to data. Looking to the future, practitioners also predicted uses that were cause for concern:

> … data-linking is one of the limitations we have to take into account, and it's one of the primary factors in terms of restricting data, because even if you've removed all the direct identifiers, maybe indirect information that could be used to identify them, even something that seems innocuous like the type of building material used in a house could, in some cases, be used to identify the specific house in a region because they use certain types of material … and the identifiers that are used, even if they're a numeric ID, could be linked to an existing dataset. It could be linked to the personal data that people have stored elsewhere than they're supposed to. (I5)

The dark side of discoverability was a phenomenon of which practitioners had a growing awareness:

> Certainly, the machine processing is, I suppose, going to be even bigger soon than it is now, and … artificial intelligence has the potential to draw new conclusions from a large amount of data, particularly unstructured data, which … until quite recent years have resisted the broader analysis … if automated tools are able to make links between those datasets and then … infer conclusions about the people, if it's identified, then there's a significant danger to them. With all the way of what Facebook does and Google does in terms of linking information together, if that becomes even more prevalent, then there are dangers there in terms of providing data. (I5)

Digital discoverability therefore exposes a dark side that archivists were used to mediating as gatekeepers of material that is vulnerable to misuse and magnifies this threat beyond the potential that can be seen unaided by digital tools. The digital era arguably raises more questions than it answers. To keep pace with this era, archival practitioners must be in dialogue with technical expertise.

How is Archival Data Use Changing? – How technical expertise is used and integrated within the archive

Calling on or integrating the expertise of technical specialists within cultural heritage institutions had exposed dissonance between visions of technical development and practice. In particular, there was said to be a "definite" division between "IT" understandings of practice between and *archival* ways of thinking. In other words: "there is a clear distinction between content and… between process, is probably more accurate, and the technical infrastructure". (I5) That archival *content* and *process* were closely linked and separate from the *technical* framework, in this case the university's computing centre that *hosts* the archival elements, further reinforces the idea that this infrastructure was not part of the fabric of cultural heritage. IT specialists were embedded at some institutions, at least physically:

> … we have someone in charge of all the computers here. But this person … has not really technical skills about metadata, but he can manage computers, he can develop software, he can do these kind of things, but he doesn't have competencies about the way that metadata works and technical aspects of archives in fact. So, it's sometimes difficult to speak with him and to be able to make him understand what we would like to develop or to get in a tool or what we have as problems for the metadata. So, it's sometimes difficult, but in the other operational directions of the [national] Archives, they have someone who was an engineer at the beginning, but who is really capable to understand all the ways that archives work and the concept of metadata and [working with them] helps us to answer some technical problems … (I4)

Proximity did not engender effective communication where colleagues had no common "competencies" but such mutual understanding was clearly possible even from the starting point of an *engineering* mind-set. Indeed, there was evidence of practice that overcame regular challenges of communication:

> For the technical side of the development, we work with a software company which customises some elements of the portal website for us, and we have an in-house IT colleague … he's the bridge between me and the IT company. If I need something or I want something, I will explain it to him first, and then we will go to the IT company to get it. it makes it easier to transfer the message, because in the beginning … I had much trouble learning to speak IT… (I3)

Even where communication was successful, the onus was clearly on cultural heritage practitioners to learn the vernacular of the *technical*, as is discussed in more detail below. A lack of permanent integration of technical expertise within the institution was often seen as limiting opportunities, as was the case when this participant's institution acted on feedback from its users:

> … it would be very interesting to know if the way they use the database is changing after we changed some things in the database. But now it's impossible to do that because we don't have the staff to do it, and the

> researchers who worked on the project are not working in the institution anymore, and so we lost the competencies in in fact. (I4)

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:

> [The national government] don't give the priority to the scientific institutions, and so the problem is also that they are not conscious that digitisation and the digital environment give to the institutions new missions and demand more skills. But these skills and these missions … are still going, the conversion of the documents, the valorisation of the documents, the reading rooms and so. But we have no new other missions linked to digital environments, and the digital projects, the digital tools, the digital conservation of digitised documents and so, ask a lot of money, and they are not really conscious of that. So, 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 … We can only think the four years coming, but not further, and that's really difficult then. So, I hope that it will change in the year coming and if it doesn't change we will really have a problem. (I4)

Insufficient investment was a significant threat to institutions' ability to plan for and respond strategically to the digital era. In the absence of long-term security, many took up opportunities to share resources across the sector with enthusiasm. This exposes a training need that was identified by survey respondents who reported that continuing support to respond to new developments in practice was rare (see Figure 16). Survey respondents also reported a lack of proactive support for training by their institutions (see Figure 17).

Figure 16: Continuing support with new developments in practice as part of survey respondents' training

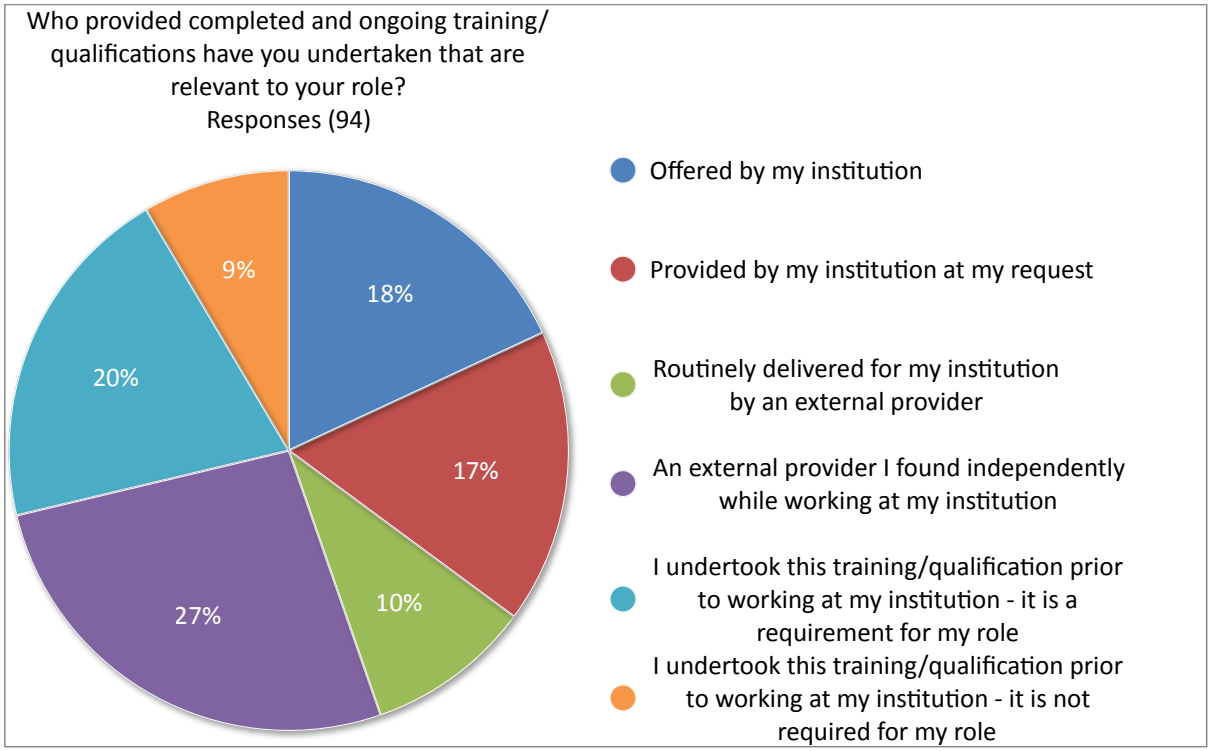

Figure 17: How survey respondents accessed training provision

As a distinct set of actors in the broader research data environment, the practices of cultural heritage practitioners have developed in tandem with, and informed by, the practices of researchers, some of which have been discussed above. It has been argued, however, that when it comes to recognising the imprint interpretation leaves on knowledge, archival practice has not reached the same maturity of reflexivity that researchers have dedicated their efforts to. Cook and Schwartz (2002: 175) identified a tendency towards *technical* explanations of archivists' role in knowledge creation:

'… even if the profession is now less passive, and more "up front" in the life cycle or continuum of record-keeping activities, or in designing new approaches to description, we believe that it does so largely in technical rather than substantial ways. The focus of most archival research over the past decade has been on creating and implementing standards, record-keeping requirements, process templates, and system architectures. It has not been on the substance or even nature of the archival contextual knowledge needed to put inside these empty shells to make them mean anything. By this focus almost exclusively on the technology and mechanics of archival processes, is not there reflected a desire to be the white-coated "scientific" clinician, unsoiled by the messy interpretation that is always endemic to performance?'

While a recognition of archivists as actors who influence the space in which they operate is a necessary companion to the conceptualisation of researchers' positionality – and in need of more attention – It would be a mistake to judge practitioners' turn towards applying certain tools to their 'inherently chaotic' (Cook and Schwartz, 2002: 176) world as a failed enterprise because of the persistence of complexity. Contrary to Cook and Schwartz's white-coated clinicians, this study's participants did not seek to separate themselves from their materials in order to maintain an illusion of unbiased 'science'. In describing their practice, rather than being preoccupied with the contamination of the essence of collections by their

interpretation, practitioners were more concerned with the cross-pollination of perspectives that would engender researchers' fruitful engagement with complexity. Fundamentally, cultural heritage practice is an ongoing dialogue that favours the latter. There is an inherent responsibility to apply a sufficiently critical research methodology when working with data from any source so that forces acting to interpret their meaning are not black-boxed (Latour, 1987) beyond the researcher's understanding. The next section discusses aspects of cultural heritage practices that are in themselves neglected facets of knowledge complexity, and explores what their hiddenness means for the possibilities of cultural heritage data use.

Common Knowledge: how does Cultural Heritage Practice Affect the Kinds of Data that are Shared Through, and Hidden from, the Historical Record?

Why do Cultural Heritage Institutions Share?

Both interviewees and survey respondents identified a public duty to share data, with which they felt engaged (see Figure 18). Both thought that the aggregation of information from different cultural heritage institutions was relevant to their institutions' operation and goals (see Figure 19) and there was a high level of involvement with, and interest in, aggregation projects (see Figure 20).

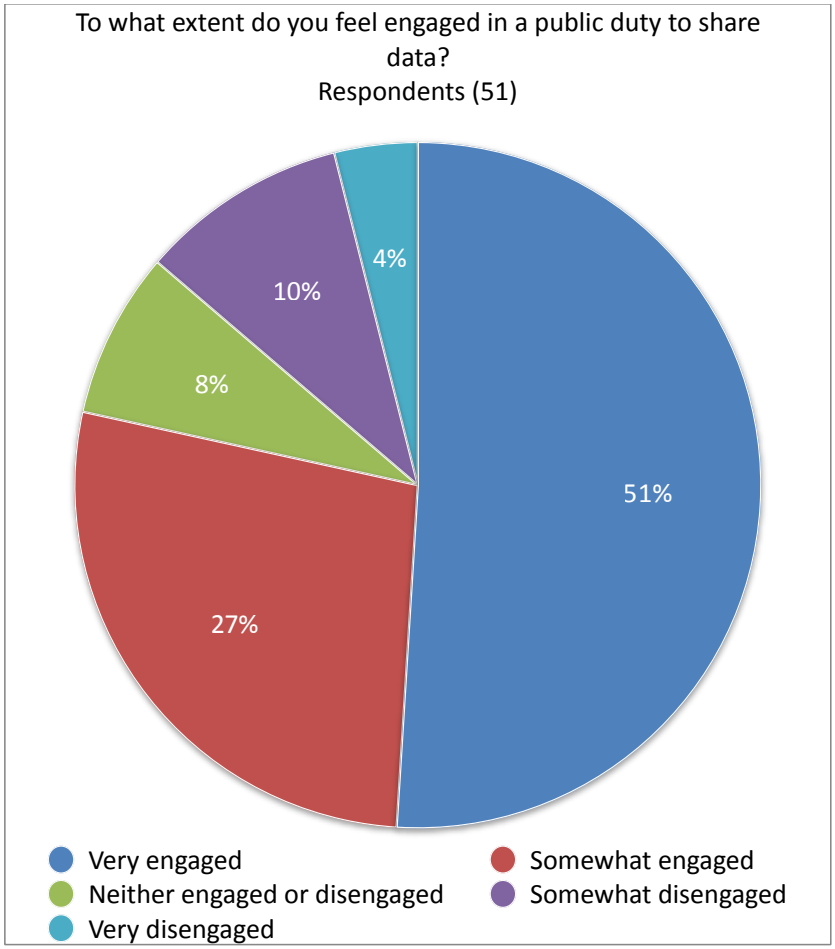

Figure 18: The extent to which survey respondents felt engaged in a public duty to share data

