Digital Collections Audit
Kevin Gosling & Gordon McKenna (Collections Trust)
Adrian Cooper (Intelligent Heritage)

TABLE OF CONTENTS

Front Cover image: Mary Fielden and others presenting petition in 1909, TWL-2009-02-214. LSE Library/FlickrCommons

Executive summary and key recommendations

Between September 2021 and the end of January 2022, Collections Trust carried out an audit of digital collections in the UK's leading cultural heritage institutions. The work was commissioned by the AHRC-funded research programme *Towards a National Collection* (TaNC). The aim of the audit was to understand the number, scale and attributes of digitally-accessible collections across the UK cultural heritage sector that might form part of a future national digital collection infrastructure.

Scope and methodology

The scope and methodology of the audit were established in an initial phase of the project between June and September 2021. On the basis of this preliminary work, it was agreed that the study would comprise three elements:

- A survey-based audit of: the collection-holding Independent Research Organisations (IROs); museums within the National Museum Directors' Council and/or Arts Council England's National Portfolio; the museums, libraries and archives within England's Designated collections and Scotland's Recognised collections; members of Research Libraries UK; other relevant national bodies within Scotland, Wales and Northern Ireland; some other major digital collections, such as the BBC Archives and Portable Antiquities Scheme. In all, this cohort included 264 cultural heritage institutions.

- Structured conversations with relevant staff from IROs and some other particularly large institutions, to gather technical information of the kind needed to inform future data-sharing initiatives.

- Seven roundtable discussions with representatives of smaller museums and archives to gain qualitative insights into their capacity to join any future national digital collection.

Main findings of the audit survey

Of the 264 cultural heritage institutions invited to take part, responses were received from 230 (87%). No previous attempt to survey the state of the nation's digital collections has come close to this response rate. The proactive, personalised approach needed to achieve this level of engagement was time-consuming, but the result is a dataset that TaNC can explore from many angles beyond the scope of this summary report. The researchers would like to thank all those who took the time and trouble to provide information.

The headline findings of the audit survey are set out below. While the summary report breaks down these headline figures by type of institution and by home nation, the usefulness of the latter is impacted by the fact that the National Library of Scotland did not engage with the audit process.

Item-level records

- The total number of item-level records reported is nearly **146m**. All but three of the 230 institutions reported that they hold records down to this level (with some caveats and clarifications, particularly from archives). The British Library (BL), The National Archives (TNA), BBC Archives and Natural History Museum (NHM) together account for just over half (51%) of this total.

- A quarter of these item-level records, **37m** in all, have associated images or other digital media. TNA, NHM, BL, the British Museum (BM) and one other IRO together account for 52% of these.

- Extrapolating from (admittedly subjective) estimates given by respondents, perhaps **28m – 30m** of the 37m records noted above (73% - 82%) have images and other media that are of publication or broadcast quality. Around 54% of these are held by TNA, NHM, BL and one other IRO.

Group-level records

- Group-level records are held by 188 of the 230 institutions (82%). Not surprisingly, far fewer group records were reported than item-level ones: just under **8m**. BL, NHM, the National Library of Wales (NLW), Bradford Museums and Galleries, and another IRO between them hold 54% of all group records reported.

- Taking the 146m item-level records and 8m group-level records together gives a total of around **154m** collection records within internal systems.

Records published on institutions' own websites

- Over **82%** of the 230 respondents (190) said they publish at least some records on their *own* website. Almost **98m** records are available via institutions' own websites, around 64% of the 154m records counted together above. BL alone has almost a quarter (24%) of these, while BL, TNA, NLW and NHM together account for 53%.

- Nearly 25m of the 98m records on institutions' own websites (26%) have images or other digital media. TNA has 36% of these; TNA, NHM and BL together have 55%.

Records only published on third-party platforms

- **Forty** of the 230 institutions (17%) *only* publish records online via third-party aggregators and other platforms (that is, not on their own websites). They include some national institutions (such as English Heritage, Historic Royal Palaces, National Trust for Scotland), and large civic services (including Glasgow City Archives, Museums and Galleries Edinburgh, Leeds Museums and Archives, and Nottingham City Museums and Galleries) as well as smaller institutions of the kind one might expect.

- A further **67** institutions reported that some of their records were *only* published on one or more third-party platforms. The researchers queried and corrected a few instances where answers given here were clearly wrong, but suspect that some other respondents may have double-counted records that are *also* published on their own websites.

- With that caveat, the total number of records published *only* on third-party sites is **3.7m**, of which **3.4m** have associated images or other digital media. Sites mentioned by at least five respondents include: Art UK (67 respondents); TNA's Discovery (11); Archives Hub (8); Find My Past (8); Bridgeman Images (7); Google Arts & Culture (7); Ancestry (5); and Flickr (5)

Policies on using images and other material for which institutions own the copyright

- Complementing more detailed studies commissioned by TaNC, respondents were simply asked to select one or more of several options to reflect their general policy on the use of images, etc, *for which their institution owned the copyright*.

- Only four respondents gave 'unrestricted' as their only answer to this question. A further nine included 'unrestricted' in answers that also mentioned one or more of the other options given.

- 180 respondents (78%) included 'non-commercial' in their answer, of which 45 (almost 20%) gave this as the only option. Most others typically combined a policy of free non-commercial use with charging for commercial use.

- Only 16 respondents gave 'commercial' as their only answer to this question

Records available only behind a paywall

- **Fourteen** institutions said that some of their records were *only* available behind a paywall. Around **5.3m** such records were reported, 98% from just four institutions: TNA (42%), Royal Artillery Museum (28%), The Box Plymouth (Archives) (16%). Note that the TNA figures include material that would normally require paid-for access, but is currently available for free during the Covid pandemic.

- As well as TNA's own Discovery portal, other paywall sites mentioned include Find My Past, Ancestry, and The Ogilby Muster, a new portal for regimental museums from the Army Museums Ogilby Trust.

Application programming interfaces (APIs)

- Asked whether their institution had 'an API that allows others to make use of your online collections', **41** respondents (21%) said they did. Assuming that institutions with APIs have made *all* the records on their websites available through them, up to **61.5m** records might already be available for re-use this way.

- A further **36** respondents (16%) said that the introduction of an API was pending. Making the same assumption noted above, it is possible that a further **13.5m** records will become re-usable once these pending APIs are implemented.

- Almost **half** (49%) of respondents said they did not have APIs, while the rest did not know.

IIIF-enabled images

- Complementing the work of the TaNC Foundation Project *Practical Applications of IIIF*, the audit survey asked whether institutions had any IIIF-enabled images. An encouraging **39** (17%) reported that they did, while a further **29** (13%) said it was pending.

- While not all the institutions that said they had some IIIF-enabled images provided even estimated numbers, at least **49m** such *images* were reported. The Wellcome Collection alone reported 44m of these (89%), relating to 367,000 online catalogue records. NHM has 5% of IIIF images reported, with Durham University Library, BL and Cambridge University Library all accounting for about 1% each.

External persistent identifiers

- To complement another of TaNC's Foundation Projects, *Persistent Identifiers*, the audit survey asked whether institutions made use of external identifiers to help minimise the problem of broken links. Only **14** respondents (6%) said they did, with a further 33 (14%) saying it was pending. The great majority said they either did not use external identifiers (57%) or did not know (23%).

Main findings of the technical interviews

To take a deeper dive into the kind of technical information needed when planning to share collections, Adrian Cooper of Intelligent Heritage conducted a total of 27 structured conversations with the relevant staff of 18 IROs and nine other cultural heritage institutions. Together these 27 institutions account for around 70% of all the online records, images and other digital media reported in the main survey.

- The IROs were: Amgueddfa Cymru (National Museum Wales); British Film Institute Archive; British Library; British Museum; Historic Environment Scotland; Historic Royal Palaces; Imperial War Museums; Museum of London Archaeology (MOLA); The National Archives; National Gallery; National Museums Liverpool; National Museums Scotland; National Portrait Gallery; National Trust; Natural History Museum; Royal Botanic Gardens Kew; Science Museum Group; and Tate. Given that the information covered often required input from several people, it was not possible to schedule interviews with all the IROs as hoped.

- The other institutions who took part in this technical supplement to the main audit were: Historic Environment Division, Northern Ireland; National Galleries of Scotland; National Library of Wales; National Museums NI; National Trust for Scotland; Public Record Office of Northern Ireland; Royal Collection Trust; Royal Commission on the Ancient and Historical Monuments of Wales; and University of St Andrews – Library.

Metadata

- There are huge variations in the structure and semantics for object-based **descriptive metadata**, especially in museums. This inhibits the ability for institutions to share and exchange information easily.

- 19 institutions stated that they routinely add embedded **technical metadata** to digital resources as part of their delivery pipelines (in some cases, along with descriptive metadata such as object title, maker, etc).

- Few institutions reported that they include any **administrative metadata** (details of asset organisation, source, ownership, etc) as part of their normal digital resource delivery. As records travel beyond current organisational boundaries and/or are made available for machine-based enquiry via APIs, such core administrative data will become increasingly vital.

- As reflected in the main audit survey, very few institutions are using third party, independent services to provide unique, **persistent identifiers** for catalogue records or digital assets. Most tend to generate identifiers (URLs) based on internal system IDs. These are open to potential change over time which would result in broken links for any dependent resources.

- Many institutions are dissatisfied with current systems and processes, especially **legacy systems** without APIs, causing problems integrating with other systems and services. Many were either in the process of, or planning, the move to a new collections management or digital asset management system (DAMS), a state of flux that will last several years.

Asset management and delivery

- 23 institutions are using a dedicated **DAMS** to provide enterprise level management of their digital resources, but only six are using in-built functionality to automate the transformation

of digital assets into different formats and sizes for public or commercial purposes, the rest using bespoke tools and services.

- Only four institutions are actively using or connecting their DAMS to a dedicated **digital preservation** system.

- **Delivery** of digital assets relies on the ability to integrate catalogue data (normally held in a CMS) and technical metadata about the digital assets (held in a DAMS). Ten institutions use a middleware solution to do this, three said the integration was entirely manual, and the rest had commissioned or developed bespoke integrations.

- 15 institutions reported that the processes to deliver digital assets online are fully automated, consistent and repeatable. For institutions already using IIIF, the generation of manifests was either dynamic or fully automated.

Access and availability

- Catalogue records and related digital resources that are approved for online delivery are usually published (copied) to a dedicated webspace hosted by the institution and all public users have the same access level.

- The British Library is the only institution interviewed that uses a controlled API to provide different types and levels of access to partner organisations.

- Six of the institutions interviewed have an Open API to allow third parties to undertake machine-based enquiry or integration with external systems. However, the existing Open APIs have low usage levels and some struggle to perform or stall completely when attempting higher-volume queries for collaborative research projects.

Rights management

- Most of the institutions assign licences and usage rights to individual digital assets, starting with records in the related catalogue systems. These are then cascaded to digital assets (held in separate DAMS). This workflow is often manual and time consuming.

- Some use bespoke scripts and pipelines to apply a range of licence types based on publishing status flags and/or other logical parameters applied to the catalogue data. Only seven stated they could assign rights to group level records in order to automate and speed up the rights management process.

Main findings of the roundtable discussions

In a third strand of activity, seven roundtable discussions were convened with people from a range of smaller museums and archives to gain qualitative insights into their capacity to join any future national digital collection. As even a cursory glance of the roundtable quotations makes clear, a few key themes dominated most of the discussions.

- Time and money are scarcer than ever, with documentation low on most people's 'to do' list.

- Many staff and volunteers struggle to access and use their own data due to the limitations of legacy collections management systems.

- Poorly-documented and hard-to-manage images are a big problem, but often the most useful and potentially most used part of local collections.

- Real, practical help from national institutions such as Art UK and TNA's Manage Your Collections is hugely welcome.

- Small institutions want to benefit from cutting-edge technologies, such as AI-based services and managed digital preservation solutions, too.

- Institutions are not making the best use of existing infrastructure such as TNA's *Discovery*, missing a trick by leaving out even basic collection-level descriptions that might lead researchers to them.

- There is confusion and some cynicism about TaNC and the concept of a 'national collection', but this could be turned around with a change of emphasis on the genuine, practical benefits collaboration at national scale might bring all institutions.

Conclusions and recommendations

This final section draws a few headline conclusions and, where it seemed useful, recommendations to guide future planning and action.

The UK's digital collections are concentrated in a few institutions

Although the audit survey revealed some impressively large numbers of records and digital assets, a consistent pattern emerges across several survey results: that around 10% of the surveyed cultural heritage institutions account for around 80% of the records and assets, both online and offline. Often, a handful of mega-collections hold more than half the total.

The concept of a 'national collection' is problematic

The roundtables with smaller institutions revealed the phrase 'national collection' to be deeply problematic, even when tempered with 'towards'. It sets up expectations that may be hard to fulfil and prompts questions about which institutions are in the 'national collection' and which are not.

It is recommended that, in scoping what might happen after TaNC, the unhelpful connotations of the phrase 'national collection' are avoided.

Few institutions use persistent identifiers

Both the audit survey and technical survey echo the findings of TaNC's Foundation Project on *Persistent Identifiers* (PIDs): few institutions are using PIDs, or even know what they are. This situation should set alarm bells ringing, as there is a real risk that the millions of online records reported to this audit will, over time, turn into millions of broken links.

The technical survey showed that adopting external identifiers was a low priority even amongst the IROs and other large institutions. There seems even less chance that smaller institutions will give this matter any thought at all.

It is therefore recommended that, as a priority, a sector-level strategy should be developed for assigning and supporting PIDs, taking the onus off individual institutions.

Support at scale helps institutions share their data

In the roundtable discussions with smaller institutions, data-sharing *per se* came across as a low priority, particularly if it involves extra work for no perceived benefit. Frequently mentioned services like Art UK and TNA's Manage Your Collections and Discovery offer not only a route online, but also practical help. The barrier to taking part is low, and the incentive to do so is clear. Among other things, services like these offer a potential solution to the low take-up of PIDs by individual institutions noted above.