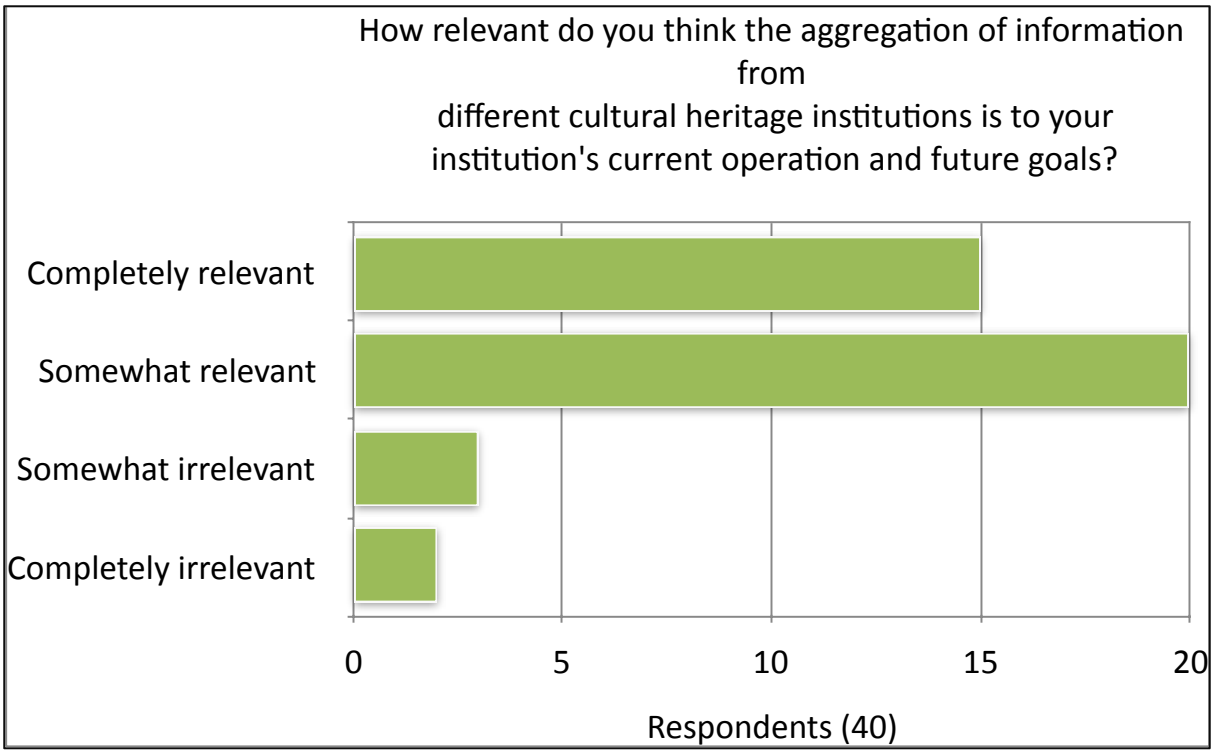

Figure 19: The relevance of aggregation projects to cultural heritage institutions' current operation and future goals, according to survey respondents

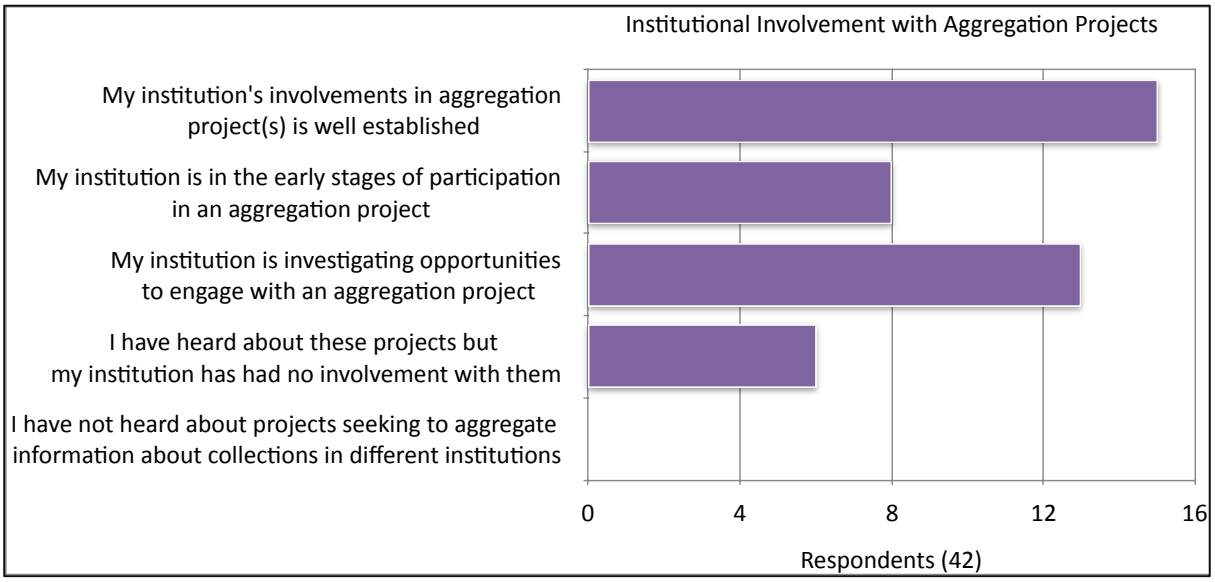

Figure 20: survey respondents' institutions' involvement with aggregation projects

The introduction of new technologies of knowledge could be seen to represent a moment of *breach and repair* in cultural heritage practice, whereby practitioners' reactions to the disruption of their practice offers the opportunity to glimpse the norms that may have been invisible until they were disrupted (Goffman, 1967; Sacks, 1974). This study found evidence of cultural clashes between *archival thinking* and *computational thinking* but, as Star (1990) suggests in referring to 'the myth of "two cultures" of those who work on machines vs. those who study or work with people', practitioners' accounts tended to be more nuanced, with

disruption more likely to be described as “evolution” in line with their vision of the mission of cultural heritage. A 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. When you participate, because we are all Europeans, 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. (I7)

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. Practitioners were keen to take a “global view” as well as reaping the benefits of an “outside opinion” (I6), which was useful when positioning their institutions within the constellation of the sector. Being “forced to evolve” provided “technical advantages” (I7) such as adopting metadata standards through learning from others, which allowed them to conform to widespread practices and not be left “behind” (I4). Adapting to sharing on this level could transform practice, as this practitioner describes:

> In 2010, we had a system where there was one 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 [X] 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 [X], and in 2014, we had a new work method. We learnt from working with [X] what the gain is from working with external partners. (I3)

Participants expressed openness to changing their practice when they were confident of the benefits of sharing. Infrastructure projects were seen as both “a good way to see the importance of standards and norms” and “to have a larger view about our field and other scientific fields” (I2). Once they had “learnt very quickly what the value is of collaboration”, archivists at one institution were keen to join the second phase of an aggregation project when asked:

> … if we would be interested in again being a guinea pig to implement tools to export valid [metadata] files which would then be uploaded onto the [X] portal website, which is a protocol that we’re still developing at the moment, but we should be able to create a sustainable connection between our portal website and the [X] portal website which would also be accessible for other partners to harvest the descriptions. (I3)

The “two-way” (I7) advantages of sharing were widely acknowledged. While conformity changed institutional practice, it also provided space for reaffirming institutional identities, which was a key motivation for joining infrastructure projects:

> 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. (I2)

Exposure beyond an institution's familiar catchment of researchers was therefore a prominent goal. Furthermore, participants were expansive in their visions of the value of sharing, which one archivist described as emanating from "the enrichment of knowledge" through historical and "scientific research … using scientific in a wide sense" (I8). The motivation for sharing was also seen as reflecting the nature of cultural heritage:

> … so our holdings are known better because we have some information that completes the holdings of the other institutions" (I1).

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 (I2). Participants were in unequivocal agreement that they were providing a service of public knowledge. This was rationalised thus:

> If we keep the information here in the documentation centre, what's the use? We get high school students or schools that visit us, that after their visit sometimes will start an educational project … we have different workshops. You can only create awareness if you share the information. Keeping it here doesn't really make any sense. That's why I think I'm pro putting the collection descriptions online, to get them out there and to show that there is something to discover and to learn here, that there is information available. (I3)

Dissemination of knowledge was also a perpetuation of institutional purpose, with visibility growing "enormously" through participation in aggregation networks, leading to increasing numbers of users both from afar and in reading rooms (I3). It was hoped that this could be built on to stimulate more researchers to undertake comparative studies based on cultural heritage items that would not been discovered in parallel previously. Changing research was seen as an inevitable consequence of changing archival practice:

> I think it will, perhaps, change the way that, even the subjects that [researchers] want to work on... Because, in the past, the holdings were not described and there was no way to even know that there was documents about a particular subject. So, I think it will give ideas to the researchers. (I2)

Practitioners were largely optimistic about the profound changes to research they believed were afoot. They were also broadly supportive of the cosmopolitan, democratic spirit of sharing embodied by infrastructures, as one archivist conveyed as giving them "the opportunity to connect to different worlds", through which they discovered new connections with other institutions (I3). This was aided by what another participant described as the "ego-less" institution whose founding tenets were "peace and openness and sharing" (I6). Stumbling blocks like differences in metadata schema continued to get in the way of closer

co-operation but “evolving” with other institutions within an infrastructure stimulated a general harmonisation of goals.

In studying how new technologies breach or strengthen the established social relations between institutions, it is possible to gain insight into the complexities of both the technological and social dimensions, in other words, we can perceive a reassembling of the social through changes in practice driven by seemingly mundane technologies. For the purpose of this study, the impact of these changes on the potential for complex knowledge to be discovered and used is the focus and survey data suggested a number of challenges for greater sharing (see Figure 21).

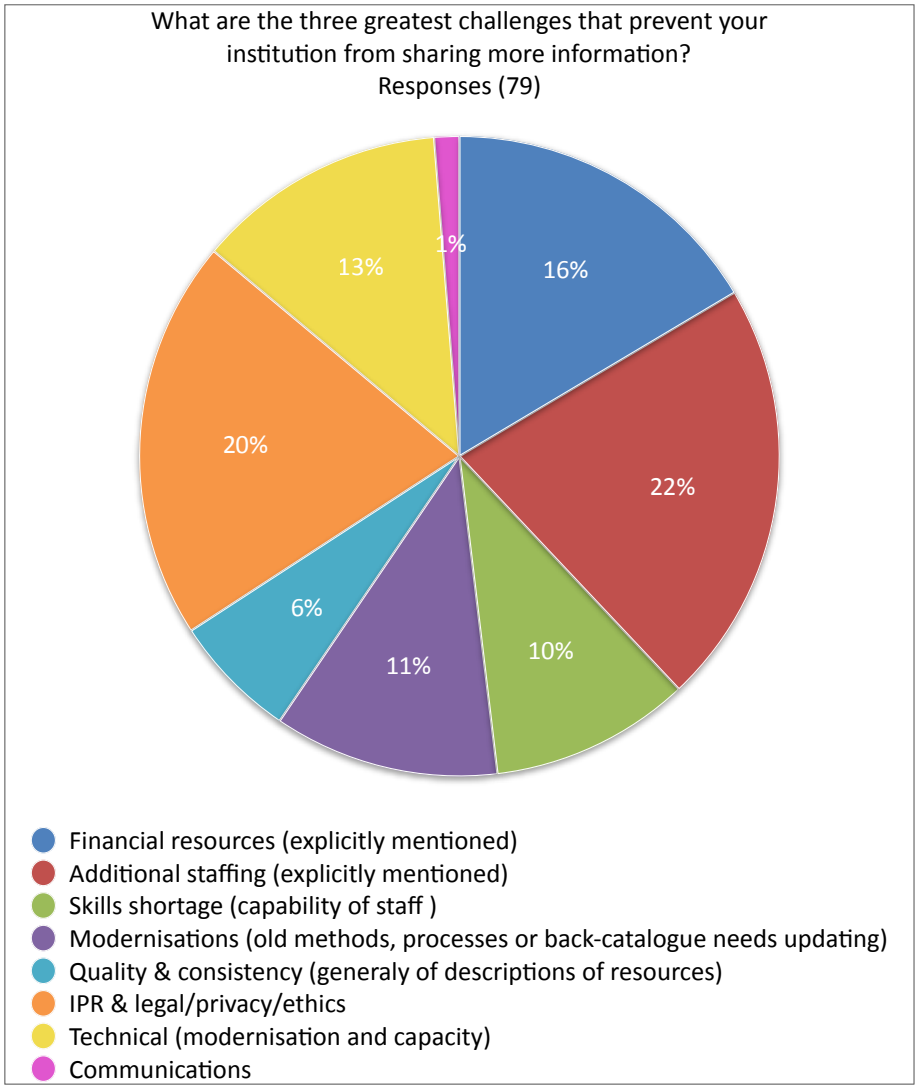

Figure 21: The greatest challenges that prevent survey respondents’ institutions from sharing more information

Cultural Barriers to Sharing

Survey respondents felt there was common ground between cultural heritage institutions in terms of shared goals across national and international contexts (see Figure 22).

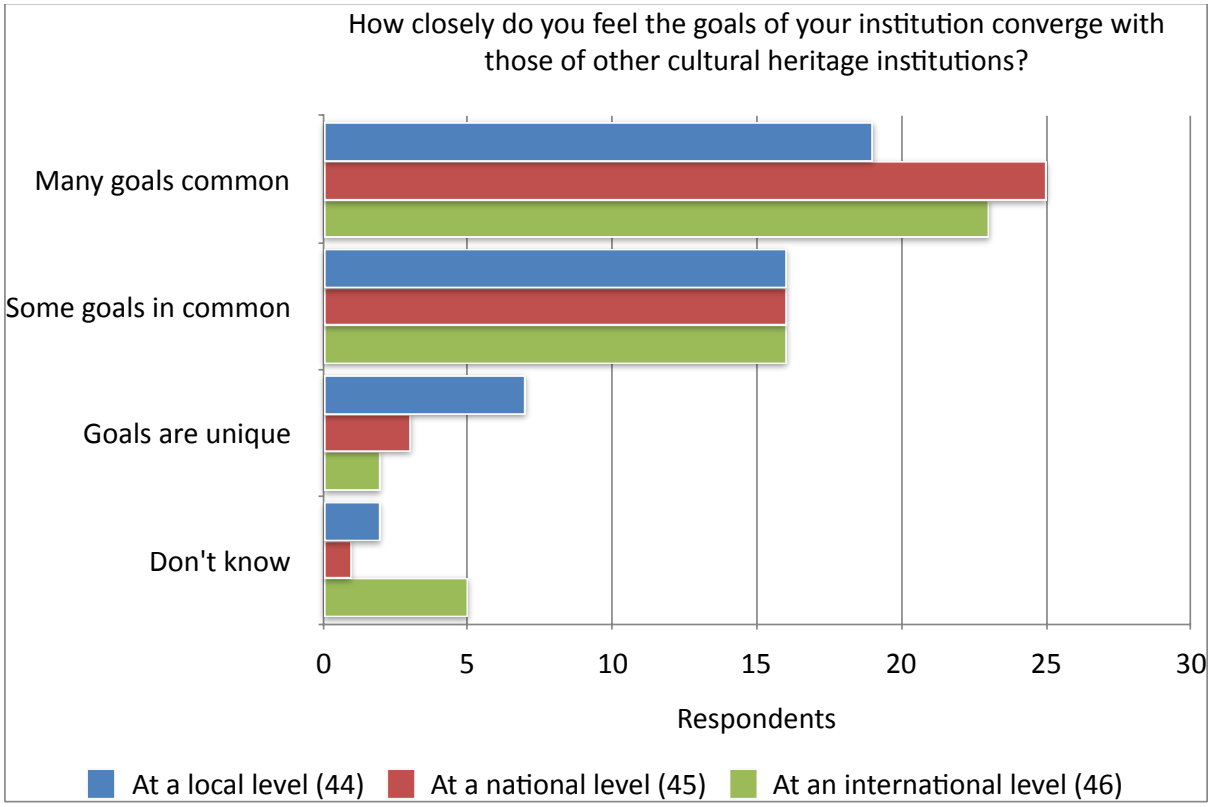

Figure 22: Survey respondents' feelings of common goals amongst cultural heritage institutions at local, national and international levels

Relationships with other institutions are integral to a functioning cultural heritage sector and practitioners were accustomed to collaborating as their collections were inevitably linked to those of other institutions:

> We don't take any audio-visual archives films, for example, because we don't have the material to conserve them. So, in this case we have collaboration with the [national] film archives … (I4)

Relationships with other institutions were fundamental to the operation of some institutions, which could not exist in isolation. As well as cultural heritage institutions, these networks included educational institutions:

> Sometimes we have specific collaboration with the different departments in the universities, and so we get support to students in the context of specific practical work. (I4)

Some networks were extremely localised. One participant described "good" relationships within the large city in which her archive was based but ones that were "not that deep" in the wider region, let alone the rest of the country or beyond. Connections could come down to the structure of the archivist community, as she explained:

> Quite a lot of archivists in [country] went to the Archival School in X. Right now, none of us here went there. So, we don't have these close connections that the people get there when they are in a class together. I get to know them in a conference or I get to know them from other things or projects or

> because we are friends, for example. But it's different, it's not that close. We didn't sit together in a class and try to solve the same problems. (I1)

A lack of conviviality could lead to superficial communication with other institutions, in one case limited to "twice a year they are getting a newsletter" (I1). On the whole, in-country relationships tended to be strong. There were also examples of effective international partnerships:

> There were co-operations in the past with the [A], with [B], and there are current excellent co-operations with several institutions. So, I mean, it's not as bleak as it might sound, but it does require, I mean, us to talk with a lot of people in order to understand what other people want. (I6)

Dedication was required to make these relationships productive and to fully grasp the overlapping interests of users of different institutions. Even where institutions were able to partner with others with more resources, they faced barriers to realising their technical ambitions. One participant expressed her disappointment in finding that the resource problems she encountered at her own institution were mirrored at institutions she had hoped to draw on after being incorporated in the same group of archives:

> … we thought that they would have tools with more possibilities, more modern tools, and that's absolutely not the case, so that was not an answer to our problem as institutions. And they have exactly the same problems, and so they are now developing a new tool for the collection, a new online interface, but they're also understaffed, and so they also have problems of money and available capable people to develop a new tool. And so, a long time, maybe, we hope now that in three years still … and three years is very long when you want to go on and to answer to the needs of your users. (I4)

Pooling resources had not offered any windfalls but practitioners nevertheless saw the value of collaboration. The common goal of digitisation often brought institutions into dialogue and this had generated networks where practitioners could learn from each other's experiences through regular contact via social media or Jisc lists. Institutions of the same type tended to collaborate most intensively and the weakest relationships were said to be across institution types, as this participant explained:

> … even though there was some collaboration, it wasn't as much as it probably should have done. Galleries talked to other galleries, libraries talked among themselves, museums, because they were operating on a different scale, just kept a watch out, an eye on everything that was going on, but really didn't have the time to do things. (I5)

Dialogue across institution types was not embedded in practice and seemed to be less valued. The reputation and approach of an institution was seen to be key to fruitful relationships:

> … meeting the people, knowing people, having like, some kind of a network does help you with trust, I mean, again and if you ask me a very

> unintimidating institution helps as well. I know other institutions you know … even if there are good intentions, tend sometimes to come across as … very large and very strong. The [A] is very different and it's very co-operative and that helps … I do believe that every institution has some kind of an image and some kind of a DNA which it operates. It can change with time, so it's not something that stays forever and forever but there are a lot of differences in the way that the institutions present themselves and work with each other and what they think about co-operation. (I6)

The absence of feelings of rivalry or competition therefore promoted productive relationships. It was felt that this could be a strength of the sector:

> As sort of cultural organisations in that space we share a lot of the same ethos. We … all report to the same government ministries so we're effectively in the same boat and I think all of these organisations have their own reputation. And it's just simply if you get organisations like the X Museum or Y Library together because we also share a past history then I think we both recognise that we deal with an organisation that's world leading in that field but is also non-commercial. And I think that's the critical part and effectively if you are sort of all arm's length government bodies who share the same ethos, who face similar challenges, all non-commercial and who are generally collaborative in nature that just sets the ground up very nicely. But the key bit is it's being collaborative because while we might maybe on occasion compete on the ground, overall there isn't really that competition in that sense. I think if government cuts happen they tend to hit all of us. It's I think probably rarely that one would suddenly get more money and the others would all be cut. So, we're not really sort of competing in that sense and the rest is then sort of normal personal relationship management. (I10)

It was therefore felt that a collaborative spirit should underpin cultural heritage practice. This was also said to depend on the characteristics of the institution and its field:

> … there are a lot of practices in the way our people are working that are different and it's also I think there's a bit of focused difference in that every museum will know it's a museum. We are a library but we're also a bit like a museum; we run exhibitions, we have sort of museum type spaces in here and therefore we are sort of working in different spaces or contexts that makes it a bit more harder to focus the organisation on something … It's a bit different with archives because I often find archives emphasise the preservation more than the access. I, partly also from personal experience, I generally find libraries a bit more open to letting people play with their stuff as opposed to archives. Although some of my archivist friends will tell me off and say that just is my skewed perspective … (I10)

Fora of collaboration were prized as opportunities "to understand how things are done in other places in order to build … a personal network of connections and that is a good start" (I6). Formal attempts to encourage collaboration were rare, however, and personal styles of working were often more influential, as in this account:

> The persona of the director has had a lot to do with it. Unfortunately, but it also, I mean, trickles down to middle management … and to the workers, it's important that we be open and to listen and to be able to communicate and to leave some prejudices aside and just listen. But again, I think the A is a good example because we do take part in a lot of international and national calls, we do try to listen. We are quite humble, I mean we don't believe that we can solve our problems for our self and that's why we do tend to co-operate quite a lot and … other institutions could learn from us. (I6)

Building good working relationships was said to be important at all levels:

> I think personal relations are important. You have to get to know the people … I think that's the first step … When you have [personal] contact: oh, you have this? Okay we can use it. So, we have something interesting for you … And it takes time, but I think the return is much… And it's a process and you have to build it … certainly to have some trust I think it's important. (I7)

Practitioners had little faith in technologies aimed at replicating this kind of rapport-building, based on experiences of infrastructure projects with no personal communication:

> It's difficult, it's not enough talking, there's not enough talking via Skype and sometimes if you say something during a meeting and even if these things are reported in the minutes, it doesn't mean that that thing that you're scared of to do, will be done. And this is a problem not of will in doing things, but it's a problem of dispersions between a lot of documents and you don't know where to turn. You can't remember exactly where that information is or it's a more complex problem, probably. The organisation of the information, the process of and sometimes also there are tools like management tools … they don't work. My experience is that they don't work, maybe they work for a very small group of people or the very, very well organised group of people but when you deal with a lot of people in not the same level, okay with different backgrounds, different way of working, they don't work. This kind of communication is not efficient ... (I8)

Personal contact was seen as very important and often a stumbling block for projects that tried to bring partners together without face-to-face relations. Where an institution had had little success forging connections on its own, however, a kick-start of contact was appreciated:

> … it's difficult because there is not the mentality, there is no… I can say that relationships of my institute with other institutions are at this moment limited. [infrastructure project] is a door open … (I8)

There were some differences in institutions' interpretations of accessibility, which nudged their approach to sharing knowledge away from a purely utilitarian stance. Ultimately, however, institutions were felt to share goals:

> So, there are differences between all these organisations, but I think in terms of the overall values, what we're here for, I think all of these organisations … would say they're here to preserve a, in some cases not insignificant, share of the nation's or global memory or cultural objects and artefacts, and to make them available. It's just the way, how we do these two things vary but in general and certainly from discussions I've had over the year I've been here I've never felt there was any clash of what we're here for. (I10)

Practitioners who had worked in partnership felt comfortable that values and goals were shared: "I think mostly we have a common view, and I think we want to do the same thing, in fact. And, that's important" (I2). Practitioners therefore sought to overcome challenges to sharing practice and data across institutions but it was felt that there was much work still to do to make collaboration the norm:

> I think they want to, I think there are a lot of fora … who bring together likeminded people, but we all work with different… we don't all work with different standards, and the international standards are getting more and more embedded in our work, but some countries are way ahead, others are far behind. I think in an ideal world we would all be working in the same way, but practically I think it's a matter of doing what we can at the moment. Thanks to [infrastructure project], we are evolving in a certain way, and we're also finding connections with other institutes, but before we are all there, it will take us … some time. We use EAD[15], other institutes are focussing on DDI[16]. We're … not all in agreement on how to get where we want to be. I think we're all in agreement on where we want to go but not on how to get there. That's still debatable. (I3)

The Challenges of Standardising Practice

There was a general consensus that practice across the cultural heritage sector was more standardised than in the past. Some practitioners felt that there was a limit to their ability to achieve the ideal level of parity of practice:

> Another thing that's maybe also important. We don't use, for example, standardised personal identification things … We don't co-operate here. It might be a problem in the future because they can't connect with us … Because, well, we started our descriptions in the 1950s and no-one thought about that. And if we would change here, we would have to go through all our descriptions until now and we would have to identify every person that was ever described or noted on descriptions. We can't do that, it's too much work. And also, we only have the names and nothing else that identifies the person. (I1)

[15] "Encoded Archival Description (EAD) is an XML standard for encoding archival finding aids" https://www.loc.gov/ead/

[16] "DDI is a free standard that can document and manage different stages in the research data lifecycle, such as conceptualization, collection, processing, distribution, discovery, and archiving." https://www.ddialliance.org/

Attaining uniformity across practice within an institution was an ongoing process for some and regulating this standardisation 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. (I7)

Even where commonly used practices had been adopted to the degree necessary for the inclusion of an institution's data in a sharing portal, some were still working to standardise metadata to ensure it complied with the requirements for full discoverability. Thus, there were many steps to the goal of standardisation. Facilitating computational methods was not the first priority, as this practitioner at a university's data repository explained:

> Primarily, I'm just concerned about making it available. I do work to make sure that it is more interoperable, meeting the I of FAIR in terms of making sure it's machine-process-able and that it's stored in a standard format that can be used by a wide range of tools. I do some work into encouraging the use of the main specific standards that are around that academics should be creating … using from the start of their research. I have looked at, at times, about trying to work with, say, a teaching course here or academics here to improve the quality of the data that we're making available, produce a new version that complies with an updated version of a standard that's out there or makes it easier to interact with it in some way. That hasn't quite come to … I haven't had time to do that just yet but maybe in the next few years.
> (I5)

Practitioners were responsive to the particular challenges of their holdings and some had developed their own tools to address common problems, such as a plugin for geographic co-ordinates that allowed depositors to select where their data was collected, "rather than manually entering co-ordinates, [which] means that it's a lot more accurate, we don't have the standardisation issues" (I5). Even where expertise and resources to advance practice were available though, standardisation could be superficial, with non-standardised "variable-level metadata" sitting below standardised collection descriptions. This 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. This participant felt fortunate to have:

> … the persons and the money to co-operate with something like [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. (I1)

Smaller institutions were therefore at risk of becoming marginalised as they drifted away from the orbit of standards used by better-resourced institutions. Larger institutions had their own challenges:

> 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 is 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, and if you look at boring but still useful things like mark standard for library metadata that doesn't seem to work that well for archival material. 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 system's not suitable or the system in the way how we are currently forced to use it not being suitable. The old metadata not quite right and yes, moving things from different domains effectively or making them useful across different domains is probably a better way of putting it. (I10)

Incorporating diversity as well as specialisation were therefore inherent problems of knowledge complexity. There were significant challenges across the range of institution size. Some institutions found that individuals' styles of working conflicted with the application of standard practice to fundamental tasks such as formatting dates, languages and the biographical content of item descriptions. Despite a theoretical commitment to standardising practices to enable knowledge sharing, the characteristics of institutions, their present staff and their history of practice could curtail such ambitions.

While standardisation of metadata may seem relatively unproblematic and indubitably offers many benefits, a wider view of standardisation, of 'linkages between this phenomenon and the broader conditions of modernity, bureaucratization, and globalization' posits 'that standardization cannot be understood outside of historical and political processes' (Duff and Harris, 2002: 281). Taking a view beyond the scope of KPLEX then suggests that standardising knowledge practices may *limit openness* by proscribing the legitimate jurisdiction of sharing so as to push back and harden the barriers of this walled garden, unless what is meant by 'sharing' and 'openness' precludes this.