As AHRC scopes what might come after TaNC, it should engage with the problem of helping the majority of institutions share their data, working in partnership with existing services that already do this and, where needed, with new ones.

Collection-level records would be a useful starting point for many

Much of the discussion in the roundtables centred on the perceived impossibility of ever having enough resources to work on collection catalogues to the point where they could be published online. However, if AHRC's main aim is to make collections more accessible to researchers, good collection-level descriptions would be better than no online information at all.

It is recommended that TaNC supports a wider sector drive to create (or update) a useful collection-level record for every institution, especially those represented only by minimal listings in existing databases, and to re-purpose legacy datasets from past projects.

1. Introduction

Between September 2021 and the end of January 2022, Collections Trust carried out an audit of digital collections in the UK's leading cultural heritage institutions. The work was commissioned by the AHRC-funded research programme *Towards a National Collection* (TaNC). The aim of the audit was to understand the number, scale and attributes of digitally-accessible collections across the UK cultural heritage sector that might form part of a future national digital collection infrastructure.

1.1 Scope and methodology

The scope and methodology of the audit were established in an initial phase of the project between June and September 2021. On the basis of this preliminary work, it was agreed that the study would comprise three elements:

- A survey-based audit of:

    - the collection-holding Independent Research Organisations (IROs);

    - museums within the National Museum Directors' Council and/or Arts Council England's National Portfolio;

    - the museums, libraries and archives within England's Designated collections and Scotland's Recognised collections;

    - members of Research Libraries UK;

    - other relevant national bodies within Scotland, Wales and Northern Ireland;

    - some other major digital collections, such as the BBC Archives and Portable Antiquities Scheme.

    In all, this cohort included **264** cultural heritage institutions.

- Structured conversations with relevant staff from IROs and some other particularly large institutions to gather technical information of the kind needed to inform future data-sharing initiatives.

- Roundtable discussions with representatives of smaller museums and archives to gain qualitative insights into their capacity to join any future national digital collection.

1.2 Outputs

In order to increase the likelihood of institutions agreeing to provide information for the audit, they were not asked to agree to their individual responses being published, only made available for restricted use by TaNC. This means that the present report is just the public summary of a wealth of detail available to TaNC in a full dataset comprising:

- An Excel workbook with all the data submitted except longer texts that made the spreadsheets unwieldy. The file also has the working copies of the tables of aggregated results presented in this report, and other useful annotations.

- The longer texts extracted from the survey responses, compiled into two Word files: *Future digitisation plans by institution* and *Notes by institution*.

- The detailed information provided for the technical survey, as an Excel workbook and formatted PDF versions of all the responses, both individually and compiled into one document.

2. The audit survey

This section of the report describes how the audit survey was carried out and presents a summary of key data resulting from it.

2.1 Survey methodology

At the start of November 2021, after two months of identifying and reaching out to specific individuals within the target cohort of 264 institutions, each received a personalized email with a request to provide the data needed by mid-December. The email included a PDF version of the survey questions and guidance (**Appendix D**), together with a link to an online form hosted on the Collections Trust website. Further gentle reminders were sent, and after the initial deadline passed others who had agreed to provide data but had not yet done so were given until mid-January. Respondents were encouraged to make contact if they had any questions about how to interpret the questions, and many did.

The researchers would like to thank all those who took the time and trouble to provide information. A few institutions submitted more than one response (eg a museum that provided separate data for its archive and library collections). In these cases, the data has been merged into one response for the institution as a whole.

2.2 Institutions that responded to the audit survey

Of the 264 cultural heritage institutions invited to take part, responses were received from **230** (87%). The full list of institutions that responded is given in **Appendix A**.

No previous attempt to survey the state of the nation's digital collections has come close to this response rate. The proactive, personalized approach needed to achieve this level of engagement was time-consuming, but the result is a dataset that TaNC can explore from many angles beyond the scope of this summary report.

Table 1: Summary of survey respondents[1]		
All respondents	230	
IROs	20	9%
NMDC	46	20%
Designated	72	31%
Recognised	33	14%
NPO	55	24%
RL-UK	27	12%
Others	29	13%
England	176	77%
Scotland	40	17%
Wales	9	4%
Northern Ireland	5	2%
Archive	41	18%
Historic Environment	6	3%
Library	30	13%
Museum	153	67%

[1] Table 1 shows the breakdown of the responses received by institution type and home nation. All but two of the 22 collection-holding IROs responded (two of them anonymously). Unfortunately, the (non-IRO) National Library of Scotland did not engage with the audit process, so the figures for Scotland presented throughout the rest of this section should be read with that major omission in mind.

2.3 Summary of key data from the audit survey

This section presents a summary of the key data resulting from the survey. It follows the order of the questions as they appeared in the online survey form. The PDF version sent to respondents in advance, including guidance notes, is appended (**Appendix D**).

Q2.1 Do your collections database(s) include records down to the level of individual items?

All but three out of 230 institutions reported that they do indeed hold records down to the level of individual items. Some caveats and clarifications were noted by respondents, particularly from archives. These are in the detailed notes document included in the full audit dataset.

Q2.2 How many catalogue records of individual items are in your collection database(s)?

The total number of item-level records reported is nearly **146m**. The distribution of these across the home nations, and across the main types of institution, is shown in **Table 2**.

Table 2: Summary of item-level records reported		
Total	**145,772,632**	**100%**
Of which exact number provided	59,580,100	41%
Of which estimated number provided	86,192,532	59%
England	125,320,547	86%
Scotland	7,587,026	5%
Wales	9,636,178	7%
Northern Ireland	3,228,881	2%
Archive	49,109,930	34%
Historic Environment Record	8,677,260	6%
Library	48,052,064	33%
Museum	39,933,378	27%
IRO	85,779,279	59%
Non-IRO	59,993,353	41%

Table 3: Distribution of item-level records by number held[2]				
Range (records)	**No. of CHIs**	**% of CHIs**	**Total records**	**% of total records**
>= 1,000,000	17	7%	111,365,483	76%
500,000 - 999,999	19	8%	12,835,684	9%
250,000 - 499,999	26	11%	9,617,901	7%
100,000 - 249,999	48	21%	7,519,513	5%
50,000 - 99,999	37	16%	2,811,727	2%
10,000 - 49,999	53	23%	1,468,819	1%
1 - 9,999	27	12%	153,505	0.1%
0	3	1%	0	0%
Total	**230**	**100%**	**145,772,632**	**100%**

---

[2] Table 3 aims to show the great variance across the cohort of respondents in the number of item-level records held. For example, 17 institutions (7% of respondents) each have more than 1m item-level records. In fact, between them, they hold 111m records, 76% of the total across all respondents. At the other end of the spectrum, 27 institutions (12%) have fewer than 10,000 records (excluding the three with none).

Other key points:

- Excluding responses of '0', the median average number of item-level records held is **94,779** (compared to a mean average of **642,170**).

- Ranking the answers to Q2.2 in descending order, the top 10% of respondents (23 institutions) hold **80%** of all item-level records reported.

- Across the UK, four institutions (the British Library, The National Archives, the BBC Archives and Natural History Museum) together account for just over half (51%) of all item-level records reported.

- In Scotland, National Records of Scotland stands out, with almost 40% of the records reported by the 40 Scottish respondents, followed by National Museums Scotland (11%). (The National Library of Scotland did not respond to the audit.)

- The National Library of Wales accounts for 60% of the records reported by the nine Welsh respondents, with 30% held by Amgueddfa Cymru (National Museum Wales).

- The Public Record Office of Northern Ireland holds 50% of the records reported by the five NI respondents.

Q2.3 How many of these records have images (or other digital media) associated with them?

The total number of item-level records with associated images or other digital media is almost **37m**, only a quarter of the 146m records noted above.

Q2.5 asked for 'brief details of any current or planned projects likely to add a significant number of additional images or other media to your collections database'. Encouragingly, 184 respondents provided such details, which are recorded in the full audit dataset.

The distribution of existing records with images, etc, across the home nations and institution types is shown in **Table 4**, while **Table 5** shows the variance in numbers held by institutions across the cohort.

Table 4: Summary of item-level records with images or other digital media		
Total	**36,856,294**	
Of which exact number provided	25,806,784	70%
Of which estimated number provided	11,049,510	30%
England	34,011,034	92%
Scotland	1,480,186	4%
Wales	746,333	2%
Northern Ireland	618,741	2%
Archive	13,582,658	37%
Historic Environment Record	2,666,675	7%
Library	4,073,192	11%
Museum	16,533,769	45%
IRO	24,281,323	66%
Non-IRO	12,574,971	34%

Table 5: Distribution of item-level records with images, etc, by number held				
Range (records)	No. of CHIs	% of CHIs	Total records	% of total records
>= 1,000,000	9	4%	23,675,762	64%
500,000 - 999,999	6	3%	3,585,962	10%
250,000 - 499,999	8	3%	2,509,324	7%
100,000 - 249,999	21	9%	2,981,941	8%
50,000 - 99,999	25	11%	1,724,883	5%
10,000 - 49,999	82	36%	2,048,472	6%
1 - 9,999	73	32%	329,950	0.9%
0	6	3%	0	0%
Total	**230**	**100%**	**36,856,294**	**100%**

Other key points:

- Excluding responses of '0', the median average number of records with associated images or other digital media is **23,368**, compared to a mean average of **164,537**.

- Ranked in descending order by answers to Q2.2, the top 10% of respondents (23 institutions) hold **81%** of all the reported item-level records with associated images or other digital media.

- Five institutions (The National Archives, Natural History Museum, British Library, British Museum and one other IRO) between them hold 52% of records with associated images and other digital media.

- Whereas museums hold only 27% of all item-level records, they have 45% of those with images or other media.

Q2.4 Approximately how many of these images or other media would you describe as 'publication' or 'broadcast' quality?

Respondents were asked to estimate what percentage of the images and other digital media referred to in the previous question were of publication or broadcast quality. This is obviously quite a subjective matter, as one respondent noted: "I would say 90%, our marketing department would say 5%."

Nonetheless, respondents were invited to pick a band of ten percentage points and these were used to extrapolate a minimum and maximum estimate of higher-quality media from the overall numbers provided for question 2.3.

Accordingly, as shown in **Table 6**, it may be that between **28m – 30m** of the 37m images and other media noted above are of publication or broadcast quality, between 73% – 82% of all images and other media reported.

Taking the lower estimate of the range, **Table 7** shows the pattern of variance across the cohort.

Table 6: Summary of images or other digital media of publication or broadcast quality		
	Lowest estimate	Highest estimate
Total	26,869,312	30,263,930
England	24,884,753	28,023,922
Scotland	941,163	1,073,755
Wales	639,487	706,657
Northern Ireland	403,909	459,596
Archive	11,981,571	13,194,649
Historic Environment Record	2,154,891	2,394,892
Library	3,597,408	3,963,995
Museum	9,135,443	10,710,394
IRO	18,370,732	20,541,638
Non-IRO	8,498,580	9,722,292

Table 7: Distribution of publication- or broadcast-quality images, etc, by number held				
Range (records)	No. of CHIs	% of CHIs	Total records	% of total records
>= 1,000,000	6	3%	17,582,297	65%
500,000 - 999,999	4	2%	2,981,479	11%
250,000 - 499,999	5	2%	1,719,938	6%
100,000 - 249,999	12	5%	1,626,348	6%
50,000 - 99,999	19	8%	1,281,998	5%
10,000 - 49,999	59	26%	1,366,125	5%
1 - 9,999	109	47%	311,127	1.2%
0	16	7%	0	0%
Total	230	100%	26,869,312	100%

Other key points:

- Again taking the lower end of the estimated range, and excluding responses of '0', the median average number of records with publication- or broadcast-quality media assets is **9,418** (mean average: 125,558).

- Some 87% of such records are held by 10% of respondents. Four institutions (The National Archives, Natural History Museum, British Library and one other IRO) hold 54% of such records between them.

Q3.1 Do your collections database(s) include records relating to groups of items?

Out of 230 respondents, **188** reported that their institution holds group-level records. Once again, the answers to this question were in many cases accompanied by explanations, recorded in the detailed notes document that forms part of the full audit dataset.

Q3.2 How many catalogue records relating to groups of items are in your collections database(s)?

Not surprisingly, far fewer group records were reported than item-level ones: just under **8m**. **Table 8** shows the distribution across the home nations and institution types, and **Table 9** the variance of numbers held across the cohort. To judge from accompanying notes, many of the group records in museums relate to natural history and archaeology collections.

Table 8: Summary of group records reported		
Total	**7,803,872**	
Of which exact number provided	2,901,069	37%
Of which estimated number provided	4,902,803	63%
England	6,600,907	85%
Scotland	242,770	3%
Wales	921,506	12%
Northern Ireland	38,689	0.4%
Archive	1,645,813	21%
Historic Environment Record	325,828	4%
Library	2,787,963	36%
Museum	3,044,268	39%
IRO	2,637,707	34%
Non-IRO	5,166,165	66%

Table 9: Distribution of group records by number held				
Range (records)	**No. of CHIs**	**% of CHIs**	**Total records**	**% of total records**
>= 1,000,000	1	0%	1,000,000	13%
500,000 - 999,999	4	2%	3,118,280	40%
250,000 - 499,999	3	1%	929,792	12%
100,000 - 249,999	3	1%	509,877	7%
50,000 - 99,999	14	6%	983,780	13%
10,000 - 49,999	40	17%	1,034,690	13%
1 - 9,999	108	47%	227,453	2.9%
0	57	25%	0	0%
Total	**230**		**7,803,872**	

Other key points:

- Excluding responses of '0', the median average number of group records held is **4,000** (mean: 45,109).

- Ranked in descending order of group record numbers, the top 23 institutions (10%) hold 83% of the total number of such records.

- Five institutions (British Library, Natural History Museum, National Library of Wales, Bradford Museums and Galleries, and another IRO) between them hold 54% of all group records reported.

Q4.1 Are catalogue records (individual or group) published on your own website?

So far, the questions have all been about the records held within institutions' own internal systems. Taking the 146m item-level records and 8m group-level records together gives a total of around **154m** collection records within internal systems.