Hidden Knowledge – How data that is not shared is at risk of disuse

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" (I3) when asked how far her role as an archivist could promote the use of her institution's collections. Legal barriers were encountered both at the national level and through working across national borders. At one institution, which had partnerships with others outside Europe, sharing was complicated:

> It is difficult because of the privacy regulations. We can't be sure, especially in America, that they will treat the documents with the kind of carefulness that we use … we're now trying to get A and B to become partners in our portal website, but the signing of the contract itself is very difficult because it refers to European privacy regulations and the new General Data Protection Regulation, so that's a big no-no. (I3)

Issues of privacy and copyright could also prevent sharing across infrastructures within Europe, facing institutions with a question of priorities:

> … these kinds of problems asked us to be able to make a choice between the collections, the metadata, which can be shared and the other ones and that took a lot of time. We weren't able to do that automatically, so these kinds of things, and it was totally impossible for us. So, for example, for [portal], to share metadata or to share documents with [portal]. It wasn't possible because of copyright issues or privacy issues … (I4)

Sometimes compliance with laws incentivised reduced sharing as to satisfy the requirements of both openness and some degree of non-disclosure proved too much of a stretch for an institution's resources. It was widely accepted that cultural heritage institutions 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. But with the new GDPR implementation, we have a privacy officer from … May, and so this person will be capable to make an analysis of all the problematics with the privacy issue in our institutions and with the linked data. (I4)

Legal expertise was therefore drawn on to judge the balance between open and closed that was best for the institution to strike. Keeping up with legalities was only part of this struggle, however, as the future of data-linking research was uncertain:

> Even if there's a numeric identifier that links to the original spreadsheet, even if we don't have that at a future date, we scramble that so it can't be traced back, because obviously it's a requirement of data protection. But it is a concern for future publication of datasets, because you don't know what is going to be made available in the future … 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 … (I5)

Adhering to legal and ethical principles cannot guarantee outcomes in this unknown landscape, which unsurprisingly led to data being held back out of fear of its potential uses. Conversely, some participants felt empowered to test the limits of legal restrictions where they thought this was ethical:

> 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. (I3)

Legal barriers to data sharing were therefore present but not as clearly defined as might be expected, as the intentions of regulations and the consequences of compliance could diverge with shifts in the knowledge environment. Legality was also raised as an issue in relation to data that may have commercial applications, particularly where collections spanned subjects used beyond the humanities:

> There is an issue that a lot of the… not all of the research is public funded, it is funded by charities that don't have a public remit, it's funded by philanthropic organisations that are working upon their own health focus but don't have a commitment to making this, the data, available. There is a balance needed in terms of the academic principles of publishing your research and making the data available, and also, say, obtaining a patent for some kind of new drug, and it can be difficult to balance for many academics. (I5)

Some practitioners were apprehensive about the potential for commercial interests to keep data out of the public realm. Despite these fears, participants were hopeful that the tension between opening access and protecting knowledge would be eased:

> The future of archival practice? Well, my hope is that I think we have… There's some kind of tension between privacy restrictions on the one hand … and legislation that is quite strict on controlling, destroying and these kind of access aspects. And then we have the other side, where we want to be open, transparent, and able to share. And my hope is that these two things will be in balance. So that we don't have to destroy … what could be very interesting as historical information and I hope that we will be able to keep that information. And, of course, the privacy aspect is very important, but my hope for the future is that these two things will be… That we get balance into that. (I9)

There was agreement that a balance between "the two tendencies" was needed, although practitioners felt that "the prevailing one is openness" (I6).

The idea of public knowledge being locked away because of commercial interests provided the starkest example of deliberately preventing the use of data that had high reuse potential. The obstruction of reuse more commonly played out in a less direct fashion.
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 were not dynamic enough to offer institutions capacity to fully progress their ambitions, from digitisation projects that had to be restricted to "the most important, or the most requested document", as it was "not really possible for everything" (I2), to the arrested development of tools, which

made it “not possible to do anything and so it's not sufficient, it’s not enough” (I4). 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 to 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 concretise what we would like to do, and that’s really the biggest problem. (I4)

Opportunities to reap significant returns were therefore missed for want of 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. Practitioners were forced to narrow their horizons, as this participant described:

> … the priority now here is to have tools here inside the institution, working correctly, that’s even not the case for the institution itself. So, it doesn't have any sense for us to begin to export metadata, to share metadata to aggregators, to other platforms, if we here don’t even have tools practical and able to answer to [internal needs] so that’s our first priority now. (I4)

Institutions were therefore taking the smallest of steps towards what they saw as the future knowledge landscape. The alternative was to over-stretch:

> … the regular staff is composed by eight/nine people. Then we have outsourcing people, volunteers, part-time personnel staff and for sure … the current staff is not enough for the … activities that we are doing. (I8)

Current working practices were often described as unsustainable as institutions struggled to keep up with changes in practice. A core concern was that progress could be squandered:

> I8: … the problem is to keep alive all these projects, the direction is very good, is right, but the problem is when project ends, and you don’t have the possibility to go on because there is no money anymore, or I don’t know. The European projects are typical from this point of view, no. After four years, eight years so the project ends and after we don’t know, but it was, originally it was a very good project …
> Interviewer: *So, do you see that as the biggest challenge to developing practice?*
> I8: Yes … So, you make a big effort to do things that when the project is finished became obsolete very quickly and you don’t have the possibility to go on with the work …

Even at comparatively well-resourced national institutions, a “still reasonably generous but constantly shrinking budget” meant getting used to taking their cue from what knowledge tended to be used previously:

> So far in the past we've bought everything we could on the chance someone might want to use it and now we're being a bit more focused on specific and also in relation to usage data. (I10

Technical Barriers

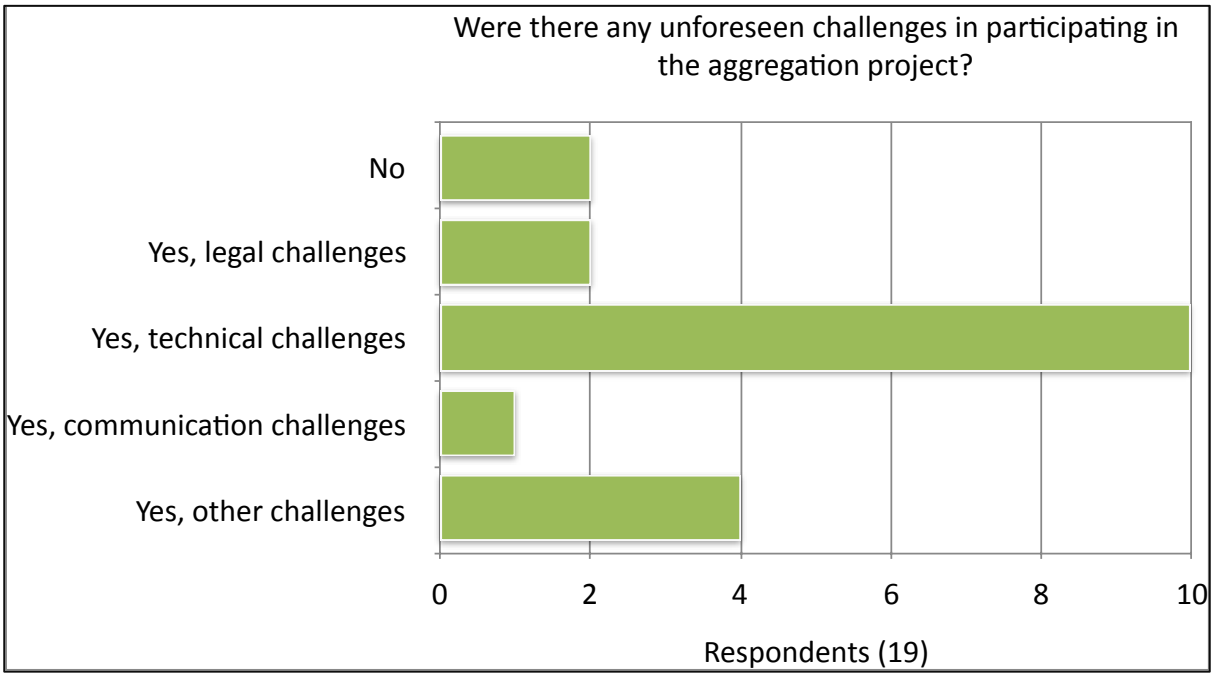

Figure 23: Types of unforeseen challenges in participating in aggregation projects, according to survey respondents

Participating in aggregation projects often presented unforeseen technical challenges (see Figure 23). While strained resources worked against knowledge sharing by relegating work beyond day-to-day operational functions to a low priority, these efforts could be directly undermined where they were contingent on technologies that did not fit smoothly into practice, as this archivist related:

> ... perhaps the biggest limit is that we have a lot of technical steps to publish finding aids, for example. And, sometimes, it doesn’t work because of technical problems, and I’m not a technical person, at all. So, it’s frustrating, sometimes, because it's only technical problems. That’s a limit. Sometimes, I would like to go the server room and make it work again. I think that’s a big limit. (I3)

Introducing new technology often created practical barriers to disseminating knowledge, as well as disrupting practitioners’ work methods. Sometimes, inappropriate technologies were adopted, as this archivist described:

> ... our system was built for in-house use and was never really considered as something which would be acceptable to the public. (I6)

Some participants had faced cultural challenges when they tried to implement changes to make their holdings more accessible, as in this practitioner's experience:

> I insisted we make available all this information on the website ... what we have done, what we do every day is useful for the researchers. So, this is the philosophy behind this project that I manage and support. And it was not easy, it was not easy because people, old generation stuff, in the institution where I work, not accept and doesn't accept this kind of philosophy of openness. And it's a matter of the mentality, it's a matter of cultural divide, professional divide and age divide and many elements that affect this ... But it was a fight, a little fight. But I won. (I8)

Opposition to change was sometimes *philosophical* but other participants gave examples of their colleagues' aversion to openness that were based on specific fears, as in the case of this archivist's plan to create a database of Holocaust victims:

> Interviewer: *Was that kind of an inevitable change?*
> I2: Not in the mind of older people in the institution ... it was not easy for everybody to understand that we will make databases with names, for example. Some people were afraid, perhaps, of the consequences of the databases. I have read some words like, it can be hacked and used against the Jews, or something like that, in fact. I think the databases frightened some people. (I2)

Convincing some colleagues sometimes involved demonstrating the value of new methods:

> I think that the head of the [memory institution] has agreed with the databases only when we made the [monument] ... And, we need all the databases to make such a [monument], so it became okay. But, before that, I'm not sure it was very clear for even the head of the [memory institution]. (I2)

Practitioners engaged in much unseen work to bring about incremental change within their institutions. When it came to collaboration, reaching out to colleagues at other institutions often exposed challenging differences in practice:

> [big data approaches are] very interesting because when I started, okay we can publish our data as linked open data. Okay, I said, it could be a very good opportunity in the future to interlink our data with those from other [related] institutions ... because of the enrichment of the knowledge on the [Holocaust] victims in general. But after four, nearly five years nothing happened in this direction. Because other institutions are not interested in opening their data. And so I felt frustrated from this point of view because my original ideas to give information coming directly from the format of publishing of data, didn't come ... And I think there are a lot of possibilities, a lot of opportunities for

> matching our data with those from other institutions that for now it's difficult. (I8)

This archivist acknowledged that cultural objections to sharing had a direct impact on her ability to lay the foundations for computational research methods to be applied to her institution's collections. Differences between institutions were not always understood by agencies that the cultural heritage sector was dependent on:

> ... the problem is, for example, we are now thinking about the possibility of having a valorisation tool to valorise the digitised collection of all these scientific institutions. And the fact is that the [national] government is ready to give us money to develop this tool, but the [national] government feels that we have only one tool for all the institutions. And the problem is that each institution has different needs, and each institution is at a different level in the tools that they already have. And then we don't ask the same things, we don't want the same things, and it's really difficult to collaborate on this kind of tool, for example, and there were many examples where we are not able to collaborate because we are too different. (I4)

Without an appreciation of these differences, attempts at encouraging collaboration in order to make efficiency savings risk creating a narcissism of small differences, dissuading institutions from finding common ground lest it overshadow their fundamental uniqueness. Another instance of institutions consciously taking the decision to not share was in the case of varying access requirements:

> Certainly, I can see benefits in a cross-institution support service and a cross-institution repository system, but our needs are quite specific in terms of the metadata that we have and the level of access that we provide. Many of the data that we have can't necessarily be made available to other researchers except in very secure environments, we'd have to guarantee that, and many of the academics here, they choose to deposit it in our repository rather than other systems … because it is internally managed and they feel that they have greater control over it. I think there may be some kind of resistance to … a shared service across different institutions. (I6)

In this case, resistance to sharing was borne out of what was seen as an existential threat to data acquisition. The origins of holdings are therefore an important factor when considering the positionality of institutions sharing 'their' data.

The Politics of Sharing

Differences between institutions were not always immutable. Some were felt to be imposed by external forces that stood in the way of "inventive" practice:

> … the key point is the money, and that's really our problem. I was talking with a colleague yesterday, and when we see the budget of the [national library], for example, because we have a lot of collaboration with [them], it doesn't have anything to compare with our budget. Proportionally, yes, they have

> bigger holdings, much more collections, much more documents, but even proportionally, if we divide their budget per document, we have a really small budget compared to these kind of institutions, and that's a problem of priority of our [national] government. (I4)

Funding is, of course, distributed according to national and regional boundaries, but there were other influential aspects of place that practitioners reported having an impact on their work. This archivist described power dynamics that frustrated her acquisitions:

> ... it's a bit political, because the district archives are a public archive. And, we have to sign an agreement with the head of the department in [X], who is elected. So, if the head of the department doesn't want to sign then we have to wait until the next one is elected! And, to see if it's not the same political side sometimes… it depends … Sometimes, they just want to keep the documents in the ... it's a public service, and they think it's only a public archive and they don't want to share with us because we're not a public archive, we're private … And, also, it's something to do with geography, because we are in [the capital], and, sometimes, those districts are very far from [the capital], and sometimes they might think if we give this archive to [X], people won't want to come and see the archive in our own reading room, perhaps. (I2)

Relationships with other institutions could then be compromised by political barriers that were detrimental to institutional missions and not conducive to fostering future collaboration. Maintaining physical control over holdings is perhaps the most easily identifiable way of precluding sharing. At some institutions, decisions were made to restrict access to certain users, with online data access limited to depositors and reading room access providing a mediated gateway for other users.

There was a general resignation to a slow pace of change:

> We would like really to go on with the possibility to link metadata and holdings with aggregator or databases, but I think it's not for today, it's not for tomorrow, it's maybe in two years, I would say, something like that, one or two years. (I4)

There was also a degree of trepidation among some participants, which tempered their desire to expand access:

> Interviewer: *do you have any reservations about expanding what you share and why?*
> I2: Sometimes, I'm afraid of the bad use that people can make with our documents, of course, because it's not hard to catch a document with no context and to say everything. But, I think if we can expose a lot of documents and make the research about all of the studies very vivid, I think the bad uses will decrease. So, we have to be clear and to share, because that is the only way to crush the [Holocaust deniers] (I2)

Barriers of a legal nature, those imposed through a lack of resources, technical expertise and cultural buy-in within institutions, had all contributed to a lower level of sharing knowledge than that which participants would otherwise support. While the potential for data to be left out of knowledge creation processes through an absence of sharing is relatively obvious, reservations expressed about the big data era brought to light another crucial risk to knowledge complexity: that data might become hidden as a consequence of certain types of sharing.

Jumping Google

Google was consistently referred to as a knowledge sharing platform that represented a threat to greater *openness*. For instance, the traditional hierarchical structure of collections through which contextual connections could be traced was reported to be losing significance as it was undermined by Google-style keyword searching:

> People are so adapted to the Google search that they don't even know anymore that there is a different way of searching. Therefore, we also chose to put in key words and categories that allow researchers to browse through collections based on key words and categories … That can give you a very interesting perspective as well because it can show you connections between collections and can show you different cases of the same thing happening, and it can help you find more proof for the point you want to make. It's another way of searching, but we feel that this is more connected to the way in which researchers today, especially the young generation, use search engines and tools that we developed in the past. (I3)

There was broad acknowledgement that the changing use of collections disrupted some of the fundamental tenets of archival cultural heritage preservation practice both through Google's presence as a rival source of knowledge and in the ubiquity of search engine use infiltrating archival research methods. Practitioners' support for integrating search engines into the research process was at least partly motivated by a concern that the *satisficing* return of a Google result might attain a hegemonic position that ultimately marginalised the role of the archive in knowledge creation. In this sense, mimicry was a less than sincere form of flattery. Concerted efforts were therefore made to stay abreast of developments in research practices beyond observations of institutions' own users:

> … the way in which researchers and students search for information is changing. Yesterday, I took a course on electronic records management, and the professor told us that nobody uses advanced search methods anymore, and I was shocked because I just put €5,000 into the development of an advanced search for the portal website. I was disappointed in that. But I think more and more people are looking for things that are available online. Unfortunately, we can't put everything online. (I3)

As well as the limitations to what data could be made searchable and available online, discussed in the previous section, there was principled resistance to fully aligning with the new paradigm:

> … I never use the search engine when I want to do a research. Because in the end you have this, and you see things, or maybe I can use that. And when you go the straight way, and even when the search is still methodological, then you miss things, because you don't look for other sources, which can be important. You have to get to know your material for good research… (I7)

> Rivalling Google was an ambition of many participants and a common motivation for joining aggregation projects was to counter the danger of sources being hidden by an unseen algorithm: "to be stronger than Google … I think this is the challenge for that kind of project. To be very attractive for people and push them to jump Google and go directly to your portal". (I8)

> It was felt that the influence of Google went beyond the user-friendly interface of the search box as the culture of search engine use was "like driving now with a Tesla, but with an [automated] one. So, you have to give in and the car drives. And for some things it can be good because people have more flaws, but on the other hand, can I trust a machine, because it was someone who programmed the machine, and I can't control on what basis the machine gives the result". (I7)

Looking to the near future, participants envisioned research methods becoming further removed from the researcher's hand as automated tools, machine learning and AI play an increased role. There was a fear that increasing data linking could identify previously anonymous data subjects, revealing data that had been 'hidden' for ethical reasons.

> Certainly, the machine processing is, I suppose, going to be even bigger soon than it is now, and … artificial intelligence has the potential to draw new conclusions from a large amount of data, particularly unstructured data, which is… until quite recent years have resisted the broader analysis. … if automated tools are able to make links between those datasets and then … infer conclusions about the people, if it's identified, then there's a significant danger to them. With all the way of what Facebook does and Google does in terms of linking information together, if that becomes even more prevalent, then there are dangers there in terms of providing data. (I6)

The risks of such computer-driven research were not thought to be purely ethical. In subordinating knowledge complexity to corporate interests, there was thought to be a danger of reducing the scope of research:

> … if Google can do better at discovery than we can then I mean by all means use Google. However, these organisations have commercial interest that we don't necessarily share and there're various biases built into neutral like looking technologies that will probably skew the perception of what knowledge is and introduce all sorts of bias into it. Even though people believe they see everything, they might see even less than before because they're only being shown the things that the algorithm believes they want to see. So, I'm really concerned with that increasing dominance of these organisations that commercial interests will increasingly drive knowledge

> creation … [the reproductive functions of filter bubbles] make Facebook richer but society drifts apart much more and I would like to avoid that … and not be stuck in something that looks a bit like Facebook large for cultural content. (I10)

Whether or not commercial organisations were allowed to set agendas for research, it was generally agreed that intervention would be necessary to maintain appropriate standards of rigour in incipient generations of researchers to allow them to maintain control over their own research and not fall victim to the blind spots created by a curated 'cultural Facebook'. One archivist was working on a project to help students in their thinking about methods:

> … so that you know, what are you looking for, what are the possibilities? Certainly, you have to know what you can't find, and it's one of my projects. I'm working on something to put a methodology used by historians into the search engine, to give a certain transparency in the search results, but also it has to be… It's one of the differences with IT. It has to be for example IT people say okay we want something Google, we want to have the result. I say no, it has to be correct, and the method has to be historically correct. But it's difficult to find something. But that's one of the projects I'm working on, to make it easier to connect with the university, again with the students, to warm them up for historical research. (I7)

The "quick wins" of Google's immediacy and familiarity were a constant thorn in the side of practitioners concerned with upholding rigour in research methods and this struggle was played out in differences in the visions of "IT" and archival colleagues. There was hope that, if adequately trained, researchers would use all tools at their disposal appropriately and ultimately apply a brake to Google-centric methods:

> I think the cultural heritage landscape will change and it will be hopefully in a more accessible direction, but I hope that the mentality of the researchers will also fling it back a bit to where we were twenty years ago, that you physically go to a place to look for something. I think students today don't realise how much there is hidden offline, I don't think they really understand. That's why [infrastructure project] is also very important, to learn what is out there; not online, physically. I hope that will be the direction. I'm afraid not, though. (I3)

There was a real fear that the spectre of openness was working as a diversion away from both the complex material excluded from it and any awareness that this phenomenon of hiddenness through eclipsing 'openness' was happening. It is clear that the new normal of the Google paradigm is having a direct effect on how researchers conceptualise their work. It was also suggested that changes brought about in this era would ultimately have a profound effect on the future of cultural heritage institutions:

> … even to the point where you wonder what of our current activity will humans still do in ten years from now and what will be… I mean how many human cataloguers will you still need if you have machine learning that's good enough that they can do a very sizable job? When will we get to that point, what will it mean for what these colleagues do? I'm not sure that

> understanding of how some of these things will potentially impact us is very sort of evenly shared across the library. We do have, partly to tackle this, there's a specialised team of digital curators. So, each of the larger curatorial teams have a digital curator assigned to them. They all form part of a separate team but they all work at one curatorial area to sort of raise awareness, train colleagues and make sure that they get better understanding of what potential is out there … 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 … There would still be enough work and I'm sure they could do something else really useful, but I think manage that transition in a way that we can do some innovation, provide good service, not let our staff down and yes, keep that commercial interest under control. (I10)

In being driven by users' expectations, this practitioner at a national library foresaw the foregrounding of technology necessitating a relegation of the embodied knowledge of some of his colleagues to an unknown, presumably less visible space in the workings of the institution. In this privileging of the *technical*, the 'clinicians' that Cook and Schwartz (2002) envisioned taking over archives are more likely to be those maintaining the technology than archivists themselves. Practitioners were motivated by the potential for technology to put researchers in control:

> If we get the collection descriptions online, we have… a description can never be neutral, but then researchers can read into them and make up their own minds. If they go into dialogue with us, my colleague and I, we have a certain way of thinking because we know the materials in a certain way. If we eliminate the staff member element in the equation, we will probably get different results. There's a lot to do, we won't be eliminated, but I think our role will also change once we put the descriptions online. (I3)

Even if the balance does not swing all the way from a staff of "historically correct" archivists to one of results-orientated clinicians, it was agreed that the growing presence of colleagues working from a position of computational thinking (Williamson, 2016) presented challenges for conventional practitioners.

Translation of Data Needs

Practitioners had embraced working with colleagues and partners with the expertise to realise their technological needs as they sought to expand the use of their collections. As discussed in the previous section, there was a clear division between roles, even where technical specialists were physically embedded in institutions. Following Latour and Callon, this study looked for instances where the interests of cultural heritage institutions might be *translated* into technical problems that technical specialists could then apply technological 'solutions' to. Latour (1987) described such a process of translation from the discourse of one world to another, followed by disciplined normalisation in order to stabilise the power

shift, as ‘intéressement’. Clarke and Fujimura (1992) describe how scientists define disciplinary boundaries while using the ‘dynamic interface’ of ‘standardised packages … to translate interests between social worlds’. Participants related their experiences of both dealing with standardised packages that required them to step out of their realm of familiar discourse and achieving some success in opening black boxes that stood between them and technical colleagues they felt essentially separate from. This division manifested as a challenge of communication whereby archival needs were not understood when “competencies” were related but not shared across the divide. Even where there may have been a shared vision, there was often a ‘language barrier’, as this participant explained:

> For an archivist, it’s not the same as for IT persons. I had major discussions with [project partners]. In the beginning, we were talking about the same thing but we were not understanding that the other person just had another word for it. (I3)

Superficial differences between the computational and the conventional cultural heritage perspective could be overcome. It was suggested that practitioners could even benefit from experiencing jarring approaches:

> [knowledge infrastructures] give archivists the opportunity to connect to different worlds: IT, information management, et cetera. Technology can help us streamline processes and can help make things findable, but they can’t do everything, and we have to put thought in it before you develop a system. I don’t know, I think of information systems and data management systems as more of a tool than as a solution for the problem. Tools in themselves cannot be the solution, they are the tools to get to a solution, but still people are on the base of thinking about a solution. We’ll still have to come up with the answers ourselves. I was actually very disappointed that it’s still necessary. (I3)

Casting a wide net of contacts could reaffirm cultural heritage practitioners’ place in solving problems of knowledge complexity, rather than allowing institutions to buy in a technological solution that usurped their role. Despite this realisation, such projects could still be initially unsettling to practitioners’ authority:

> It was very difficult, because in the beginning … I was thinking, why are you… why do I have to put my time into this, why do we have to change everything, that we’ve been doing the same thing for so many years, but… in the beginning I was very… and also, I didn’t understand why I had to lead the project, but now I understand that it’s because the people who were here longer even had a much more difficult time adapting to the change than me, and I had been here for two years. It was hard, but … 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 … I think stepping away from the fear, that's the most difficult part. Working around the fear and the resistance of the colleagues and of yourself, especially yourself, is very difficult. (I3)

Translating the terms of reference used by different communities of practice could then dissolve strongly held reservations based on misunderstandings about the benefits or disadvantages of sharing knowledge. The malleability of less experienced workers facilitated the uptake of new technologies but there was little mention of technical specialists' flexibility. There was a consensus that the onus was on cultural heritage practitioners to learn the language of technology. The danger that this translation could install a power shift towards technical solutions through interessément or the normalisation of standardised packages then looms as practitioners grow used to having their original goals displaced. Where technologies are a black box, displacing one goal with another that more closely fits the solution offered derails arguments that that method is not the best choice (Latour, 1987). This power disparity could take root as practice changed despite more utopian visions:

> Interviewer: *What are your thoughts about how the landscape of cultural heritage knowledge might change in the future?*
> I9: Well, I think that we have these divisions of library, archive… I mean, the clans and with all these specialisations, I think that will be more one: cultural heritage. One information landscape and information managing landscape. So, we will become more of one expertise rather than all these different layers or boxes.

Such a melding of professional identities carries the promise of smoother communication and a greater fluidity of practice but practitioners' experiences of the power dynamics at play when other specialists enter heritage spaces suggest that their perspective is the unique contribution most at risk of being squeezed out of the discourse. An awareness of a shift in the direction of technical control was behind some acts of resistance to the adoption of technological terms:

> … with IT and digitisation, you control things, and I was the expert in letting the people of Brussels think that I was the [gamekeeper rather than a poacher]. (I7)

Power struggles of this nature resulted in playing a game of apparent co-operation in place of genuine dialogue about points of disagreement. It was felt that in the long-term, building communication between those engaged in technical work and conventional practice based on mutual appreciation was the key to progress:

> … the IT personnel, half of them, there are only two, once worked in the Archive. So, we have the same level of communication. He knows what I mean when I try to explain or I don't have to explain how we make our descriptions or why an archive is different than for example the administration of a city or of an organisation. (I2)

Elsewhere, meaningful communication could be hit and miss and was more likely to be precipitated by meeting at the technical specialist's level. This participant perceived his

productive relationship with engineers – who he felt could contribute more if they were “more archivist” – to have emanated from his individual initiative:

> … I was a geek … but you have to know all the processes, the tools, and you don't have to program. I don't have the need to program them, but I'd want to know how they work and what the result is. I think [the archive’s tools created by engineers function well] but there is also a degree for improvement, I think. But I think researchers have to be aware, and that's one of the tasks of the technical things. (I7)

In this way, it was possible for cultural heritage specialists to strengthen their grip on the workings of the institution and having to adopt the vernacular of technology was a price worth paying. The idea of gatekeepers retaining control is also key to cultural heritage practitioners’ digital future, discussed in more depth in the next section.