Section four of the audit survey asked about records published by institutions themselves on their own websites. Over **82%** of the 230 respondents (190) reported that they do publish at least some records on their own website, though the accompanying notes sometimes stress that these are on pages of collection highlights, rather than searchable online catalogues.

Q4.2 How many catalogue records are published on your own website?

Almost **98m** records were reported as available via institutions *own* websites, around 64% of the 154m records counted together above. Note that the audit survey also asked about records *only* published *elsewhere*, as discussed under Q5.1 below. **Table 10** below gives a summary of records published on institutions' own websites, by home nation and institution type. **Table 11** shows how the numbers vary across the cohort.

Table 10: Summary of records published on institutions' own websites		
Total	**97,674,653**	
Of which exact number provided	50,845,305	52%
Of which estimated number provided	46,829,348	48%
England	82,626,792	85%
Scotland	5,232,064	5%
Wales	8,342,664	9%
Northern Ireland	1,473,133	2%
Archive	34,680,272	36%
Historic Environment Record	3,108,860	3%
Library	34,934,886	36%
Museum	24,950,635	26%
IRO	61,963,906	63%
Non-IRO	35,710,747	37%

Table 11: Distribution by number of records published on institutions' own websites				
Range (records)	No. of CHIs	% of CHIs	Total records	% of total records
>= 1,000,000	16	7%	74,370,171	76%
500,000 - 999,999	13	6%	9,094,887	9%
250,000 - 499,999	15	7%	5,508,389	6%
100,000 - 249,999	34	15%	5,809,194	6%
50,000 - 99,999	27	12%	1,909,036	2%
10,000 - 49,999	30	13%	797,602	1%
1 - 9,999	54	23%	185,374	0.2%
0	41	18%	0	0%
Total	**230**	**100%**	**97,674,653**	**100%**

Other key points:

- Excluding responses of '0', the median average number of records published on institutions' own websites is **65,441** (mean average: 516,797).

- Ranking these answers in descending order, the top 10% of respondents (23) have published 82% of all the records available on institutions' own websites.

- The British Library alone has published almost a quarter (24%) of these online records. It is one of four institutions (with The National Archives, National Library of Wales and Natural History Museum) that between them account for 53% of all the records reported in answer to this question.

Q4.3 How many catalogue records on your own website have images or other media?

Nearly **25m** of the 98m records on institutions' own websites (26%) have images or other digital media. This total is broken down by home nation and institution type in **Table 12**, with the varying numbers across the cohort shown in **Table 13**.

Table 12: Summary of records with images, etc, on own website		
Total	**24,797,917**	
Of which exact number provided	10,047,676	41%
Of which estimated number provided	14,750,241	59%
England	23,282,039	94%
Scotland	683,080	3%
Wales	725,804	3%
Northern Ireland	106,994	0%
Archive	10,476,373	42%
Historic Environment Record	1,473,338	6%
Library	3,232,080	13%
Museum	9,616,126	39%
IRO	18,301,566	74%
Non-IRO	6,496,351	26%

Table 13: Distribution by number of records with images, etc, on own website				
Range (records)	**No. of CHIs**	**% of CHIs**	**Total records**	**% of total records**
>= 1,000,000	4	2%	14,987,483	60%
500,000 - 999,999	5	2%	3,221,796	13%
250,000 - 499,999	4	2%	1,642,056	7%
100,000 - 249,999	13	6%	2,147,628	9%
50,000 - 99,999	19	8%	1,398,839	6%
10,000 - 49,999	47	20%	1,115,277	4%
1 - 9,999	89	39%	284,838	1.1%
0	49	21%	0	0%
Total	**230**	**100%**	**24,797,917**	**100%**

Other key points:

- Excluding responses of '0', the median average number of records with associated images (or other digital media) published on institutions' own websites is **11,000** (mean average: 137,005).

- Ranked in descending order by answers to Q4.3, the top 10% of respondents (23 institutions) hold 87% of such records.

- The National Archives accounts for 36% of all the records with associated images, etc, published on institutions' own websites, followed by the Natural History Museum (11%) and British Library (8%). Together, these three institutions are responsible for 55% of all such records.

Q5.1 Are any of your catalogue records only published elsewhere (that is, not also on your own website)?

Out of the 230 respondents, **107** (47%) reported that *some* of their records were *only* published on one or more third-party platforms.

The aim of this question was to explore the extent to which some records were *only* available via aggregators and other third-party platforms. This is clearly the case for the 40 institutions that also reported (in answer to Q4.1) that they did not publish *any* records on their own website.

That sub-group of 40 includes some national institutions (such as English Heritage Trust, Historic Royal Palaces, National Trust for Scotland), and large civic services (including Glasgow City Archives, Museums and Galleries Edinburgh, Leeds Museums and Archives, and Nottingham City Museums and Galleries) as well as smaller institutions of the kind one might expect.

However, it may be that some of the 67 respondents who *do* publish records on their own site have read this question as 'are any of your catalogue records published elsewhere', rather than '…*only* published elsewhere'. Some obvious instances were queried and corrected by the researchers, but some of the remaining answers to this question may be wrong, with implications for the figures provided in answer to Q5.3 and Q5.4 below.

Q5.2 Where are these catalogue records published?

The following sites were mentioned by more than one respondent:

- Ancestry

- Archives Hub

- Art UK

- BFI Player

- Bridgeman Images

- Europeana

- Find My Past

- Flickr

- Google Arts & Culture

- GB3D Type Fossils

- JSTOR

- National Biodiversity Network and Global Biodiversity Information Facility

- Portable Antiquities Scheme

- SCRAN

- TNA Discovery

- The Ogilby Muster

- VADS

- Watercolour World

Other sites mentioned only once are captured in the detailed notes included in the full audit dataset.

Table 14 below shows the number of times each of these sites was mentioned by firstly, all respondents, and also by respondents who reported (in answer to Q4.1) that they did not publish any records on their own website.

Art UK was mentioned the most: by 67 institutions, 22 of which do not publish records on their own website. The National Archives' Discovery portal was mentioned by 11 respondents (including seven museums with archival holdings).

Table 14: Number of times sites mentioned in response to Q5.2

	Ancestry	Archives Hub	Art UK	BFI Player	Bridgeman Images	Europeana	Find My Past	Flickr	Google A&C	GB3D Type Fossils	JSTOR	NBN/GBIF	Portable Antiquities	SCRAN	TNA Discovery	The Ogilby Muster	VADS	Watercolour World
Total	5	8	67	2	7	3	8	5	7	2	4	3	2	4	11	2	2	2
No online records on own site	0	2	22	1	3	0	1	1	2	2	1	3	0	2	1	1	0	0
England	4	4	49	2	6	3	7	2	7	2	2	3	2	0	10	2	2	2
Scotland	0	2	16	0	1	0	0	1	0	0	1	0	0	4	1	0	0	0
Wales	1	2	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
NI	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
Archive	3	2	4	2	0	0	6	1	1	0	1	0	0	1	1	0	0	0
HER	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0
Library	0	3	6	0	0	0	0	1	1	0	0	0	0	0	3	0	1	0
Museum	2	3	56	0	7	3	2	2	5	2	3	3	2	2	7	2	1	2
IRO	2	0	3	0	0	0	1	0	1	0	2	1	0	1	0	0	0	0
Non-IRO	3	8	64	2	7	3	7	5	6	2	2	2	2	3	11	2	2	2

Q5.3 How many of your catalogue records are only published elsewhere?

As noted under Q5.1, 67 institutions that publish records on their own websites reported that some other records were *only* published on third-party platforms. The researchers queried and corrected a few instances where answers given here were clearly wrong, but suspect that some other respondents may have double-counted records that are *also* published on their own websites.

With that caveat, the total number of records published *only* on third-party sites is **3.75m** as shown in **Table 15** and **Table 16**.

Table 15: Summary of catalogue records only on third-party sites		
Total	**3,750,753**	
Of which exact number provided	2,974,470	79%
Of which estimated number provided	761,152	20%
England	3,643,778	97%
Scotland	22,613	1%
Wales	64,305	2%
Northern Ireland	20,057	1%
Archive	1,051,391	28%
Historic Environment Record	110,128	3%
Library	131,347	4%
Museum	2,457,887	66%
IRO	741,899	20%
Non-IRO	3,008,854	80%

Table 16: Distribution by number of catalogue records only on third-party sites				
Range (records)	No. of CHIs	% of CHIs	Total records	% of total records
>= 1,000,000	1	0%	1,502,000	40%
500,000 - 999,999	2	1%	1,432,338	38%
250,000 - 499,999	0	0%	0	0%
100,000 - 249,999	2	1%	268,000	7%
50,000 - 99,999	2	1%	155,000	4%
10,000 - 49,999	11	5%	278,372	7%
1 - 9,999	78	34%	115,043	3.1%
0	134	58%	0	0%
Total	**230**	**100%**	**3,750,753**	**100%**

Q5.4 How many of these records published elsewhere have images or other media?

As **Table 17** and **Table 18** show, **3.4m** records - 92% of the 3.7m records published only on third-party platforms - have associated images or other digital media.

Table 17: Summary of records with images, etc, only on third-party sites		
Total	**3,426,397**	
Of which exact number provided	2,747,377	80%
Of which estimated number provided	674,370	20%
England	3,369,108	98%
Scotland	25,159	1%
Wales	19,779	1%
Northern Ireland	12,351	0%
Archive	1,043,840	30%
Historic Environment Record	102,244	3%
Library	65,270	2%
Museum	2,215,043	65%
IRO	725,418	21%
Non-IRO	2,700,979	79%

Table 18: Distribution by number of records with images, etc, only on third-party sites				
Range (records)	No. of CHIs	% of CHIs	Total records	% of total records
>= 1,000,000	1	0%	1,502,000	44%
500,000 - 999,999	2	1%	1,432,338	42%
250,000 - 499,999	0	0%	0	0%
100,000 - 249,999	1	0%	150,000	4%
50,000 - 99,999	2	1%	155,000	5%
10,000 - 49,999	3	1%	77,949	2%
1 - 9,999	78	34%	109,110	3.2%
0	12	5%	0	0%
Total	**99**	**43%**	**3,426,397**	**100%**

Q6.1 What is your general policy on how people can use images, etc, for which you own the copyright?

Complementing more detailed studies commissioned by TaNC, respondents were simply asked to select one or more of the following options to reflect their general policy on the use of images, etc, *for which their institution owned the copyright*:

- Unrestricted re-use allowed

- Free non-commercial re-use allowed, with some conditions (eg attribution)

- Commercial re-use allowed for a fee

- No re-use allowed

- Other

Key points:

- Only four respondents gave 'unrestricted' as their only answer to this question. These included Birmingham Museums Trust, the Wellcome Collection and York Museums Trust. A further nine included 'unrestricted' in answers that also mentioned one or more of the other options given.

- The majority of respondents, 180 (78%) included 'non-commercial' in their answer, of which 45 respondents (almost 20%) gave this as the only option. Most others typically combined a policy of free non-commercial use with charging for commercial use.

- Only 16 respondents gave 'commercial' as their only answer to this question, most notably the BBC Archives and Government Art Collection, but also some smaller archives and museums too.

- A few respondents included the options 'none' or 'other' in their answers, some with accompanying notes suggesting that the question had not been interpreted as intended.

No attempt has been made to extrapolate what these responses might mean in terms of numbers of digital assets available for, say, unrestricted or free non-commercial use, as it cannot be assumed that the institutions themselves own the copyright for all the assets reported in answer to Q4.3 and Q5.4.

Q6.2 Are any of your catalogue records available online only behind a pay wall?

Only a few institutions - **14** - said that some of their records were *only* available behind a paywall. As well as The National Archive's Discovery portal, other paywall sites mentioned include Find My Past, Ancestry, and The Ogilby Muster, a new portal for regimental museums from the Army Museums Ogilby Trust.

Q6.3 How many of your catalogue records are only available behind a paywall?

Around **5.3m** records are only available behind a paywall, as summarised in **Table 19.**

Table 19: Summary of records only available behind a paywall		
Total	**5,278,622**	
Of which exact number provided	2,377,407	45%
Of which estimated number provided	2,901,215	55%
England	5,278,122	100%
Scotland	0	0%
Wales	500	0%
Northern Ireland	0	0%
Archive	3,121,406	59%
Historic Environment Record	0	0%
Library	51,941	1%
Museum	2,105,275	40%
IRO	2,833,471	54%
Non-IRO	2,445,151	46%

From **Table 20** below it is clear that (excluding responses of '0') the distribution of these records is split between four institutions with large numbers, 10 more with relatively few, and nothing in between.

Table 20: Distribution by number of records only available behind a paywall				
Range (records)	**No. of CHIs**	**% of CHIs**	**Total records**	**% of total records**
>= 1,000,000	2	1%	3,738,000	71%
500,000 - 999,999	2	1%	1,432,338	27%
250,000 - 499,999	0	0%	0	0%
100,000 - 249,999	0	0%	0	0%
50,000 - 99,999	0	0%	0	0%
10,000 - 49,999	4	2%	100,050	2%
1 - 9,999	6	3%	8,234	0.2%
0	216	94%	0	0%
Total	**230**	**100%**	**5,278,622**	**100%**

Other key points:

- Almost all, 98%, of all the reported records behind a paywall originate with the four high-volume institutions noted above: The National Archives (42%), Royal Artillery Museum (28%), The Box Plymouth (Archives) (16%) and Royal Museums Greenwich (11%).

Q7.1 Do you have an API that allows others to make use of your online collections?

Asked whether their institution had 'an API that allows others to make use of your online collections', **41** respondents (21%) said they did. Between them, those institutions have **61.5m** records published on their own websites, though it is not known how many of their records are available via the reported APIs.

Other key points:

- A further **36** respondents (16%) said that the introduction of an API was pending. Making the same assumption noted above, it is possible that a further **13.5m** records will become re-usable once these pending APIs are implemented.

- Almost **half** (49%) of respondents said they did not have APIs, while the rest did not know.

Q7.2 Are any of your online images IIIF-enabled?