Ruppert (2013: 3) avers that humanities and social science researchers should develop the technical skills required to engage with big data because otherwise, ‘the humanities and social sciences will become ever-more alienated from the creative power of software analytics in *formatting* their working practices’ [emphasis added]. This study suggests that the corollary for cultural heritage practitioners is the danger that their exclusion from the technologies opening up their collections may result in a *translation* of their complex knowledge of material to a purely *technical* understanding of its existing functions. Ruppert (2013) calls for a fine balance to be struck whereby humanists and social scientists emerge beyond discipline-discrete debates, not to become computer scientists but to ‘socialise’ the practices they have a stake in, and this study posits that cultural heritage practitioners should take a similar direction. Leading the conversation with technical specialists and putting their perspectives on their users’ needs in the foreground throughout should address practitioners’ feelings of being on the back foot in responding to technologies as they become available to them or having to take up resources that do not fit existing needs because alternative models have not been worked up. The historical record is a process, not a product and so this dialogue between knowledge complexity and the technologies that might be applied to it should be ongoing. Practitioners’ resistance to translation should not be seen as a refusal to communicate but rather a statement of their particular contribution, which positions them as ‘resting uncomfortably but content with that which is wild’ to their role (Star, 1990).

Some degree of translation is to be expected when skills are usefully combined, as we have seen in the earlier discussion of practitioners’ persuasion of donating researchers and institutional management. Indeed, some practitioners emphasised their desire to tip the balance in favour of their users. In the case of the examples above, however, practitioners experienced an asymmetric filtering from the technological towards them and this raises questions of who will control cultural heritage knowledge in the future.

The Digital Future of the Historical Record

In speculating on a future of openly shared public knowledge, practitioners were drawn to reflect on their institutions’ pasts, how far they had come in their mission to disseminate knowledge and the work that was still to be done. Some institutions had undergone

transformations that greatly expanded their capacity to share, including an archive that was "completely closed and it's a secret archive and nobody is allowed access" (I6). Nevertheless, many institutions were still in a transitional period that was characterised by the additional complexity of materials and users interacting according to both established and new conventions. This was the case at an archive that was in the process of changing its entire "back office and front office" online system, resulting in the current availability of "10%" of their holdings online and the rest still requiring a visit to the archive:

> You can access only the part of the documents that are tagged that it's okay to be online. It's a pyramid on the website, you will have only a small part of the description and the documents. In the reading room, you will have this box in addition. And, if you ask the librarian, you will get access to everything. (I2)

With what was available online representing the tip of the knowledge iceberg, there are clear implications for the material buried at deeper levels of accessibility during this transitional period, especially when a significant proportion of users' first contact is fully automated (see Figure 11). As one of the ways in which practitioners create value, 'that is, an order of value, by putting things in their proper place, by making place(s) for them' (Brothman, 1991), de-prioritising some knowledge by dint of elevating other knowledge over it has a lasting effect. Researchers will continue to engage with the historical record whatever state a gatekeeper institution's housekeeping is in, indeed their encounter with a particular institution may be brief, limited to this transitional snapshot. A decision to choose one collection for digitisation and online access over another therefore privileges that knowledge directly in that it increases its chances of being studied by researchers but also once that research begets further research, whose orbit is then further away from alternative sources. This process of transition is then a significant event in its own right, as decisions made along the way reveal 'tacit narratives' (Ketelaar, 2001) that become more deeply inscribed on material whose place in the historical record is reaffirmed or questioned by the new paradigm.

Of course, to attribute such powers to cultural heritage practitioners is not to suggest that they are exercising control over the entire process of knowledge creation according to their vision of material from before becomes part of the historical record to some predetermined end. Rather, their role is to bring order to 'inherently chaotic' (Cook and Schwartz, 2002) knowledge whose journey to them is beyond their control and influenced by interests whose priorities may shift, as in the case of a practitioner leading a project to open up "two kilometres of archives", which were currently inaccessible:

> But since the university wants to be open, transparent, more than they did in the past, and they are also proud of their history, right? So, this combination of things makes that they say, okay, well, we make sure that we have money for this project and that we have our history being part of the [national] heritage context in the City Archives … And since we're now preserving the archives since 1880, which is quite a long time ago, I think that you could argue that this cultural heritage side of things is not very… Not everyone in the university is very aware of that or has not been aware of that until a couple of years ago when we started with this project. So, you could say that that was hidden for a lot of years at the university. (I9)

Whether taking on previously hidden collections, incorporating newly produced material, or selecting items for digitisation and online access, practitioners will continue their work to represent the value of cultural heritage in line with the interests acting on their institutions. On one hand, the computational turn represents an opportunity to take data from complete obscurity to public knowledge. In practitioners' experience, however, the work left to do to achieve such openness meant that, for much of the knowledge they held, creating the possibility of 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 years, that will be the focus point, and I think it will take me five years. They asked me if it would be doable within [infrastructure's] second phase, and I said yes, but I think it might be a bit longer. That's our future. (I3)

In terms of predicting the digital future of the historical record, the key aspect of the discrepancy in expectations between infrastructure engineers and cultural heritage practitioners is how this gap is addressed. As discussed in the previous section, if engineers seek to close the gap by offering *technical solutions* that do not fit practitioners' ways of working and understanding their material, goals emanating from practitioners' vision of the future of cultural heritage will be displaced.

Another potential problem for those concerned with accurately representing the life cycle of cultural heritage data is "almost entirely hidden" nature of aspects of preservation, with metadata creation described as "something that few people care about" (I6) at a digital repository and a hidden, semi-voluntary, workforce helping to digitise holdings at an archive:

> They are not really volunteers, in the sense that they are placed here by the [national] social services. They're mostly refugees from Syria, from Iraq, and they are placed here so they can learn the language and they can integrate in a work environment and learn how teams operate, etc. That's more of a social project; they digitise. Our other volunteers, they are involved in indexing, creating name indexes. They also partially digitise documents, research files, or original documents from the wartime, and they are involved in cataloguing library books. They do a number of things. (I3)

The vital role of volunteers in plugging the gaps of under-funded cultural heritage work is widely recognised but the different aspects of casualised labour in this institution raises further questions about hidden stakeholders in the value of cultural heritage and the implications of societal shifts for the future of the historical record that deserve to be addressed in more depth. What is particularly interesting in this account is how an influx of resources in the form of free labour had not lightened the workload of the archivist:

> … to have everybody who works on [creating descriptions] do it in the same way, that's very challenging. At the moment, I do it, because I don't feel comfortable having interns working on them because everyone has his or her

> own way, work method. And you can see that even if I wrote the brochure for these interns so that they would all do it in the same way, and even though it's 12 pages for 20 fields, you can see there are differences between the fields. It's difficult. (I3)

The strength of professional identity is evident in the boundaries between these roles. It is insightful to contrast the authority exercised to maintain standards within the archivist's familiar dominion of describing collections with the relative deference to the transformative potential of *technical solutions* applied elsewhere. As a qualifier to the earlier discussion of the influence of (inadequate) resources on practice, it is especially noteworthy that a core concern like standardised description was placed beyond resource-driven decisions as practitioners prioritised maintaining their professional control. Consequently, it was widely acknowledged that such practice contributed to knowledge becoming isolated in individual practitioners:

> Interviewer: *You talked about people being quite specialised in their roles. Does that perhaps carry a risk that knowledge might get lost if people leave or if people move around?*
> I1: Oh, talk about it. We had some personnel change this year especially. And of course, there is a big, big danger that knowledge is lost. We try. We ask the people to document what they do when they go and also while they are working. But of course, if someone works with a special collection for about twenty years you can't write down everything. It's even worse when it comes to the photo collection because this colleague is very, very specialised also in the technical side of taking a photo. He takes photos on a nearly professional level. And so, I can't talk with him about details I don't understand. That's a real, real problem. But we don't have a better way right now. We try really to document things. Sometimes it works better than other times. It also depends on how people can structure their own work and also how well they can document or they can write it down. If they write cryptic notes, well, it's lost.

The phenomenon of embodied knowledge being lost was recognised by all participants and a lack of documentation of work was felt to lead inevitably to this labour, and its value, becoming hidden. This archivist suggested that even his colleagues:

> … don't know that 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 [X], twenty years ago, made the first list. Archivist A adapted it. It was colleague [X] who typed the inventory, read it and so on and so on. But the people who made the [finding] aid are not mentioned. So, I'm not mentioned. But every [particular language] inventory passes through me … (I7)

It was this kind of individual labour of activating embodied knowledge that was felt to the most hidden as its value was largely neglected and participants gave many examples of the continuing need for human contact alongside technological developments. This archivist described this synergy:

> We have both automatic contact and personal. It's mostly automatic for the first contact, because on our … institutional website, you have a [request] form … And then, when they are in the reading room, it's mostly human. They come to the desk and ask for help with their research … I think that they always have a contact with the desk and the archive, with our librarian. And, my colleague will give the context, in fact. We always explain a bit about the collection … [human contact is] very important (I2)

The "dialogue" through which practitioners helped to guide researchers continued to be highly valued while it may be augmented by technology. Participants spoke of a balance between automation and human labour, as in this example of archivists producing "packages" for users instead of offering a personalised service:

> The difficult part is to be fair and be able to offer the same thing to everyone who requests it. That is maybe one of the big limitations, because you can easily notice that when it's a very busy time, you have the sense of trying to do everything very quickly and standardise your requests, and sometimes you can't offer the same quality to everybody, and that's very frustrating for us sometimes. (I3)

The two elements of personal contact and technological expediency were said to support each other, with the feeling that "it's very human behind the portal" (CL) critical to an infrastructure's success and acceptance by the cultural heritage community. It was generally felt that supporting computational research methods required both technological tools and personal support and so, while practitioners' roles would not be "eliminated" in the foreseeable future, users' techno-self-sufficiency was predicted to foment significant change:

> It's scary. I've been here for seven years; my colleague has been here for twenty-three. The way in which we work will be different, but I don't know if it's necessarily a bad thing. I think it will give us more time for other tasks, to create more collection descriptions, also to create more access points … to be able to actively go out and obtain collections. It will give us the opportunity to refocus what we do, but still researchers will need us in the sense that interpreting a source will also need some kind of experience which we have and they don't. There will always be questions but there will be different questions probably. It doesn't really scare me at the moment. Ask me again in five years. That might be a different story. (I3)

A fundamental reorientation of the archive was therefore anticipated with some trepidation. Whatever the implications for job security, practitioners felt they were becoming more detached from the knowledge they presided over. Being "very highly digitised" was frustrating for some who "studied history and really [like] old paper" (I6) and relationships with the historical record were said to be changing as a result. This participant spoke about a sense of loss of contact with materials:

> Interviewer: *And does moving from one approach to another change your relationship with the data, with the collections?*

> I3: It does in the way that you are not working on item level anymore, you are trying to subtract the general meaning, the general line from a collection … For every new collection that comes in, you can't go in depth, reading every page in detail, you skim through and you seek the major subject. We're not as close to the items anymore, we're close to the collection as a whole.

Changing practice therefore carries risks of *skimming* over knowledge complexity to produce a simulacrum that represents less of an item's *deviation* from the collection in which it has been placed. In this way, differences between collections may become exaggerated as practitioners' 'closeness' reinforces the unique value and identity of a collection as the smallest unit in their purview, while the complexity that distinguishes the unique value of items may be hidden. The digital future of cultural heritage institutions raises questions about gatekeeping and control over more or less porous boundaries as the historical record becomes untethered from a physical location. This practitioner at a digital repository described how opening access to research data resulted in a weak relationship between the institution and its users:

> … the request form has a question about… what they intend to use the data for, we also ask them about their academic background or their research background. They can provide details about their expertise in the area. It is an optional field, not everyone completes it, but it was seen as important to actually give the data creators confidence that the people who are obtaining the data do have the necessary expertise to analyse it correctly, which is a big concern. (I6)

Even though practitioners were concerned about the final interpretation of data, digital discoverability meant that their control over the representation of complexity was exercised further upstream of users' interaction with it than if they had engaged with the institution in person. While a researcher might view avoiding a question on an online form expedient to answering her research questions, this displacement of the relationship between user and cultural heritage expert in which a *dialogue* is nurtured (and questions are less likely to be ignored) poses risks to the best interests of both the researcher and the historical record.

Opening and simplifying access to complexity to the extent of imitating the functionality of Google presents its own dangers of obfuscation, of which practitioners are well aware:

> Sometimes they of course compare us to big players like Google or something like that. 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] 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 realize that we might have more. (I1)

Some practitioners expressed concerns that a partial approach to digital discoverability would lead to a Matthew Effect (Merton, 1968), at least in the short term when only a very small percentage of their total collections were digital and an even smaller proportion directly available online:

> … 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. And that then leads to a process where that then attracts more funding and then more digitisation and improvement of metadata, and then it attracts even more visitors. (I10)

This account from a national library is indicative of how complex knowledge held alongside a huge number of other holdings is at risk of being marginalised, but it also highlights how valid and pressing concerns of inadequate resources across cultural heritage institutions are, as delays in sharing whole collections, let alone full holdings, push some material into researchers' consciousness while holding other material back, causing potentially irreparable damage to the public profile of some 'public' knowledge.

As a 'mediated and ever-changing construction' (Cook, 2001), the historical record cannot be an 'artefact with fixed boundaries of contents and contexts' (Ketelaar, 2001). Indeed, its fluid nature dictates that certain knowledge will be more prominent or more hidden at any given instance of its consultation. 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-contextualisation (Ketelaar, 2001). In the era of big data, this study suggests that the application of *computational thinking* to research methods and the archival practices of cultural heritage institutions represents a shift in the scope of re-contextualisation, as computational manipulation of the data representing knowledge complexity redefines the terrain of researchers' exploration. While the extreme of the unknown potential connections of data-linking is yet to be realised, practitioners' consciousness of what big data approaches could mean for the use of their material stimulated an increased concern to assert the essential, if sometimes hidden, elements of the process of the historical record that were part of their professional identity.

For some time, it has been seen as practitioners' responsibility to adapt cultural heritage practices to the computational turn by reasserting their role in inscribing meaning within the historical record (Nesmith, 2002). In 1998, Hedstrom urged archivists to 'teach the users of electronic archives how to be discriminating and sceptical consumers of digital information' so that they might 'approach digital evidence with a questioning mind about how it was generated, why it was preserved, and how it might be interpreted. Until most members of society feel as comfortable with electronic evidence as they do with traditional forms of documentation, archivists will have a responsibility to help users evaluate, understand, and interpret new documentary forms'. This study argues that that time has come. Moreover, the level of 'comfort' that users feel now threatens any headway that might have been made in the promotion of scepticism and it is now the responsibility of those designing the technologies to which researchers comfortably turn to reveal the blind spots they create.

While this study found that routinised techniques were less prominent than the application of individual practitioners' specialist skills (see Figure 15), the growth of routinisation and its 'embedded knowledge model' (Susskind and Susskind, 2015) across sectors certainly represents a threat to practitioners' control over how data is manipulated. Theimer (2016), drawing on Susskind and Susskind, points out that the conceit of routinisation also affects

practices that are non-routinisable by displacing them in favour of routinised practices, which necessarily offer different outcomes. While appearing not to cause significant disruption, this kind of routinisation in fact represents a subversion of the aims and identity of the profession it replaces.

There has been criticism of the absence of standards requiring archivists to document the methods at the core of their work and the thinking behind the decisions they make on a daily basis that leave an imprint on the historical record (Kaplan) and this study supports that provocation. When it comes to the assumptions, methods and decisions of data scientists, however, a naïve deference to the scientific transcendence of *technical solutions* still lingers. The digital future of the historical record depends on opening these black boxes so that practitioners' commitment to openness and relinquishing control of cultural heritage knowledge has the intended effect of transferring power to researchers' imaginations, not the unseen machinations of algorithms and their unaccountable programmers.

Conclusion – Hidden Data and the Historical Record

Hiddenness is part of the story of the making of the historical record. The recognition and inscription of values and meanings in cultural heritage are in never-ending dialogue and practitioners in archival institutions 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 knowledge necessarily re-inscribes the meanings given to it and value it appears to have for future use. This knowledge has already escaped the fate of all the 'lost data' hidden or destroyed according to law or left languishing in the 'grey area'. It may be deemed to fall on one side or the other of the tugging tension of cautious protection or transparent sharing, depending on the prevailing propensities of the era in which it was gathered.

Practitioners' crafting of narrative through their core responsibilities of creating descriptions of, and organising, material to preserve the context of a resource revealed how their *archaeological* knowledge of context, which is more complex than a continuous narrative of the origins and subsequent treat of a source, is critical to interpreting how it might be used in relation to other research sources. If practitioners' knowledge of context is no longer activated in the creation of descriptions or the organisation of knowledge this may impoverish researchers' understanding of potential uses of sources.

This study found a desire for balance in which the historical record was rich with data that could be discovered and used without moral panics. Institutions were responding to new data-scapes of research by looking at their capacity to support changing research methods and investments were being made in technologies it was hoped would better serve practitioners adapting their practice. The changing nature of researchers' contact with institutions presents new challenges for ensuring cultural heritage knowledge is used to the advantage, and for the advancement, of the world's knowledge, and appropriate training for both researchers and practitioners will ensure all can take an active role in this.

Supporting innovative research was regarded as vitally important there was consensus that this meant taking operational measures towards increasing both technical and personal support but some institutions continued to maintain barriers to sharing some of their data more widely, particularly where they catered to multiple user groups. Wherever an institution had progressed the sharing of only a portion of their holdings, there was a fear that the prominence of that knowledge cast a shadow over the rest. Furthermore, the path from analogue to digital to online was not assured for all holdings, as resources were finite and there was no justifying the disproportionate effort needed to accommodate the incorporation of the most complex knowledge. Eliminating hiddenness is not, therefore, a rising tide lifting all boats; rather it risks entrenching divisions that marginalise the most different institutional practices, users and material. Sometimes the technologies adopted were not fit for purpose, as evidenced by practitioners unable to find data using their institution's own search engines.

This study's findings demonstrate how strongly societal-level knowledge trends cross into the academic research domain, as in the ubiquity of Google-driven search habits. Aside from the interfaces users engaged with, much background work went unappreciated.
Human labour was found to be invaluable in unearthing the hard-to-find, as well as performing other tasks that may not be recognised as integral to cultural heritage institutions' operations. The low profile of voluntary labour was especially interesting, as was its role in demarcating the professional boundaries of practitioners. As computational thinking and archival thinking come to share a space, it would be illuminating to compare attitudes to labour and professional identity across their respective practices. Among cultural heritage practitioners, it was a common problem that knowledge was isolated in individuals and at risk of being lost.

There were also risks inherent in changing practice. As practitioners stepped away from the materiality of collections, knowledge complexity could be under-exposed and skimmed over. What was then exposed was in danger of generating a Matthew Effect (Merton, 1968) or enabling undesirable data linking. Practitioners were aware of these risks and some offered examples of how they had acted on this understanding to resist certain moves towards knowledge sharing. The struggle between *archival thinking* and *computational thinking* and the conceit of routinisation raises questions of who will control cultural heritage knowledge in the future, which is an obvious concern for practitioners. Instances of resistance to *translation* should be viewed in the context of practitioners' other acts in defence of professional and disciplinary boundaries, which suggests they may actually be relatively receptive to the application of technology to solve problems outside their expertise: they are not impervious to changes in practice steered by *computational thinking* per se. Most pertinently, practitioners' individual acts of resistance did not represent the greatest exigency for sharing knowledge.

This study found two kinds of hiddenness at work in practitioners' accounts:

Firstly, knowledge found to be hidden through institutions not pursuing opportunities for sharing could be attributed to the following factors: the use of metadata that resists sharing due to incompatibility with standards used by other institutions or aggregators with whom institutions might otherwise share; practitioners' desire to share knowledge being inhibited by an institutional culture that does not reflect this open outlook; a particular reticence about the possibilities of data-linking, which allows the sharing of some data, delays the sharing of

other data, which then suffers a comparatively reduced profile, engendering a downward spiral in use; a lack of resources, which leads institutions to focus on inward-looking, short-term priorities rather than address long-term goals such as sharing; and a lack of resources that leads to a focus on non-researcher user groups, marginalising knowledge complexity.

Secondly, knowledge found to be hidden through the consequences of sharing could be attributed to the use of methods and data that 'satisfice', leaving alternatives marginalised – this relates both to researchers and practitioners whose relationship with their material had changed; the richest (meta)data being the most likely collateral damage of standardisation for sharing, as standardisation privileges similarity and obscures difference; keyword search functionality that bypasses context, obviating the need for researchers to apply interpretive skills and obscuring the archaeological nature of the context out of which practitioners construct narrative metadata; an undue focus on researchers, excluding others who may play a valuable role in knowledge creation and dissemination.

Recommendations

ICT projects aimed at fostering increased sharing through data aggregation and infrastructures should provide long-term support to institutions to ensure developments do not stall, knowledge is not isolated in individual practitioners and technological obsolescence does not threaten progress, endanger data or discourage future participation in such projects.

Researchers should be supported to address any training needs for meaningfully discovering and engaging with data complexity at the point of access. Cultural heritage institutions have historically borne the weight of guiding researchers through their collections but the changing nature of researchers' contact with institutions, with self-guided use of technologies augmenting or replacing personal contact, presents new challenges for ensuring the optimal use of cultural heritage knowledge.

Cross-sectorial training should be considered to enable and encourage understanding between cultural heritage, ICT practitioners and researchers. This may better support research questions when the various stakeholders can contribute with mutual apprehension of the complexity of knowledge and the application of 'black-box' computational methods.

Institutions should be supported to introduce meaningful measurements of the use of their collections, to overturn the current norms of 1) an absence of data on what gets used; 2) an unmanageable blanket collection of this data, which institutions do not have the resources to analyse or 3) the collection of this data to monitor and support the popularity of collections, which risks becoming a tail that wags the dog. Instead, monitoring of use should be integrated into institutions' practices in a way that is meaningful and useful to them, for example helping to flag collections or data that are 'at risk' of becoming hidden.

Further research is required to deepen understandings of practitioners' fears about the possibilities of data linking – and to examine the validity of these concerns amid the uncertain future of the use of big data. It is entirely reasonable that practitioners are worried

about the potential for identifying and sensitive data to become public through data linking when it is not yet certain that current regulations and best practice preclude this.

Further research on institutional practices beyond the scope of this study is needed. This study has focused on European and North American styles of logocentric, text-based archival practice. Given the difficulties in rolling out commonly agreed standard practices in this sample, a study of the global cultural heritage knowledge environment is likely to uncover even further complexity, which should be combined with existing knowledge of the presentation, representation, and archival practice and management of cultural heritage digital objects sooner rather than later.

Bibliography

Berry, D.M. (2011). “The Computational Turn: thinking about the digital humanities,” Culture Machine, vol. 12, 2011, pp.1-22.

Bolin, G. and Schwarz, J.A. (2015) “Heuristics of the algorithm: Big Data, user interpretation and institutional translation,” Big Data & Society, vol. 4, July–Dec. 2015, pp. 1–12.

Borgman, C.L. (2015). “Big data, little data, no data: Scholarship in the networked world”. Cambridge, MA: MIT Press.

boyd, d. and Crawford, K. (2012) “Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon,” Information, Communication, & Society, vol. 15, May 2012, pp. 662-679.

Broadfield, A. (1946). The Philosophy of Classification, London: Grafton.

Brothman, B. (1991). “Orders of Values: Probing the Theoretical Terms of Archival Practice”, Archivaria 32.82.

Butler, J. “Preface (1999)”, in Gender Trouble: Feminism and the Subversion of Identity New York and London: Routledge pp. xiv–xv

Cook, T. “Archival science and postmodernism: new formulations for old concepts”, Archival Science. 1 (2001) 3–24.

Cook (2011) ‘We Are What We Keep; We Keep What We Are’: Archival Appraisal Past, Present and Future. Journal of the Society of Archivists. 32(2).

Cook, T. and Schwartz, J.M. (2002) Archives, Records, and Power: From (Postmodern) Theory to (Archival) Performance Archival Science 2: 171–185.

Clarke, A. and Fujimura, J.H. (eds.). 1(992) The Right Tools For the Job: At Work in Twentieth-Century Life Sciences. Princeton, NJ: Princeton University Press.

Crouch, D. The Perpetual Performance and Emergence of Heritage in Watson, S., Waterton, E. (2010). Culture, Heritage and Representation. London: Routledge.

Derrida, J. (1996) Archive Fever. Chicago and London: University of Chicago Press.

Duderstadt, J.J., Atkins, D.E., Seely Brown, J., Fox, M.A., Gomory, R.E., Hasselmo, N., Horn, P. et al. (2002). Preparing for the Revolution: Information Technology and the Future of the Research University, Washington, DC: National Academies Press.

Duff and Harris 2002 Stories and Names: Archival Description as Narrating Records and Constructing Meanings. Archival Science 2: 263-285, 2002.

Edwards, P.N., Jackson, S., Chalmers, J., Bowker, G., Borgman, C.L., Ribes, D., Burton, M. and Calvert, S. (2013) Knowledge Infrastructures: Intellectual Frameworks and Research Challenges, Ann Arbor: University of Michigan School of Information.

Foucault, Michel. 1969. The Archaeology of Knowledge. Trans. A. M. Sheridan Smith. London and New York: Routledge, 2002.

Fuller, S. (2010) “Humanity: The Always Already – or Never to be – Object of the Social Sciences?”, in The Social Sciences and Democracy, Ed. W. Bouwel, London: Palgrave, 2010.

Geoffrey C. Bowker and Susan Leigh Star, Sorting Things Out: Classification and its Consequences. Cambridge MA: The MIT Press.

Goffman, Erving (1967): On Face-Work. An Analysis of Ritual Elements in Social Interaction. New York: Doubleday.

Hedstrom, M. (1998) “How Do Archivists Make Electronic Archives Usable and Accessible?” Archives and Manuscripts 26(15).

Hess, C. and Ostrom, E. (2003) “Ideas, Artifacts, and Facilities: information as a common-pool resource,” School of Law, Duke University, vol. 66 Winter/Spring 2003.

Kaplan, E. (2001) Practicing Archives with a Postmodern Perspective. Paper presented in the seminar “Archives, Documentation and the Institutions of Social Memory”, organized by the Bentley Historical Library and the International Institute of the University of Michigan, 24 January 2001.

Karup, T. and Block, A. (2011). “Unfolding the Social: Quasi-actants, Virtual Theory, and the New Empiricism of Bruno Latour,” The Sociological Review, vol. 59, March 2011.

Ketelaar, E. (2001). Tacit Narratives: The Meanings of Archives. Archival Science 1: 131–141.

Kitchin, R. (2014a). The Big Data Revolution. London: SAGE.

Kitchin, R. (2014b) “Big Data, new epistemologies and paradigm shifts,” Big Data & Society, vol. 1, April–June 2014, pp. 1–12.

Latour, B. (1987). Science in Action: How to Follow Scientists and Engineers Through Society. Harvard University Press.

Latour, B. (1999) Pandora's Hope: essays on the reality of science studies. Cambridge: Harvard University Press.

Lave, J. and Wenger, E. (1991) Situated Learning: Legitimate Peripheral Participation, Cambridge: Cambridge University Press.

Manovich, L. (2011). Trending: The Promises and the Challenges of Big Social Data In Gold, M.K. (ed.) Debates in the Digital Humanities. Minnesota: University of Minnesota Press.

Manovich (2014) The Exceptional and the Everyday: 144 Hours in Kiev. 2014 IEEE International Conference on Big Data.

McPherson, T. Why Are the Digital Humanities So White? or Thinking the Histories of Race and Computation, Debates in the Digital Humanities, 2012. http://dhdebates.gc.cuny.edu/debates/text/29

Merton, Robert K. (1968). "The Matthew Effect in Science". Science. 159 (3810): 56–63. doi:10.1126/science.159.3810.56

Olson, D.R. 1994. The World on Paper. Cambridge: Cambridge University Press.