Complementing the work of the TaNC Foundation project *Practical Applications of IIIF*, the audit survey asked whether institutions had any International Image Interoperability Framework (IIIF)-enabled images. As shown in **Table 21**, an encouraging **39** (17%) reported that they did, while a further **29** (13%) said it was pending.

Table 21: Summary of institutions with IIIF-enabled images		
Total	**230**	
Yes	39	17%
No	116	50%
Pending	29	13%
Unknown	46	20%
England	29	13%
Scotland	7	3%
Wales	2	1%
Northern Ireland	1	0%
Archive	6	3%
Historic Environment Record	0	0%
Library	13	6%
Museum	20	9%
IRO	5	2%
Non-IRO	34	15%

Q7.2.1 How many of your images are IIIF-enabled?

While not all the institutions that said they had some IIIF-enabled images provided even estimated numbers, at least **49m** such *images* were reported, as shown in **Table 22**.

Table 22: Distribution by number of IIIF-enabled images				
Range (records)	No. of CHIs	% of CHIs	Total records	% of total records
>= 1,000,000	2	1%	46,692,352	95%
500,000 - 999,999	3	1%	1,600,000	3%
250,000 - 499,999	1	0%	321,530	1%
100,000 - 249,999	2	1%	367,916	1%
50,000 - 99,999	2	1%	168,213	0%
10,000 - 49,999	7	3%	172,584	0%
1 - 9,999	13	6%	21,423	0.0%
0	9	4%	0	0%
Total	**39**	**17%**	**49,344,018**	**100%**

Other key points:

- The Wellcome Collection alone reported 44m of these (89%), relating to 367,000 online catalogue records.

- NHM has 5% of IIIF images reported, with Durham University Library, BL and Cambridge University Library all accounting for about 1% each.

Q 7.3 Do you make use of external identifiers for your online records, to help minimise the problem of broken links?

To complement another of TaNC's Foundation projects, *Persistent Identifiers*, the audit survey asked whether institutions made use of external identifiers to help minimise the problem of broken links. Only **14** respondents (6%) said they did, with a further 33 (14%) saying it was pending. The great majority said they either did not use external identifiers (57%) or did not know (23%).

3. Technical survey

This section of the report, by Adrian Cooper of Intelligent Heritage, summarises the second strand of the digital collections audit commissioned by TaNC: a deeper dive into technical aspects of the systems and processes relating to collections management and digital asset management in the IROs and some other big cultural heritage institutions.

3.1 Scope of the technical survey

Adrian Cooper conducted a total of 27 structured conversations with the relevant staff of 18 IROs and nine other cultural heritage institutions. These are listed in **Appendix B**.

Given that the information covered often required input from several people, it was not possible to schedule interviews with all the IROs as originally hoped. Nevertheless, together the 27 institutions that did take part in the technical survey account for around 70% of all the online records, images and other digital media reported in the main audit survey discussed in the previous section.

3.2 Technical survey methodology

Once participants had agreed a date for an online meeting, they were sent a list of questions that would be covered in the structured conversation and also a worksheet for them to start compiling the information needed. The questions are set out in **Appendix E**. They are grouped into five areas, as shown in the table below.

Area	Purpose
Format	Understanding types and formats of digital resources made available to researchers and public users
Metadata	Understanding what types of metadata are created, the schemas and standards used, and how institutions manage the relationship between digital assets and physical artefacts
Asset management	Understanding systems, scope, and service models for managing digital assets and associated pipelines for asset delivery to researchers and public users
Access and availability	Understanding how institutions control access to digital resources, levels of access, and the capacity for machine-based enquiry
Rights management	Understanding how licences and usage rights are assigned to digital assets

3.3 Summary of technical survey findings

This section of the report sets out some of the main findings of the technical survey. The complete responses received from all participating institutions are included in the full audit dataset.

Format

- The following types of digital asset are created, managed and made available:
  - Images
  - Documents
  - Video

- Audio
- Datasets
- Web archives
- GIS data
- 3D models

- Images make up the overwhelming majority of the total digital resources in the sector.

- 15 institutions are starting to make video files available but the volumes are comparatively low. The only institution with a substantial digital video repository is the BFI. However, this resource is restricted to streaming via the BFI website only.

- Only six out of the 28 institutions included in the audit are currently making digital images available using the International Image Interoperability Framework (IIIF) - https://iiif.io/.

- Six institutions are actively implementing IIIF and a further five are hoping to start a IIIF project in the near future.

Metadata

Digital asset metadata is essential to content discovery and all institutions are providing core metadata alongside their resources.

Descriptive metadata

- Of the institutions surveyed, digital resources are often surrogates of physical artefacts in the collection. As such, records managed within either a collections management system (CMS) or an archival management system are usually the basis for descriptive metadata of any related digital surrogates.

- Some of the institutions interviewed manage both collections and archives. In most cases these different record types are managed in separate systems.

- Whilst museum collections management systems are often Spectrum compliant, this is not a metadata standard. This results in huge variations in the structure and semantics for object-based descriptive metadata. Institutions have tended to develop their own structures (based on historic, in-house cataloguing guidelines) rather than follow or work towards a standard. On the other hand, archives are more likely to use a specific standard (such as ISAD(G)).

- The lack of a shared structure and semantics for collection records inhibits the ability for museums to share and exchange information easily. This reduces the ability to develop third party, machine-readable services that can scale or operate across multiple collections simultaneously.

Technical metadata

In addition to descriptive metadata (derived from catalogue records), digital assets often contain technical embedded information.

- 19 institutions stated that they routinely add embedded metadata to digital resources as part of their delivery pipelines.

- In some cases, institutions embed elements of descriptive metadata into their digital files (such as object title, maker, object type etc.).

Administrative metadata

Administrative metadata (details of asset organisation, source, ownership etc.) is very important but often overlooked.

- Few institutions reported that they include any comprehensive administrative metadata as part of their normal digital resource delivery.

- Lack of administrative metadata is likely down to the fact that, for most institutions, they are only delivering digital resources to their own website. However, given the increasing desire to collaborate and exchange data this will become more important. As records travel beyond current organisational boundaries and/or are made available for machine-based enquiry via APIs, including core administrative data will be vital.

Persistent identifiers

- Very few institutions are using third party, independent services to provide unique, persistent identifiers for catalogue records or digital assets. Most institutions tend to generate identifiers (URLs) based on internal system IDs. These are open to potential change over time which would result in broken links for any dependent resources.

- Adoption of a third-party PID system was considered to be a low priority for most institutions interviewed. There are many other infrastructure or process related improvements to consider first.

Legacy systems

- The audit process revealed that many institutions are dissatisfied with current systems and processes. In most cases, this was down to frustrations with legacy systems that didn't have APIs resulting in difficulties integrating with other systems and services.

- A large number of institutions stated that they had either just embarked on the process of implementing a new collections management or digital asset management system or had budget approval to start the process.

- Several institutions acknowledged that they were in a state of flux, knowing that the transition from current to new systems would take several years.

Asset management and delivery

- 23 institutions are using a dedicated digital asset management system (DAMS) to provide enterprise level management of their digital resources.

- 20 institutions are using commercial products and the other three have developed their own bespoke DAMS solution. Bespoke solutions were developed due to a combination of sizing, scale and functionality issues experienced with existing off-the-shelf solutions.

- Four institutions reported that they were using DAMS functionality within their exciting collections management system. Collections management systems used as DAMS offer a relatively easy way to reference associated digital resources and share metadata.

- CMS systems used as DAMS lack the overall range of functions usually associated with the full digital asset management lifecycle. This has resulted in a greater reliance on the development of bespoke scripts and tools to support digital asset transformation and delivery for online research or commercial licensing.

- Most institutions use their dedicated DAMS to ingest, store and manage their digital assets. However, only six institutions reported that they were using in-built functionality to automate the transformation of digital assets into different formats and sizes for public or commercial purposes. The rest are relying on the development of bespoke tools and services. Larger institutions have tended to build and manage their own digital asset pipelines.

- Only four institutions reported that they are actively using or connecting their DAMS to a dedicated digital preservation system. The majority of institutions are still investigating a proactive approach to digital preservation. A few institutions are currently embarked on projects to implement and/or test a new digital preservation system.

- Very few (3) institutions said that they use their DAMS to manage and control delivery of digital assets. The majority of digital asset pipelines and workflows are developed in house.

- Delivery of digital assets relies on the ability to integrate catalogue data (normally held in a CMS) and technical metadata about the digital assets (held in a DAMS).

- Many institutions have commissioned or developed bespoke one-to-one integrations between their CMS and DAMS systems. Three institutions said that the integration between CSM and DAMS was completely manual.

- 10 institutions are using a middleware solution to aggregate, synchronise and simplify the integration between catalogue records and digital assets. Eight of these are using the same system (Knowledge Integration's CIIM - https://www.k-int.com/products/ciim/). The other two are using services developed by third parties.

- 15 institutions reported that the processes to deliver digital assets online are fully automated, consistent and repeatable.

- For institutions already using IIIF, the generation of manifests was either dynamic or fully automated.

Access and availability

- All institutions use their internal catalogue systems to control public and research access to digital resources. This is normally via a range of licensing types or publishing status flags applied directly to individual catalogue records.

- Catalogue records and related digital resources that are approved for online delivery are usually published (copied) to a dedicated webspace hosted by the institution and all public users have the same access level.

- The British Library is the only institution interviewed that uses a controlled API to provide different types and levels of access to partner institutions.

- Six of the institutions interviewed have an Open API to allow third parties to undertake machine-based enquiry or integration with external systems. The Open APIs offer a limited range of fixed endpoints (type of query) and resource records are based on bespoke schemas. These bespoke schemas reflect internal data structures rather than any agreed third party metadata standard.

- Whilst some institutions provide documentation for their APIs it does mean that third parties will need to work out, apply and store organisational specific mappings before being

able to integrate data into their systems and services.

- Existing Open APIs tend to have low usage levels. Several institutions reported that their APIs struggled to perform or stalled completely when attempting higher volume queries for collaborative research projects.

Rights management

- The majority of institutions assign licences and usage rights to individual digital assets.

- In many cases the rights management workflow starts with assignment of rights to records in the related catalogue systems. These are then cascaded to digital assets (held in separate digital asset management systems). This workflow is often manual and time consuming.

- In some cases, institutions are able to use bespoke scripts and pipelines to apply a range of licence types based on publishing status flags and/or other logical parameters applied to the catalogue data.

- Only seven institutions stated that they were able to assign rights to group level records in order to automate and speed up the rights management process.

4. Roundtables with smaller museums and archives

In a third strand of activity, seven roundtable discussions were convened with people from a range of smaller museums and archives to gain qualitative insights into their capacity to join any future national digital collection.

4.1 Scope of the roundtables

The dates of the roundtables and names of those taking part are listed in **Appendix C**. Some responded to invitations sent out via museum networks in Cornwall and Northern Ireland, others signed up after reading social media posts about them from TaNC.

The open nature of the call meant that a few participants were from institutions that were also included in the main audit survey, but for the most part those attending were from small museums and archives outside the cohort defined for the survey. A few were from sector bodies, and one roundtable was just for the lead contacts of museum Subject Specialist Networks.

There is no suggestion that such a tiny and self-selecting group of people is statistically representative of smaller museums and archives, but the points they made were familiar from countless similar conversations over the years, in museums large and small. Extended, anonymized quotations are provided in **Appendix F** (lightly edited to remove the verbal ticks and errors of spontaneous speech).

The quotations are best read as vignettes that reflect the daily reality of most people working or volunteering in collections-related roles in most UK museums and archives. They serve as a bracing reality check to all of us deeply engaged in TaNC and related initiatives.

4.2 Key themes

As even a cursory glance of the roundtable quotations makes clear, a few key themes dominated most of the discussions.

Time and money are scarcer than ever

"The bottom line is that over years of austerity, sadly documentation is one of the last things that people get to do, it's always on their To Do List. They want to do it, they see the benefit of doing it, but if you're the curator plus the learning officer plus the person that's opening the door, and the person is cleaning the toilets and welcoming people to the museum, then actually working on your collection documentation can be the last thing on your list."

"Just getting money to do basic cataloguing is a nightmare but for us there's another issue terms of resourcing: while I would love to get more people to use our collections, I do not have the time - and the organization does not have the resources - to employ someone to meet increased demand that would come from better knowledge of our collections. So if I put a load of stuff on Discovery *that was driving people towards asking me questions, or wanting to come and visit, we couldn't fund that through the charity, especially as the charity doesn't have a purpose for archiving, education or history. Our [charitable] purpose is helping older people in the community with difficulties getting into hospital or something like that, not research for academics. So that's another big barrier in that we don't want to create demands that we can't fulfil."*

Many staff and volunteers struggle to access and use their own data

"[Our collections management system] is incredibly clunky. It's a very arcane piece of client server database with a HTML front-end screwed onto it badly. It does follow the Spectrum standards and it's understood by the community, but it is not in any way ideal for a tiny museum (we're a micro museum). We have a single user license: one PC with a standalone version that we have to basically move on to other PCs when we want to do some field-based work. It's clunky but we don't have the kind of budget to invest in a full-blown system. It just wouldn't financially make any sense. However, because we're a group of volunteers that work remotely we can't really share the same PC it just doesn't work."

"We have a regular problem: we have a window and there's a group that would like to have a new exhibition in it every two or three weeks. They want to dip into local activities or into what we have in the archive to bring things in. It's very difficult. We don't do it very well; we haven't got a tool which is appropriate to the needs of the community and it's as simple as that."

Poorly-documented and hard-to-manage images are a big problem

"The thing I'm particularly working on at the moment is the [digitised] photographs. They are part of the museum collection which only exists as an image; there's no physical object associated with them. This is something that people doing exhibitions or the curator or anyone trying to work with the museum information really needs access to - in a searchable way, because a directory of jpegs with nothing to help you identify what they are is absolutely useless. I'm ploughing through those three thousand images at the moment; obviously they need to be searchable so that people [can] see all the pictures of such-and-such an event, or all the pictures of women in WWII doing ambulance work, or all the pictures of the local bowling. I'm trying to find a way that you could tag all of these photos so that they would become searchable in the way that big organizations organize their image collections."

"And that's before you start asking about the images, audio, video. I think there is a lot of non-management with this material – I'm pushing for DAMs ... because we have c. 100k images just of our collection and hours are lost searching for images. I worked on a funded project for another body and they have lost all my TIFFs because they didn't know they mattered or perhaps IT thought they didn't need them or failed to move them as they changed infrastructure…who knows? But now they have very few images of an appropriate quality to use on their new collection website. Fundamentally, if no one understands how to manage or pays attention to this material, it is lost very easily."

Real, practical help from national institutions is hugely welcome

"We got a grant from the Business Archive Council to train us in cataloguing because we have years of business records of running the pier. And we've used [TNA's] Manage Your Collections from the get-go … The National Archives are extremely generous with their training and support. Three of them came and spent three days with my team, so that we were completely trained in how to do it. I would thoroughly recommend that. One of the reasons for doing it was because it meant we would go on to Discovery and our collections could be accessed by people from anywhere in the world."

"Is the model that Art UK used something that would be useful? They put out a call for all public institutions with artwork to be part of the public catalogue programme, and they paid for the digitization, and paid for copyright clearance. They paid for photographers, and they paid for people to do the documentation, the text, the metadata creation. And then it was all put online in one central catalogue. Easy."

Small institutions want to benefit from cutting-edge technologies too

"I deal with a lot with copyright and it is unbelievably complicated, but there is guidance out there. It's just having the time and the energy to [do what the] guidance tells you is almost impossible … If there was a project which, for example, used AI to look at all copyright owners in all the archives across the country that would be a brilliant start, because just finding the [correct rights holder] can take me days sometimes, bringing up local records and finding out which newspaper bought this newspaper, which bought that newspaper. If you could create a definitive list of copyright owners in the UK that would be fantastic. I guess there must be a fair amount of duplication, that other archives may have done that research and spent the days and days figuring out what the copyright position is regarding a particular individual, and another archive also has material relating to that same individual and they're also having to do their several days' worth of work as well. Often that background information never makes it onto a catalogue and so isn't easily searchable."

"The [digitisation] we have done we hold on our own service. That's not very sustainable in the longer term. If there was a nationally-run Preservica which we didn't have to pay a small fortune for that would be massively important for us, because it's a cost that as a charity we find difficult to [afford]. I'm now running something like a terabyte's worth of information, which is more than the entire rest of the organization has in digital assets, so that's an issue for us … There are no checksums going on or anything like that; it's just a server sitting there and there's a backup tape that runs every day and, once a month, the tape comes to my house with me."

Institutions are not making the best use of existing infrastructure

"We did do a couple of projects a while ago creating pages worth of digital content involving some of our wartime records, and those are available through a CalmView site we have. The use of that material is somewhat limited because people don't necessarily know we exist in the first place, or how to come across that data, and for us that's an issue. We don't have the resources to, for example, transfer all that material to Discovery.*" [Do you have a collection-level descriptions on* Discovery?*] "Not really, not that would be useful to people."*

"We are involved in the Heritage Information Access Strategy, and some of the datasets that Historic England have been looking at. I wanted to reference the rebirth of Heritage Gateway in a different format. What it does is harvest from other databases and provides access to records that way. When we first started engaging with the project the obvious question was "is there one point of access for all information about archaeological collections held in museums?" And the answer is there is not. There is some information that it held within Heritage Gateway about archaeological collections, mainly because they've been excavated and they appear on historic environment records, so HERs are obviously held by local authorities as part of the planning process in the sense that they need to keep that information. I can tell you that we've got very expert people running those systems and those systems are still falling short in terms of whether they are accurate or not – and accuracy is an issue."

There is confusion and some cynicism about TaNC and the concept of a 'national collection'

"[TaNC] is such a big project, but actually it is quite difficult sometimes to understand where certain museums might fit within a broader picture. I don't understand what [TaNC] is there for, what is it trying to achieve. I come from a marketing background so will always go for audience first. Even having been involved in most of the meetings, so far, and the public meetings and everything else about TaNC, and talking to people who are sitting on the steering committee, I still can't get what it is."

"I think academic researchers default to thinking that the significant collections are within national museums or large regional museums when actually there are significant collections in the tiniest of museums, tucked away. Some campaign or something needs to be done to chip away at that attitude. Anyway, I think the whole TaNC programme is almost perpetuating that, and there's potential for a missed opportunity to get the local significance of collections out there."

5. Conclusions and recommendations

This digital collections audit has generated a huge volume of quantitative and qualitative data that can be analysed and interrogated by TaNC in many different ways. This summary report presents only the headline findings, so the final section draws only a few headline conclusions and, where it seemed useful, recommendations to guide future planning and action.

5.1 The UK's digital collections are concentrated in a few institutions

The surveyed cohort was deliberately designed to include the 264 cultural heritage institutions likely to have the biggest collections and the most capacity, or (in the case of small Designated and Recognised collections) have access to specific funding streams. By these, and measures such as acknowledged significance, this is probably the top tier of our cultural heritage sector.

Although the audit survey has revealed some impressively large numbers of records and digital assets, a consistent pattern emerges across several survey results: that around 10% of the surveyed cultural heritage institutions account for around 80% of the records and digital assets, both online and offline. Often, a handful of mega-collections hold more than half the total tally.

On the one hand, this suggests some 'quick' wins might be possible, bringing together a high volume of material from just 20 or so institutions. On the other, when the institutions are ranked by numbers of records or assets, the pattern consistently shows a steep drop-off after the highest-ranking 10%. If there is such a yawning digital divide even within the 264 relatively privileged institutions surveyed, what about the thousands of other museums, libraries and archives?

5.2 The concept of a 'national collection' is problematic

TaNC describes its work as taking the first steps towards "a unified virtual 'national collection'".[3] However, the roundtables with smaller institutions revealed the phrase 'national collection' to be deeply problematic, even when tempered with 'towards'.

The phrase sets up expectations that may be hard to fulfil. It prompts questions about which institutions are in the 'national collection' and which are not. Indeed, concerns were raised in the roundtables even about the criteria used to define the cohort of 264 institutions included within the main audit.

As noted above, around 80% of the records and digital assets reported to the audit are held by only a score of larger institutions. These are concentrated in London and other large cities that TaNC wants to reach beyond, according to its stated aim of "directly addressing the issues related to accessibility beyond current metropolitan centres."[4]

The quick win noted above - bringing together, say, the top 20 digital collections - would compound the dominance of these metropolitan institutions. Even aiming to unify the digital collections of the 264 institutions included in this audit would not result in a 'national collection' that the roundtable participants would recognise as such.

The data compiled for this audit will help TaNC, and AHRC more widely, scope the extent of its ambition for making digital collections more accessible to researchers and the wider public, and the line will clearly need to be drawn somewhere. These scoping exercises will be more difficult than

---

[3] https://www.nationalcollection.org.uk/about
[4] Ibid.

they need to be if they try to define any kind of 'national collection'.

It is recommended that, in scoping what might happen after TaNC, the unhelpful connotations of the phrase 'national collection' are avoided.

5.3 Few institutions use persistent identifiers

Both the audit survey and technical survey echo the findings of TaNC's Foundation Project on *Persistent Identifiers* (PIDs): few institutions are using PIDs, or even know what they are.

This situation should set alarm bells ringing, as it is a serious barrier to achieving data that meets the FAIR principles of being findable, accessible, interoperable and reusable. Indeed, the principle that '(Meta) data are assigned globally unique and persistent identifiers' is 'arguably the most important because it will be hard to achieve other aspects of FAIR without [them].'[5]

In short, there is a real risk that the millions of online records reported to this audit will, over time, turn into millions of broken links.

The technical survey showed that adopting external identifiers was a low priority even amongst the IROs and other large institutions. Even here, at the very top of the sector, any strategy that relies on individual institutions taking the need for robust PIDs more seriously, and following through with the required action, seems optimistic. There seems even less chance that smaller institutions will give this matter any thought at all.

It is therefore recommended that, as a priority, a sector-level strategy should be developed for assigning and supporting PIDs, taking the onus off individual institutions.

Workable and sustainable solutions are likely to involve a mix of:

- An agreed sector-wide approach (of the kind the *Persistent Identifiers* project is working towards).

- Automatic assigning of PIDs as a by-product of support given by third-party services (see below).

- Incentives and conditions within project grant schemes and other funding streams.

5.4 Support at scale helps institutions share their data

The roundtable discussions with smaller institutions provided a good, representative snapshot of the limited capacity many have to document their collections at all, and their reluctance to open up data they feel is not ready for sharing. Indeed, data-sharing *per se* came across as a low priority, particularly if it involves extra work for no perceived benefit.

The inescapable conclusion is that these institutions, and doubtless many larger ones, will need to see tangible benefits before they make their data available, with much financial and practical help along the way.

The frequent mention of Art UK and The National Archives' Manage Your Collections and Discovery, in both the audit survey and roundtable discussions, suggests a way forward. These platforms (and others, like the relatively new platform for regimental museums, The Ogilby Muster) offer not only a route online, but also practical help such as photographing and scanning collections and dealing with

---

[5] https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/

complex issues such as copyright that are often difficult for individual institutions to master. The barrier to taking part is low, and the incentive to do so is clear.

Many collection types fall outside the remit of the existing services mentioned by the survey respondents, and new services need to be established and nurtured.

Among other things, support services like these offer a potential solution to the low take-up of PIDs by individual institutions noted above. Assigning PIDs, and underwriting their long-term resolvability, could be done at scale by these services without the contributing institutions even being aware of it.

Solving the documentation problems of the whole cultural heritage sector is not, of course, AHRC's responsibility. That requires a strategic, joined-up approach involving many sector bodies and funders. However, if AHRC wants institutions to share their collections data for research purposes, it will need to offer tangible benefits that make it worth their while, either directly or through third-party services that already work with hard-to-reach parts of the sector.

As AHRC scopes what might come after TaNC, it should engage with the problem of helping the majority of institutions share their data, working in partnership with existing services that already do this and, where needed, with new ones.

The practical support needed does not fall easily within the scope of research of the kind funded through TaNC. As attention turns to potential future infrastructure, however, there are opportunities for AHRC to make a clear, equitable offer to cultural heritage institutions beyond those involved in funded projects.

One example might be free storage in a trustworthy digital repository in return for shared data. That kind of practical help would go a long way towards easing concerns raised in the roundtable discussions, and would be seen as a fair contribution by AHRC to the overall costs of creating and sharing collections.

5.5 Collection-level records would be a useful starting point for many

Although the focus of the audit survey was on item-level records, images and other digital media, the roundtable discussions pointed to the potential value of collection-level records. Much of the discussion in these groups centred on the perceived impossibility of ever having enough resources to work on collection catalogues to the point where they could be published online. It was striking how much of an all-or-nothing mindset was reflected in many of these conversations.

However, if AHRC's main aim is to make collections more accessible to researchers, good collection-level descriptions would be better than no online information at all. Even for the most hard-pressed curator, librarian or archivist, a few hundred words on the scope and highlights of their collections feels achievable, and has probably already been prepared for other purposes.

There is an opportunity here to make the most of existing infrastructure and data.

[Museum] Subject Specialist Networks (SSNs) might be partners in projects relating to their areas of expertise, such as the forthcoming collection mapping exercise mentioned by a representative of the Religion, Collections and Heritage Group. Other SSNs have carried out similar work in the past, but the results are no longer available due to the limited technical capacity of these groups.

It is recommended that TaNC supports a wider sector drive to create (or update) a useful collection-level record for every institution, especially those represented only by minimal listings in existing databases, and to re-purpose legacy datasets from past projects.

Appendices

Appendix A: Institutions included in the audit survey data set

The following list includes all the 230 institutions that responded to the audit survey. Note that 17 institutions did not give consent for their details to be shared with TaNC. Here, these institutions are simply identified as 'Institution A', 'Institution B', etc.

Aberdeen Archives, Gallery and Museums

Aberdeenshire Council/Live Life Aberdeenshire/Aberdeenshire Museums Service

Almond Valley Heritage Trust

Amgueddfa Cymru (National Museum Wales)

Ashmolean Museum (University of Oxford)

Bail' Ach' an Droighinn (Auchindrain Historic Township)

Bangor University (Archives and Special Collections)

Barnsley Museums

BBC Archives

Beamish Museum

Birmingham Museums Trust

Black Country Living Museum

Borthwick Institute for Archives (University of York)

Bowes Museum

Box Plymouth (Archives)

Box Plymouth (Media Archive Collections)

Box Plymouth (Museum)

Bradford Museums and Galleries

Bristol Museums

British Library

British Museum

Britten Pears Arts

Bronte Parsonage Museum

Cambridge University Library (special collections)

Cardiff University

Chatham Historic Dockyard Trust

Chetham's Library

Churchill Archives Centre (University of Cambridge)

Colchester + Ipswich Museums

Compton Verney Art Gallery and Park

Courtauld Gallery

Culture Coventry (Herbert Art Gallery and Museum, Coventry Transport Museum and Coventry Archives)

Culture Perth and Kinross Museums and Galleries

Culture Trust Luton

Derby Museums

Dulwich Picture Gallery

Dumfries and Galloway Council Arts and Museums Service (East)

Dundee Heritage Trust

Durham University Library and Collections

Durham University Oriental Museum

East Anglian Film Archive (University of East Anglia)

East Ayrshire Leisure Trust

Elgin Museum

English Folk Dance and Song Society

English Heritage Trust

Fitzwilliam Museum (University of Cambridge)

Glasgow City Archives

Glasgow Life: Glasgow Museums and Collections

Glasgow School of Art

Glasgow Women's Library

Gordon Highlanders Museum

Government Art Collection

Grant Museum of Zoology (University College London)

Groam House Museum

Gwynedd Archives Service

Hampshire Archives and Local Studies / Wessex Film and Sound Archive

Harewood House Trust

Heritage Doncaster

Highland Folk Museum

Historic England

Historic Environment Division, Northern Ireland

Historic Environment Scotland

Historic Royal Palaces

History of Science Museum (University of Oxford)

Hull Culture and Leisure Limited

Hunterian (University of Glasgow)