Robertson, H. Engines of Knowledge: the museum and the exhibit: https://discoversociety.org/2017/01/03/engines-of-knowledge-the-museum-and-the-exhibit/

Ruppert, E. (2013) 'Rethinking Empirical Social Sciences'. Dialogues in Human Geography. 3(3): 268-273.

Ruppert, E. et. al. (2015). Socialising Big Data: From concept to practice. CRESC Working Paper Series Working Paper No. 138

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. Language, 50, 696-735.

Star (1990) Power, technology and the phenomenology of conventions: on being allergic to onions. The Sociological Review. Volume: 38 issue: 1_suppl, page(s): 26-56 Issue published: May 1, 1990.

Susskind, R. and Susskind, D. (2015) The Future of the Professions: how technology will transform the work of human experts. Oxford: Oxford University Press.

Theimer, K. (2016) It's the end of the archival profession as we know it, and I feel fine. http://archivesnext.com/?cat=10

van Dijck, J. (2014). 'Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology'. Surveillance & Society 12(2): 197-208.

Vanden Daelen, V., Edmond, J., Links, P., Priddy, M., Reijnhoudt, L. et al. (2015) "Sustainable Digital Publishing of Archival Catalogues of Twentieth-Century History Archives," Open History: Sustainable digital publishing of archival catalogues of twentieth-century history archives. Brussels, Belgium, Dec. 2015. 〈hal-01281442v2〉

von Bertalanffy, L. (1949) Vom Molekül zur Organismenwelt: Grundfragen der modernen Biologie. Athenaion.

Weinberger, D. (2007). Everything Is Miscellaneous: The Power of the New Digital Disorder, New York: Holt.

Williamson, B. (2016) Political computational thinking: policy networks, digital governance and ‘learning to code’, Critical Policy Studies, 10:1, 39-58.

Appendices

Appendix 1: Interview Questions for CHI and Infrastructure Practitioners

Framing preamble – inquire into participants' background: have they worked as a researcher? etc., their frame of reference can then be drawn on in framing questions, where appropriate we can then use their terms or explicitly state that these are our terms and this is how we are using them (we might ask about fonds, record groupings etc.)
Background questions on CHI – size, specialism, funding engagement with (public?) users…

What is your working definition of (historical) [**data/holdings]**?
Probe: awareness of different users' definitions, in which direction do definitions travel? – from user to gatekeeper to system <>

What percentage of your holdings is?
Type of material % N/A
Text-based resources
…
Digitised – Borgman's categories?
'Born digital' data

Do you monitor the proportion of your [data/holdings] that get used and re-used by service users?

We're mainly interested in [data/holdings] that are used, or might be used, by researchers.
What proportion of your [data/holdings] is used by researchers?
Probe: does that mean the rest are not 'research objects'?

What proportion of your users are researchers?
Probe: how do they define 'researcher'?

How much engagement with solving researchers' problems do you tend to have?

How do (non-research) users' use of your service affect what decisions are made about the representation of [data/holdings]?
Probe: differences between GLAM in representation of data – aggregated but used differently by users looking to be informed (library approach)/inspired (gallery approach)? Museums and galleries tend towards flatter structure of metadata

Do you require accreditation of users? What does this involve? Timescale?

Do you provide information [metadata] about your holdings to an external [portal/ aggregator]? Why?

Probe: who does that serve? Levels of aggregation – local, national, international; internally/externally searchable catalogue/accessible holdings

Do you feel developments in aggregating information about research objects [**big data approaches]** apply more or less to your area of GLAM?

Have you seen an increase in data-driven research approaches?
Probe for digital archives: increase in big data approaches?
Do you feel able to support this kind of research? Do you provide services virtually? Have you been asked about access to data?
Do you have any reservations about it?

Can you talk me through the process of acquiring [data/holdings]?
Does [CHI] tend to be active or reactive in acquisition?
Is it more important to hold rare or representative [data/holdings]?
What determines whether [data/holdings] are considered of research interest?
Are you restricted by resource or time limits on the acquisition process?
How do you go about accepting and rejecting [collections/holdings]?
Probe: how are decisions made? Whose research interest is privileged? what's prioritised? how is value assigned?

Can you talk me through the process of making [data/holdings] publicly available – how do you prepare [data/holdings] for [FAIR] use?

Is standardising [metadata/descriptions of holdings] particularly challenging in your field?
Probe: names, dates (periods) and places can be imprecise. Is adequate reference info used that links all possible external descriptions in order to standardise for end user? Do you feel any *desirable* ambiguity/fluidity can be preserved within categorisation in a suitably (historically/socially) sensitive way?

What do you think might be the consequences of aggregating information about [data/holdings] for researchers concerned with the provenance of a resource?
Probe: are their users concerned about provenance? big data approaches have been said to render invisible or 'remove' 'human bias' (or tacit knowledge), which is central to historical inquiry. What does (increasingly) hidden bias mean for historical approaches to data?

To what extent do you look within your field for expertise and ideas to develop your own practice?
Probe: to what extent do they investigate practice in other fields?

How do you see your relationship with other CHIs/structures?
Probe: how visible are these relationships and what do they mean for CHIs in practice?

What are your experiences of sharing [data/holdings] across boundaries? How do you build and maintain trust at a distance?
Probe: infrastructures, aggregators, users

How does the operation of the service and its processes reflect typical structures of knowledge creation in the relevant discipline/field? RE-WORD
Probe: are structures hierarchical or team-based? is it collaborative/contingent work? Are practices gendered?

What are the limits of your role in terms of ensuring [data/holdings] are used, how does it relate to roles/data elsewhere?
Probe: how expansive is this vision, visions of connectedness, potential connections, breaks in chain? How might there be a better fit with the bigger picture?

Is maintaining the context of [data/holdings] within collections less important than it used to be?
Probe: is gatekeeper's role in keeping collections together – in context – more than administrative function?

Do you see yourself as a custodian of *public* knowledge?
Probe: thoughts about privatisation of knowledge – custodians as saviours of public knowledge

What are your experiences of changing the way material is organised? How are decisions about these changes made?
Probe: are actors' interests translated in order to enrol them to new practices?

How *active* is your role in leading/accommodating/adapting to change?

What have you seen change?

Who do you trust to lead change?

Is the organisation of material ever complete?
Probe: ideas of completeness – equilibrium in flow?

Do different approaches to managing [data/holdings] result in differences in your relationship with the [data/holdings]?
Probe: what are the dynamics of emotional relationship with material and how it is managed? how have new processes affected their relationship with their material? Effects of the internet, social media, trending topics?

Is there a division between 'technical' and 'emotional' practices in GLAM?

Do you think gender plays a role in how things are done in GLAM? At which level(s)?
Probe: positives and negatives of that

Are any aspects of your role 'hidden'?
Probe: elements of custodianship
Does knowledge ever get lost or become hidden between the cracks of the institution?
Probe: between colleagues with different pieces of the metadata puzzle?

What might the consequences of a move towards the inductive be? Or a perceived alignment with sciences or profit? TIGHTEN UP
Probe: CHIs no longer seen as geared towards enabling deductive research so use of resources for deductive purposes dwindles – more citizen researchers? – and the opportunity to pull researchers in to discovery of new data approaches is lost; or even backlash from rebel researchers operating outside of paradigm altogether if seen as linked to opposing interests, rejecting CHIs' services in favour of pure deductive approaches to keep methodologies alive? – CHIs pulled both ways?

How do IPR affect your work?

IPR – CATALOGUES HELD BY PUBLISHER ETC.

What are your thoughts about how the landscape of cultural heritage knowledge might change in the future?
Probe: how might the value of CHI's 'assets' change, be differently used or exploited? Is there an increasing threat of public goods being used for private profit?
Do you think strategic decisions are being taken at the right level?
How might decision making be improved?
Are there any developing practices to which you fundamentally object? Why?

Do you have any reservations about direct access to [data/holdings] from linked organisations?
Probe: how assured are you that privacy and research integrity can be maintained when datasets/collections are linked? Ownership, control, access.

What opportunities for broader use of data do big data approaches (e.g. visualisation) offer?
How do you feel about the potential of knowledge infrastructures to reinforce or redistribute authority, influence and power?
Probe: fears of being enrolled by discourses of data science and profit?

Do you think all colleagues in your community of practice subscribe to the same definitions and goals of CH institutions, practices and the relationships between them?

What common definitions do you accept?
Do you ever *resist* particular practices or terminology or other types of change? How?

What are your hopes/fears for the future?

How they communicate about their holdings, training on new practices – policies on these

Why isn't data *used* – not necessarily 'hidden' but not visible, validated e.g. by duplication in specialist archives? Obscured metadata and actions – Karup & Blok's quasi-actants erasing their traces – their *mediation* (hidden work, infrastructures)
To what extent do existing metadata already represent a big data approach to historical and cultural sources?

Are there any ways in which data use may be constrained by standardisation? (if standards are externally imposed)

Do you have any concerns about how big data approaches might change data use? Probe: reservations about offering up data to abuse of 'new empiricism', (whereas data-*driven* approaches might be more palatable). To what extent are data-*driven* approaches supported and pursued?

Appendix 2: Survey for Cultural Heritage Institution (CHI) Practitioners

How long have you worked in cultural heritage? [tick-box ranges]

1-5 years	5-10 years
10-15 years	More than 15 years

Country [drop-down list]

Job title, brief description of role [free text]

Organisation type [tick-boxes – allow multiple selections]

Archive	(national)	(regional/provincial/local)
Gallery	(national)	(regional/provincial/local)
Library	(national)	(regional/provincial/local)
Museum	(national)	(regional/provincial/local)
Research data archive	(national)	(regional/provincial/local)

Specialism of organisation [free text]

Please select the collection types that your institution holds

Text-based sources	Other artefacts and art works
Digitised resources	Data

What is your organisation's:

Annual budget [free text]

Number of (full-time equivalent) staff [free text]

Source of funding

National grant/subsidy	Private fund/trust
Regional/local grant/subsidy	Public/private partnership
Other (please specify)	

What relevant completed and ongoing training have you undertaken? [free text]

Is your CHI part of a formal/informal data sharing infrastructure? [tick all that apply]

At a local level	At a national level	At an international level
With an internally searchable catalogue	With an externally searchable catalogue	With remotely accessible holdings

Does your role cover… (or do you work with the role holder/is there a role)

	Part of my role	My work supports this role	This is a role in my CHI but my work is unrelated

Acquisition of collections			
Overview of collections whose descriptive data may be suitable for sharing through infrastructures			
Knowledge of processes of adhering to standards, mapping etc. in order to share data or descriptions of holdings			
Authority to share data or descriptions of holdings			
Interaction with users			

To what extent do you agree with the following statement?
In my organisation, some specialist knowledge is held by individual colleagues [sliding scale]

How would you describe your consumer (user) community?
Academic researchers
Student researchers
Businesses
Genealogists
School children
Other members of the public
[sliding scales Small % of users Significant % of users highest % of users]

How do you communicate information about your holdings? Are there policies for such communication?
To (potential/) users [free text; tick-box to indicate a related policy is in place]
To organisations that might want to share this information [free text; tick-box to indicate a related policy is in place]

Do you feel the goals of your institution converge with those of other institutions?
At a local level [sliding scale]
At a national level [sliding scale]
At an international level [sliding scale]

To what extent do you look within your field for expertise and ideas to develop your own practice? To what extent do you investigate practice in other fields?
[one sliding scale with extremes of 'within my field' and 'other fields']

How strongly do you agree with the following…

[sliding scale with extremes of ‘new practices adopted by my CHI are likely to be applied by those with specialised skills’ and ‘new practices tend to involve routinised techniques that may be employed with little prior knowledge’]

What standards do you follow? [free text]

To what extent do you feel engaged in a public duty to share data? [Likert scale 1-5]

Do you monitor what percentage of your holdings is used? [yes/no tick-boxes] How? [free text]

What percentage of your holdings is used?

Type of material	%	N/A
Text-based sources		
Non-text artefacts and art		
Digitised artefacts		
‘Born digital’ data		

How does the preservation system in your archive handle user access requests?

An automatic online query system	Human intervention
No access to AIP storage	Other [please specify]

What are the main methods of access to your collections? [free text]
What media do you use?

Online open access	FTP (file transfer protocol) networked access
Offline media (including post)	Reading room access only

Do you require accreditation of users? [free text]
What does this involve and how long does it take?

Free text	Days	Hours

How is physical access to your archive collections controlled? [free text]

What percentage of the information describing your holdings is available online to the general user? [sliding scale]

What methods can users of the archive use to request material from your archive? What finding aids are provided? [free text]

Do you provide information [metadata] about your holdings to an external portal/aggregator?
If yes: what’s the primary reason for this? [free text]

How familiar are you with such projects? [Likert scale 1-5]

How relevant to your institution do you think they are? [Likert scale 1-5]

What metadata (if any) does your institution create for inclusion in the AIP and DIP? [free text]

Does your institution comply with the general OAIS reference model logic?

Yes	No	Don't know

Do you find that areas of regulation with which your institution complies generally complement or conflict with each other?

Complement	Conflict

Can you give an example to illustrate your answer? [free text]

What are the three greatest resource barriers that prevent your institution from sharing more information? [free text]

What are (potential) assets/opportunities for greater sharing? [free text]

Branching questions from type of organisation = research data archive:

How well do you think (researchers' use of) Data Management Plans in your field address issues of…
re-use? [Likert scale 1-5] Sustained access? [Likert scale 1-5]

To what extent are you confident that your preservation planning/ the metadata you hold ensures the digital objects are independently understandable in the long term? [free text]

Do you have a disaster recovery policy? [free text]

How well equipped do you think your research data archive is for disaster recovery in the long term? [free text]

Are you familiar with FAIR access principles [yes/no tick-boxes]
If yes, what barriers to FAIR access do you think users of your holdings experience?
Findability [free text]
Access [free text]
Interoperability [free text]
Re-use [free text]