Imperial War Museums

Institution A

Institution B

Institution C

Institution D

Institution E

Institution F

Institution G

Institution H

Institution I

Institution J

Institution K

Institution L

Institution M

Institution N

Institution O

Institution of Engineering and Technology

Institution P

Institution Q

Ironbridge Gorge Museum Trust

Jewish Museum London

John Rylands Research Institute and Library (University of Manchester)

King's College London

Kresen Kernow (Cornwall Record Office)

Lambeth Palace Library

Lapworth Museum of Geology (University of Birmingham)

Leeds Museums and Galleries

Leicester Museums and Galleries

Leicestershire County Council Museum Service

Linnean Society of London

Liverpool Record Office

London Metropolitan Archives

London Transport Museum

LSE Library, London School of Political and Economic Science (LSE)

Manchester Art Gallery

Manchester Jewish Museum

Manchester Museum and Whitworth Art Gallery

Mary Rose Trust

Media Archive for Central England

Modern Records Centre (University of Warwick)

Museum of Archaeology and Anthropology (University of Cambridge)

Museum of Cornish Life

Museum of Domestic Design and Architecture (Middlesex University)

Museum of East Anglian Life

Museum of Freemasonry

Museum of Lead Mining

Museum of Scottish Railways

Museum of the Home

Museum of Zoology (University of Cambridge)

Museums and Galleries Edinburgh

Museums Northumberland

National Archives

National Football Museum

National Galleries Scotland

National Gallery

National Holocaust Centre and Museum

National Justice Museum

National Library of Wales

National Mining Museum Scotland

National Museum of the Royal Navy

National Museums Liverpool

National Museums NI

National Museums Scotland

National Portrait Gallery

National Tramway Museum

National Trust

National Trust for Scotland

National Waterways Museum, Canal and River Trust and Waterways Archive

Natural History Museum

Norfolk Museums Service

Norfolk Record Office

North East Wales Archives (Hawarden)

North West Film Archive (Manchester Metropolitan University)

Northampton Museums and Art Gallery

Northern Ireland Screen - Digital Film Archive

Nottingham City Museums and Galleries

Orkney Museum

Paxton Trust

People's History Museum

Petrie Museum of Egyptian Archaeology (University College London)

Pier Arts Centre

Pitt Rivers Museum (University of Oxford)

Poole Museum

Portable Antiquities Scheme

Postal Museum

Potteries Museum and Art Gallery

Preston Park Museum and Grounds

Public Record Office of Northern Ireland

Queen Mary University of London

Queen's University Belfast

R and A World Golf Museum

Richard Burton Archives (Swansea University)

Ripon Museums Trust

Royal Academy of Arts

Royal Academy of Music Museum

Royal Air Force Museum

Royal Albert Memorial Museum and Art Gallery

Royal Armouries

Royal Artillery Museum

Royal Botanic Gardens Kew

Royal Collection Trust

Royal College of Music Museum

Royal College of Surgeons of England

Royal Commission on the Ancient and Historical Monuments of Wales

Royal Engineers Museum

Royal Holloway, University of London

Royal Horticultural Society

Royal Museums Greenwich

Royal Scottish Academy of Art and Architecture

Salisbury Museum

Science Museum Group

Scottish Fisheries Museum

Screen Archive South East (University of Brighton)

Sedgwick Museum of Geology (University of Cambridge)

Sheffield Museums Trust

Sir John Soane's Museum

SOAS University of London

Southampton City Council Cultural Services

St Helens Council Library Service

St John's College Library, Cambridge

Staffordshire Archives and Heritage

Sunderland Museums

Tank Museum

Thackray Museum of Medicine

Tullie House Museum and Art Gallery

Tyne and Wear Archives

Tyne and Wear Museums

University College London, Art Collections

University College London, Pathology Collections

University College London, Science Collections

University of Aberdeen

University of Bristol - Art Collection

University of Bristol - Earth Sciences Collection

University of Bristol - Theatre Collection

University of Bristol Library Special Collections

University of Bristol Spelaeological Society

University of Cambridge, Herbarium

University of Edinburgh

University of Exeter

University of Glasgow Archives and Special Collections

University of Leeds

University of Leicester

University of Nottingham Manuscripts and Special Collections

University of Oxford, Museum of Natural History

University of Reading

University of Sheffield Library

University of Southampton

University of St Andrews - Archives

University of St Andrews - Library

University of St Andrews - Museums

Victoria and Albert Museum

Vindolanda Trust (Chesterholm Museum)

Wakefield Museums and Castles

Wallace Collection

Watts Gallery Trust

Weald and Downland Living Museum

Wellcome Collection

West Glamorgan Archive Service

Whipple Museum of the History of Science (University of Cambridge)

William Morris Gallery

Wiltshire Museum

Wordsworth Trust

York Museums Trust

Yorkshire Film Archive

Appendix B: Institutions included in the technical survey

The following list includes the 27 institutions that participated in the technical survey. Note that Museum of London Archaeology took part in the technical survey, but it was agreed that the main audit survey was not relevant to them so they did not submit information to that.

Amgueddfa Cymru (National Museum Wales)

British Film Institute Archive

British Library

British Museum

Historic Environment Division, Northern Ireland

Historic Environment Scotland

Historic Royal Palaces

Imperial War Museums

Museum of London Archaeology

The National Archives

National Galleries Scotland

National Gallery

National Library of Wales

National Museums Liverpool

National Museums NI

National Museums Scotland

National Portrait Gallery

National Trust

National Trust for Scotland

Natural History Museum

Public Record Office of Northern Ireland

Royal Botanic Gardens Kew

Science Museum Group

Royal Collection Trust

Royal Commission on the Ancient and Historical Monuments of Wales

Tate

University of St Andrews – Library

Appendix C: Roundtable participants

The following people took part in the roundtable discussions held to gain insights into the capacity of small archives and museums to be part of a national digital collection. All comments have been anonymised in the write-up of these sessions.

17 January 2022, Cornish museums

- Janine King, Castle Heritage Centre, Bude
- Sue Andrews, Lostwithiel Museum
- Michael Furminger, Lostwithiel Museum
- Phil Jones, Cornwall Museums Partnership
- Stephanie Clemens, Cornwall Museums Partnership

21 January 2022, Small archives 1

- Christopher Hilton, Britten Pears Arts, Aldeburgh
- Mary Ellis, Welsh Government
- Martin Wynne, Oxford Text Archive
- Lisa Di Tommaso, Morrab Library, Penzance
- Janneke Geene, Manchester Metropolitan University
- Also Diane Scott from the TaNC Discovery project *Our heritage, our stories*

21 January 2022, Small museums 1

- Lu McNair, North Lanarkshire Council
- Danielle Sellers, Royal Engineers Museum, Gillingham
- Dan Britton, Royal Engineers Museum
- Samantha Walker, St Andrews Heritage Museum and Garden
- Matthew Bellhouse Moran, HMS Unicorn, Dundee
- Jane Freel, Fife Cultural Trust
- Jacob O'Sullivan, Museums Galleries Scotland
- Raphaëlle Goyeau, Museum of the Order of St John, Clerkenwell

21 January 2022, Subject Specialist Networks

- Crispin Paine, Religion, Collections and Heritage Group
- Anna Reeve, Classical Collections Network
- Emma Halford-Forbes, Industrial Museums Scotland
- Paolo Viscardi, Natural Sciences Collections Association (NatSCA)
- Paul Hermann, Photographic Collections Network
- Hannah Jackson, Rural Museums Network
- Gail Boyle, Society for Museum Archaeology

26 January 2022, NI museums

- Claire Hackett, Falls Community Council, Belfast

- Damien Maddalena, Eluicidate

- Sinead Reilly, Fermanagh County Museum

- Bronagh Cleary, Fermanagh County Museum

- Robert Whan, Armagh Robinson Library

- Sarah Carson, Causeway Coast and Glens Borough Council Museum Service

- Jamie Austin, Causeway Coast and Glens Borough Council Museum Service

- Rhiann Coulter, Armagh City, Banbridge and Craigavon Borough Council

- David Farrell-Banks, Deborah White and Elizabeth Crooke, Ulster University

25 January 2022, Small archives 2

- Matthew McMurray, Royal Voluntary Service, Swindon

- Michael Gunton, Portsmouth Library and Archives Service

- Robin Sampson, Norfolk Records Office

- Louise Pichel, Museum of Freemasonary, London

- Juulia Ahvensalm, St George's University of London

- Lin Rice, Knowsley Archives

- Karl Magee, University of Stirling

- Gillian Fewings, Plymouth Marjon University

- Gary Collins, Coventry University

- Also Lorna Hughes and Diane Scott from the TaNC Discovery project *Our heritage, our stories*

25 January 2022, Small museums 2

- Bette Baldwin, Clevedon Pier and Heritage Trust

- Sally Lewis, Association of Performing Arts Collections

- Kate Hebditch, Nothe Fort, Weymouth

- Victoria Rogers, Welsh Government

- Chris Martin, Glenesk Folk Museum, Brechnin

- Elizabeth Selby, Dorset Museum

- Emma Halford-Forbes, Industrial Museums Scotland

- Sarah Burry-Hayes, University Museums in Scotland

- Christine McLean, Fife Cultural Trust

- Stephanie Boydell, Manchester Metropolitan University Special Collections Museum

- Jayne Evans, Manchester Metropolitan University Special Collections Museum

- Scott Byrne, Hospitalfield, Arbroath

Towards a National Collection: digital collections audit

Collections Trust is conducting a UK-wide audit of digital collections on behalf of the [Towards a National Collection](#) research programme (TaNC), funded by the Arts and Humanities Research Council. The main aim of the audit is to understand the number, scale and attributes of digitally-accessible collections across the UK cultural heritage sector that might form part of a future national digital collection infrastructure.

This document includes all the questions found in the online audit survey form. Note that certain answers will send you straight to the next section, so you probably will not have to answer every single question.

*We recommend **using this document as a guide** while you assemble all the answers before starting the online form. A cookie should save your answers to your browser's local storage if you can't get through the form in one go.*

If you are not sure how to answer any of the questions, please contact Gordon McKenna at Collections Trust. And thank you, in advance, for your help with this important audit.

Fields marked with an * are required

1 Consents and contact details

Personal information *

☐

I understand that Collections Trust will use the personal information supplied below only if needed to contact me for clarification about any of the answers supplied, or for other reasons connected directly to the audit, and that Collections Trust will not share this personal information with anyone else

Corporate information *

If you are happy for the data you provide about your institution's digitised collections to be shared in full with the Towards a National Collection research programme, please select 'yes' below. If not, please select 'no' and we will only use the data provided in aggregated form within our report. We don't think any of the questions should pose problems of confidentiality; indeed, they are questions that might reasonably be asked of public institutions as freedom of information requests.

O Yes

O No

1.1 Your institution *

1.2 Your name *

1.3 Your role *

1.4 Your contact email address *

(Please let us have an email address we can use to contact you if we need to clarify any of your answers. We will not share this number with anyone outside Collections Trust, nor use it for any purpose not directly related to this survey.)

1.5 Your contact phone number *

(Please let us have a phone number we can use to contact you if we need to clarify any of your answers. We will not share this number with anyone outside Collections Trust, nor use it for any purpose not directly related to this survey.)

2 Item-level records

*In this part of the survey, we're interested in **item-level** collection records. By this we mean records about individual objects, documents, artworks, specimens, etc. If you are unsure about how to answer the questions in this section, please contact us.*

2.1 Do your collections database(s) include records down to the level of individual items? *

O Yes

O No *(Skip to section 3)*

2.2 How many catalogue records of individual items are in your collection database(s)? *

(If all your collection records relate to individual items, this number should be easy to get from your system. If your database(s) include a mix of item-level and group-level records, you may need to provide an estimate.)

2.2.1 Is your answer above *

O Exact?

O An estimate?

2.3 How many of these records have images (or other digital media) associated with them? *

(These images and other media might be stored within your collections databases, or elsewhere. This number should hopefully be fairly straightforward to get from your system. If not, please just provide an estimate. By 'other digital media' we mean video files, audio files, 3D scans, etc, that depict or are associated with items in your collections.)

2.3.1 Is your answer above *

O Exact?

O An estimate?

2.4 Approximately what proportion of these images or other media would you describe as 'publication' or 'broadcast' quality? *

(We know this is a subjective question. We're trying to get a sense of to what extent, for example, the images you have are quick, low-resolution snaps taken for record purposes or high-quality photographs taken with skill and care. It doesn't have to be exact: your best guess will be fine.)

2.5 Please give brief details of any current or planned projects likely to add a significant number of additional images or other media to your collections database(s).

(The number of images and other media you have noted above may not reflect work in hand or planned over the next couple of years. If you have any projects underway or in the pipeline that will add significantly to the existing numbers, please give brief details here. If possible, please include the number of images or other media likely to result from the project.)

3 Group-level records

*In this part of the survey, we're interested in **group-level** collection records. In archives, this would be everything above the level of individual documents. In libraries, it might include records about series of newspapers or magazines etc. In museums it might include records about bulk archaeological material or large numbers of natural history specimens. It might also include 'holding records' created as part of backlog projects. If you are unsure about how to answer the questions in this section, please contact us.*

3.1 Do your collections database(s) include records relating to groups of items? *

O Yes

O No *(Skip to section 4)*

3.2 How many catalogue records relating to groups of items are in your collections database(s)? *

(If all your collection records relate to groups of items, this number should be easy to get from your system. If your database(s) include a mix of item-level and group-level records, you may need to provide an estimate.)

3.2.1 Is your answer above *

O Exact?

O An estimate?

4 Records published on your website

*In this section we're interested in whether your **own** institution's website has any kind of online database that allows users to explore some or all of your collections. If you are part of a larger service with a website covering several branches, please answer with reference to that website if applicable. If you are unsure about how to answer the questions in this section, please contact us.*

4.1 Are catalogue records (individual or group) published on your own website? *

O Yes

O No *(Skip to section 5)*

4.2 How many catalogue records are published on your own website? *

(In some cases, it might be clear from the website itself how many records there are in the online database. If not, you should be able to find the number somewhere in the back-end. Your system administrator or web developer may be able to help.)

4.2.1 Is your answer above *

O Exact?

O An estimate?

4.3 How many catalogue records on your own website have images or other media? *

(Again, the website itself might say how many records in the online database have images or other media. If not, you should be able to find the number somewhere in the back-end. Your system administrator or web developer may be able to help.)

4.3.1 Is your answer above *

O Exact?

O An estimate?

5 Records published elsewhere

*In this section, we are trying to establish how many collection records are **only** published on sites such as Art UK, Google Arts & Culture, National Biodiversity Network, etc. If **all** the records you publish on such third-party sites are **also** published on your own website, please answer **'no'** here. Once again, if you are not sure how to answer the questions in this section, please contact us.*

5.1 Are any of your catalogue records only published elsewhere (that is, not also on your own website)? *

O Yes

O No *(Skip to section 6)*

5.2 Where are these catalogue records published? *

(Please briefly state where these records are published, eg, Art UK, GBIF, etc.)

```
```

5.3 How many of your catalogue records are only published elsewhere? *

(We hope this should be a relatively straightforward to find, but an estimate would also be fine.)

```
```

5.3.1 Is your answer above *

O Exact?

O An estimate?

5.4 How many of these records published elsewhere have images or other media? *

(As before, by 'other media' we mean video files, audio files, 3D scans, etc, that depict or are otherwise associated with items in your collections.)

```
```

5.4.1 Is your answer above *

O Exact?

O An estimate?

6 Copyright

In cases where your institution owns the copyright to content (or has licensing agreements with copyright holders that allow you to use their content), please indicate your general policy for re-use of this content by others. You can tick more than one box (eg you might allow free non-commercial re-use and also commercial re-use for a fee). As before, if you are not sure how to answer these questions, please contact us.

6.1 What is your general policy on how people can use images, etc, for which you own the copyright? *

(You can tick more than one answer here.)

☐ Unrestricted re-use allowed

☐ Free non-commercial re-use allowed, with some conditions (eg attribution)

☐ Commercial re-use allowed for a fee

☐ No re-use allowed.

☐ Other

6.1.1 If you ticked 'other' please give brief details here.

6.2 Are any of your catalogue records available online only behind a pay wall? *

(For example, some archival material is published on family history websites that can only be accessed by paying subscribers.)

O Yes

O No

O Don't know

6.2.1 If you answered 'yes' to 6.2, please give details of the relevant paywall-protected website(s).

6.3 How many of your catalogue records are only available behind a paywall? *

6.3.1 Is your answer above *

O Exact?

O An estimate?

7 Technical stuff

If you don't understand these questions, don't worry. The chances are the answer is 'no', but please discuss them with your system administrator or other technical colleagues or suppliers. 'Don't know' is also fine as an answer in this section. We are trying to get a sense of the awareness and use of these technologies beyond the bigger institutions - or not - so all answers are helpful. If you are not sure how to answer these questions, please contact us.

7.1 Do you have an API that allows others to make use of your online collections? *

(API is the acronym for [Application Programming Interface]. Some institutions have these as a way of sharing their collections data with other websites, etc, automatically. It's the sort of thing developers would use, rather than the general public.)

O Yes

O No

O Working on it

O Don't know

7.2 Are any of your online images IIIF-enabled? *

(IIIF stands for [International Image Interoperability Framework]. This is a set of open standards underpinning tools that allow the sharing and annotation of high-resolution images, with features such as zooming.)

O Yes

O No

O Working on it

O Don't know

7.3 Do you make use of external identifiers for your online records, to help minimise the problem of broken links? *

(We're thinking here of systems for assigning unique references to digital records that are governed and managed by bodies other than your institution. For example, the [Digital Object Identifier] system.)

O Yes

O No

O Working on it

O Don't know

8 Anything else

8.1 Please use this box for any other information you think is relevant to the audit, or any caveats or qualifications you want to make regarded the information provided elsewhere. Thank you. *

Appendix E : Technical survey scope and questions

Scope

Area	Purpose
Format	Understanding types and formats of digital resources made available to researchers and public users.
Metadata	Understanding what types of metadata are created, the schemas and standards used, and how organisations manage the relationship between digital assets and physical artefacts.
Asset management	Understanding systems, scope, and service models for managing digital assets and associated pipelines for asset delivery to researchers and public users.
Access and availability	Understanding how organisations control access to digital resources, levels of access, and the capacity for machine-based enquiry.
Rights management	Understanding how licences and usage rights are assigned to digital assets.

Questions

Area	Core questions
Format	What are the key digital asset types (images, video, audio, documents, databases, web pages etc.)? What key file formats are used? Any specialised or proprietary formats? Are images IIIF compliant? If so, to what level (0, 1, or 2)?
Metadata	Do you generate metadata for digital resources? Is this made available for search and discovery? What schema(s) are used? Do metadata schemas include technical metadata about the assets? Do metadata schemas include descriptive metadata about the assets? (eg. data injected from a collection, library or archives management system) Do metadata schemas include any administrative metadata about the assets? (eg. source, ownership, access rights and restrictions etc.) Do you embed any metadata into the digital assets? If so, to what extent? Does this include any descriptive metadata? Is this a one-off activity, updated periodically or generated on demand (at the point of use)? Do metadata schemas include persistent references (URL, URI) to a range versions of a digital asset (size, format, resolution etc.)? Do metadata schemas include persistent references (URL, URI) to high quality assets (such as high resolution images, video or audio)? Can descriptive metadata be used to determine if an asset is a part of a group, set or sequence (such as pages from a manuscript, multiple images of the same artefact etc.)?

Area	Core questions
Asset management	How do you ingest, store and manage your digital assets?
	Are systems, services and resources centralised or departmental/project based?
	Do you use a DAM system to manage the overall asset lifecycle? (ingest, store, manage, transform, preserve etc.)
	Who is responsible for managing and publishing online digital resources? (IT, Library, department, project etc.)
	Are the online assets and formats generated and published automatically?
	How do you integrate digital resources with material resource data? (objects, archives, library etc.)
	What mechanisms do you use to deliver digital resources online? (eg. DAM workflows, bespoke scripts and pipelines etc.)
	Are processes fully automated and consistently repeatable?
	How much of the process relies on manual transformation of digital assets? (eg. changing file formats, resizing etc.)
	If you are generating IIIF images how are you creating manifests? Are they fully automated (using a service), partially automated or completely manual?
Access and availability	Do you control access to your primary online digital resources?
	Are there additional controls/authorisations required to access high quality digital resources?
	Do you use a controlled API or other form of gateway service to determine access to digital resources?
	Does the API/gateway allow for different levels of access to your digital resources?
	Does the API offer a range of methods to facilitate resource discovery? (ie. includes sets of technical, descriptive and administrative metadata for each digital asset)
	Does the API use bespoke or standards-based schemas for resource metadata?
	Does the API include persistent references (URL, URI) to cloud-based resources?
Rights management	How do you assign access/usage rights to digital assets?
	Are rights/licensing restrictions managed at the individual asset level or for asset/collection groups?

Appendix F: Roundtable quotes

This appendix presents extended, anonymised quotes from the roundtable discussions with smaller museums and archives, and with representatives from some of the Subject Specialist Networks. These conversations were wide-ranging, but the quotes are loosely grouped under the broad questions that prompted them. Beyond that, they are in no particular order.

The quotes are lightly edited to enhance their readability, but reflect the authentic voices of a range of people working at the coal face of collections management in smaller institutions.

What is the current state of your collections documentation?

"We're a charity archive and our resources are somewhat limited. These days there's just me working four days a week trying to do everything from the cataloguing to new accessions to mending the plumbing in the kitchen. About 7% of our collection is catalogued to ISAD(G) standards and the rest is merely box listing, which is about as useful as a chocolate teapot.

We did do a couple of projects a while ago creating pages worth of digital content involving some of our wartime records, and those are available through a CalmView site we have.

The use of that material is somewhat limited because people don't necessarily know we exist in the first place, or how to come across that data, and for us that's an issue. We don't have the resources to, for example, transfer all that material to Discovery."

[Archive]

"Truthfully, I don't really know [if there is a backlog of undocumented material]. I don't think so. Our archives are in a store and we're slowly working through them. Remember it's a volunteer-based system; I don't have staff and your process for then computerizing the paper-based records is scanning and OCR-ing, it is typing and transcribing. I inherited work done by others some time ago and it's conscientious, it's complete, it is good work. It's just that sometimes you've got holes in it. Our current challenge is around things like where loans have been made, and making sure that's been handled correctly and properly. Items which have come to the museum and not been fully documented? There probably are one or two. If they're of obvious value, yes: definitely documented, definitely accessioned correctly they put onto the database. If it's something that isn't really of huge value but we don't want to say no, for all sorts of reasons, then it's probably up in the office ready to be filed."

[Museum]

"We're an arm's length trust, now, from a local council, and used to be three separate council areas which had come together in the 1990s... We've got information on the database largely at inventory level. We don't always have the entry information attached; it's on paper still, a lot of it. Also different types of collections are better documented; so the art is really well documented, it's really well photographed, mainly because of Art UK. But other things, like the natural history collections are fairly sparsely documented."

[Museum]

"We have very small collection, probably only about 5,000-8,000 items. We are very small museum - two members of staff - and we are currently catching up after a period of next to no proper documentation. So in a way, we've got a clean-slate start. We're currently transferring a couple of thousand handwritten paper records that were done in the last two years before we managed to get

Axiell Collections, which is what we have now. We only have a couple of hundred records up there, but they're all fairly good quality because they've not been rehashed and sent through however many exports and transfers. And so we haven't got a very exciting collection for digitisation in that regard. What we do have, however, is the ship itself, which has been laser scanned two or three times over the last five years, and has had structural analysis and digital wireframe modeling done to an equivalent standard to HMS victory. So we might be looking at ways to share that."

[Museum]

"Unlike some of the other museums, we have very large academic teams in some of the university museums. In some cases the collections databases are quite extensive, but the level of information available on the collections is, in some cases, quite limited. We're quite privileged to have quite good provenance knowledge about a lot of our collections, but that research is an ongoing thing within Scotland. There's a massive project at the moment about African Caribbean collections because obviously that's particularly topical at the moment, and what has been recorded about those objects may well change.

Having said that, we have smaller museums that don't have the level of documentation of the rest of the larger university museums. We still have half of the smaller collections not documented, not to the extent that potentially an external database might want, and certain audiences might want. Research audiences obviously need far more detailed information about collections than the public."

[Museum]

"I was brought in as the first paid curator we've had in years. And part of that was to fix a lot of the issues that we have with documentation and object accessibility. At the moment we have an Adlib system (which is itself a problem because Adlib is now being scrapped) so one of the issues we're going to face is changing to a new system. We are an independent museum (independent only for the last two years) so financially that is going to be difficult. At the moment we're facing an object database that has been variously put together by a series of well-meaning volunteers who did not have good database training."

[Museum]

"So, a lot of what I'm trying to do is consolidate object information and make sure it actually matches up with what is in the store. A secondary issue there is that our museum was knocked down and rebuilt and a lot of the object locations have not been updated. We've got locations that are pretty vague in the first place and relate to a building that doesn't exist anymore. And I am only now, for the first time, consolidating any paper documentation. There are various archives scattered around the museum: within the object collections, within special documentation put together for previous exhibits, or just pieces of research that people have done themselves and sent into the museum. So all of that is just around the building, there's no real organisational structure that is consistent, and it's a massive job to try and bring that under control."

[Museum]

"The Welsh museum sector doesn't have those large regional museum services. We go from the National Museum with its seven sites across the country to what would be termed a smaller local museum in England. Across that you've got a real mix of museums working on a paper basis, in spreadsheets, and then those with some proper 'collections management systems'. And even those probably haven't migrated all of the data that they've got. The data that's held within any of those systems is going to be talking to the museum person rather than in a way that will be accessible to the public or a non-museum person."

[Museum adviser]

The issue in my field is that small museums often don't have good quality data on their collections. They don't know what they've got, in effect. We're often talking about legacy collections from the nineteenth century that perhaps have never been fully identified and described, so there's little point sharing data when the data is of such poor quality that it can't be put to any particular use. From my perspective the first step is to improve the quality of the information that museums hold, and due to curatorial cut-backs and so on that requires funding to bring experts and collections together. That has manifold benefits: it means that data can be produced on what there is and where it is around the country. It helps the museum use its objects for connecting with audiences and interpretation, but it can't just happen. It's not really going to happen organically, so that's step one from my perspective.

[Subject Specialist Network]

"My collection is, I suppose, a bit different because it's an oral history collection. I couldn't say that I would be completely happy with it because there's a big backlog. It's partially catalogued, and it's insufficiently backed up. And also there's an ongoing issue around getting permission forums and stuff like that."

[Museum]

"I think from our point of view I think no matter what museum you work in, there's always a museum backlog. And, and we do have a backlog. We have just actually got a new collections management system. It's work in progress. It's a fantastic system. Improvements? In an ideal world, I would love, love to have every single object photographed in the collection so that we can share more online. I'm happy with the system, but in terms of a backlog, yes, if we had more time, maybe more staff resources, that'd be great in terms of trying to tackle that."

[Museum]

"Historically, some of our collections would have come in a big bulk. For example, we had a fabulous folk life collection coming in … and when you're registering 1000 - 2000 objects that are suddenly coming into your museum, your descriptions, the information is very basic. We will have lengths and dimensions and the basics, but we don't necessarily have that detailed caption text that will engage an audience. And the ones we do have that for are items that have part of an exhibition, where caption was written, and added [to the database] at a later date."

[Museum]

What is the current state of your collections management system?

"I'd say [our system] generally is pretty good. We've got a paper-based system as well but we do use Modes Complete. We upgraded a couple of years ago. We've always found it OK and I've got archive volunteers who are very experienced. However only one or two really know how to use it and if I'm honest we probably don't use half of it, because there's so much more we could do. We just use the basics but I think our whole collection is pretty well catalogued."

[Museum]

"I'm the [volunteer] curator. I've been doing some work bringing their collections from being exclusively paper-based onto digital systems within the museum, and also trying to encourage a more interactive online presence. I come with fairly significant IT background; I'm not really a museum person as much as a techie, so [my knowledge of] technology is fairly adequate.

I came to the museum about two years ago. They had serious technical issues: they had a broken computer; they had Modes [Complete] as the database system, but that hadn't been used for some years. So resurrected the system and tidied it up. It's there and it's working. It does hold the catalogue-able information. I had a learning curve myself as I didn't thoroughly realise when I started that the primary database is in fact paper-based (of all museums, I'm told). We have a selection of record systems: paper-based systems; the card-based system; there was the old computer; there's a newer computer; and so on.

Modes itself is incredibly clunky. It's a very arcane piece of client server database with a HTML front-end screwed onto it badly. It does follow the Spectrum standards and it's understood by the community, but it is not in any way ideal for a tiny museum (we're a micro museum). We have a single user license: one PC with a standalone version that we have to basically move on to other PCs when we want to do some field-based work. It's clunky but we don't have the kind of budget to invest in a full-blown system. It just wouldn't financially make any sense. However, because we're a group of volunteers that work remotely we can't really share the same PC it just doesn't work. So one of my wishes would be that we had a proper online based system for logging and sharing data, logging our own data as well as sharing with others. That is wish number one of many."

[Museum]

"This is pretty representative of what's been going on. Basically the information is there but it is still not as accessible as it needs to be for people to really use it. It's the spectrum from completely paper-based, and having to look up files and things, to museums maybe who've managed to get the basic catalogue onto Modes so it is retrievable. I'm hearing exactly the same things everywhere I go to a different degree. In-house, people are generally using Modes and nobody likes it at all, no-one says it's fit for purpose.

Modes is so flexible you can make it do exactly what you want. When you want to pour that into something, or some kind of consistent framework, that's when it starts falling down. Also, when you want to move Modes data between installations it doesn't seem to work at all. So you're constantly being forced to work in your local Modes bubble. The interfaces in and out of Modes are very poor, but as far as actually capturing data in immense detail in a very flexible way, if you had easy ways to make that happen then the actual database structure itself is excellent. They've thought of everything; they just haven't thought about how to get it in and out."

[Museum advisor]

"We didn't inherit any collections management software or procedures, or any documentation really, when we took on [the site] about six, seven years ago. That's changed and we have done a lot of work in that regard. We couldn't afford to buy proprietary software to manage our collection so we went down the open source software route, and we found that to be feasible and useful and actually much easier. So there's an opportunity for those organisations who, like us, work with an Excel spreadsheet, or through an Excel spreadsheet, to migrate it to something that's cross-referenceable and more accessible for researchers and the public, by looking at open source options as well."

[Museum]

How do you deal with images?

"There's two different sets of images. First of all there's the image library associated with the Modes collection database. The images there are actually an illustration of an entry for an item in the collection. Separately to that the museum also has just under 3,000 actual digitised images of photos of life in [X], going back to the 1860s, whose value to the museum is as a digital image rather than a physical object.

The first set is all captured as jpegs attached to entries in the Modes database. It is just jpegs, nothing fancy. It's searchable in as much as it's in the database already. The thing I'm particularly working on at the moment is the [digitised] photographs. They are part of the museum collection which only exists as an image; there's no physical object associated with them. This is something that people doing exhibitions or the curator or anyone trying to work with the museum information really needs access to - in a searchable way, because a directory of jpegs with nothing to help you identify what they are is absolutely useless.

I'm ploughing through those three thousand images at the moment; obviously they need to be searchable so that people [can] see all the pictures of such-and-such an event, or all the pictures of women in WWII doing ambulance work, or all the pictures of the local bowling. I'm trying to find a way that you could tag all of these photos so that they would become searchable in the way that big organisations organise their image collections."

[Museum]

"We've got some pictures on Modes but not every object has had a picture uploaded. That doesn't necessarily mean it hasn't got a picture, just that it hasn't been attached to that file, and that could be due to lack of knowledge and confidence.

A few years ago one of our volunteers photographed all of our pictures (we have quite a large collection of paintings) and she actually physically printed them off on A4 sheets of paper and documented them that way. We still use that system if we're trying to look up something quickly. There are other pockets of photographs, part of the collection that might be on a spreadsheet somewhere, so it's slightly chaotic if I'm honest.

Several years ago we were part of a project which resulted in a website called cornishmemory.com, that was a digitalisation of our entire photograph collection, which is really useful. Quite often I use it myself to find things and search for things. Not only did they digitise all of our photographs (quite a large collection) but also transparencies and negatives and things like that. So we do have that as a resource in the public domain, but in-house our images collection isn't quite as concise as it could be."

[Museum]

"We're trying to categorise our image database, which is a collection of photographs, some of them digital, lots of them on paper. We want to integrate that into the touch screen system I developed some time ago to make things interesting when you go into the museum if you want to extend your knowledge. This is not the same as an online based system because that is obviously widely shared. With this the idea is you're inside, you want to learn a bit more, you want to be engaged a bit longer."

[Museum]

"And that's before you start asking about the images, audio, video. I think there is a lot of non-management with this material – I'm pushing for DAMs ... because we have c.100k images just of collection and hours are lost searching for images. I worked on a funded project for another body and they have lost all my tiffs because they didn't know they mattered or perhaps IT thought they didn't need them or failed to move them as they changed infrastructure…who knows but now they have very few images of an appropriate quality to use on their new collection website. Fundamentally, if no one understands how to manage or pays attention to this material, it is lost very easily."

[Museum]

"There's a lot of pressure, particularly after lockdown, and a lot of assumption that it's very easy to create digital images of your collections and put them online. Of course it isn't. These things take time."

[Museum]

"In terms of photographs as well like we might have a record to indicate a particular photograph, but the photograph hasn't been scanned, you know, whether it's negative or a slide. We would love to be able to go back. Yes, we have a record, we know where the slide is physically, but in order to get the slides scanned or the negative scanned or the image scanned, we have a lot of work to do in that regard."

[Museum]

How do you deal with copyright?

"Our organisation has been going for a very long time, entirely volunteer-run, so we didn't actually have any staff. So in theory [copyright in] every single thing in our archive belongs to the individual who wrote it, which is a big nightmare. We tend to try and slightly ignore it."

[Archive]

"People want to use our images and a lot of them are orphan works. Lovely as the [Orphan Works Register] is, it's so labour-intensive to get the required level of evidence to register an orphan work (for which you have to pay more for the licence) that it's actually not worth doing. If there was an orphan works programme which was basically insurance (you pay a small fee and then you'll be covered if you license something which is owned by someone else) that would make a million times difference, but I don't think anybody's going to do it."

[Archive]

"I've found Art UK's copyright thing has taken a huge kind of weight off my shoulders. When enquirers get in touch I can just direct them to Art UK, I don't have to try and find out myself."

[Museum]

"I think it's not just the what-is-it kind of data, but in terms of copyright - and where the images have come from - that's a whole chunk of data that hasn't historically been recorded. That [kind of information] is difficult to pick up again so I think there's huge anxiety over copyright issues as well."

[Museum advisor]

How easy is it for people within your organisation to access and use collections information?

"We have a regular problem: we have a window and there's a group that would like to have a new exhibition in it every two or three weeks. They want to dip into local activities or into what we have in the archive to bring things in. It's very difficult. We don't do it very well; we haven't got a tool which is appropriate to the needs of the community and it's as simple as that. Modes has everything in, we just can't get the data in or out. It's just very, very old-fashioned technology."

[Museum]

"Another barrier to access, even within institutions, arises when only a limited number of users are allowed to use the collections management system simultaneously because of software licensing restrictions."

[Museum]

"A couple of years before I joined the Trust a non-museum member of staff and trustees wanted to put all of our collections online. Although our records are OK, they do need tidying up, but that process was never thought about because, at the time, there were no curatorial staff at the museum. So there was this push to put a lot of collections online … but if I go through the records I cringe because there are typos or there's information that's not public facing. Like why would you have the dimensions of a photograph in the description? That type of thing, because none of the records were cleaned up. So I guess I'm a bit sceptical about having our database completely accessible by everyone within the organisation because I feel like it just can open up quite a few issues."

[Museum]

"On the one hand, everybody who's in the museum team can access [our database] although the level of confidence of users is variable. I'm probably the member of staff who knows how to use it the most, but there's others who will use it occasionally and, because it's complicated, they're like, 'what do I do?', 'how do I do this again?'

We'd like to expand it to front of house staff so they can see things, and all the comms department and marketing to get images, but licensing is an issue: we only have 15 concurrent users, so we're just conscious we don't want to reach that. And also, we probably would want to tighten up a little bit, just so people can't get into donor information; it wouldn't be appropriate for comms to do that."

[Museum]

How easy is it for you to share collections data with people from outside your organisation?

"It's pretty close to impossible [for anyone who wants access to collections information to get it]. I would sit down with whoever wanted to find something out show them how to use the database. I tend to focus on the image library which we've got tidied it (we had a young student-type person to do the work). They can have a look at that but they tend not to, they want to handle things and not data, so it's usually with help and local knowledge stuff is identified."

[Museum]

"To be honest the majority of exhibitions would be my responsibility anyway, along with the volunteers, so it would be quite rare for somebody to come in who really didn't know anything to look at our system. But if that was to happen it would be difficult. I or one of my volunteers would have to sit down with them and physically find the information that they needed. We wouldn't just be able to say here is a password, off you go. It would be far more complicated than that."

[Museum]

"We've recently moved to a centralised store, which means it's a lot easier for us to get researchers and enquirers in, and we've been working with volunteers on various projects. It's been fantastic because they've got expert knowledge about pottery collections, or archaeology, that we just don't have as staff. We're doing it on a very personal level at the moment, but if we can get the data out there and tap into expertise from wherever that be fantastic. And getting something back is always helpful as well."

[Museum]

"I started out as a researcher I did my masters and doctorate on museum collections. One was Perth Museum and Art Gallery, with its Northwest Coast collection, Recognized for its national significance. I was only able to do that because the amount of information that they had in place already was sufficient to build a project around. And since coming [to my present museum] we've had people come along, interested in building projects around our collections, who have pretty much turned and run away, because when ask what we have on X, you say "you're gonna have to look for it, because it's somewhere in here. Maybe."

And so we are not appealing as a research location, because if you go to AHRC and say I want to build a doctoral project around something that may not even exist when I get into it, they're going to laugh you out the building."

[Museum]

"Universities who deal with archaeological teaching across the UK have an organisation that surveyed how archaeological collections had been used for PhD and post-doctoral research between 2010 and 2020. It was a desktop survey. They came up with some really useful advocacy for how people were accessing information and also what the level of financial investment was in that research process. One of the most useful findings was that researchers were not referencing the material properly, and they weren't acknowledging that use. Which comes back to 'who are the users?' One of the biggest requirements was that people wanted to get that global sense of where stuff was, and access it from one portal – and those tended to be the academic researchers who didn't want to go and physically visit museums, or go through 20 or 30 different types of databases."

[Subject Specialist Network]

"If anybody's working on an exhibition and they request [something] either we do the search for them, or we can get a day when they can come and sit down and go through our Modes [database] as well to do their own research on the topic. We also have the Northern Ireland community archive, which I suppose is how we're trying to make our collection accessible online. There's a lot needs fixing on it, it could do with being modernised, but that's how we're trying to show what each of our museums hold, what collections they have, and then we get requests through from people wanting to see objects or get more additional information. So that's been beneficial to us because, as the others were saying, unless people physically come in and go through the collections they're not going to see it or not know what you have."

[Museum]

"We did a bit of work with [another] museum service a few years ago, but we would have relied on them to send us through any relevant collection stuff. We wouldn't have had access to their database as such. So we would have been relying on each other to send through the relevant stuff, as a Modes PDF [export] or something, from a search. So it was a bit held back."

[Museum]

"I'm part of the Sporting Heritage Network, and that was something that was raised there: knowing what other sporting material was in the other museums. Other museum collections could have items that we could loan for displays, but we just don't know where they are."

[Museum]

"In a small place like Northern Ireland, there are personal relationships, so if a [need] arises I would be able to contact the relevant people in local museums. Maybe it's a non-sophisticated system, but I would do that quite frequently."

[Museum]

"A group here put together a memory map of their community, it's deep mapping. It's a physical geographic map on to which you're dropping objects, and history about them, people talking about them, and it's all given a geographic reference. They won at the Archaeological Awards this January, for the best engagement project but also the outstanding award."

[Museum]

"I would get constant research queries, on average, probably one a day. And there are different smaller discrete platforms; Art UK is as an obvious one, and yesterday we had a gentleman come in to see something he'd seen on the Art UK website. But there are other queries, you know, 'what do you have about.. a place, a family, a subject?' And that involves a bit of work on our end and creative querying of the database to give the person the answer."

[Museum]

Do you have any kind of higher level finding aids and collection level descriptions?

"Not really, not that would be useful to people. There would be in some sense if you found our website because you were researching the Women's Royal Voluntary Service. Then you would eventually come through to me and we could have discussion about what we had, but to the casual researcher it's very difficult to find us."

[Archive]

"We've got at least collection-level descriptions online which people should be able to find, although we've got an accessioning backlog, so there's an element of catching up there. Where we fall down is a lot of our older catalogues aren't ISAD(G) compatible and so there's a whole effort to re-catalogue those to modern documentation standards.

We're at the beginning of our digitisation program. We're awaiting delivery of an ingest machine, we've got a digital asset register, and so we've got a fair idea of what we've got in terms of born-digital and surrogate material. Some material is catalogued into Calm and is viewable online. The photographic collection is pretty well described in Calm, but where we've got other collections which form their own little digital collections, they're still just within the digital asset register, so that's not visible. We're just trying to find our way really."

[Archive]

"One of our major intentions is a mapping project which will try to identify the collections of religious material anywhere in the UK. Not only in … museums, but also in private collections, in places of worship, and so on. It's absurdly ambitious, of course. We certainly wouldn't aspire to go below collections level. There can scarcely be a museum in the UK that doesn't have items in its collections of religious interest, though they may be there primarily as social history items, as works of art, as 'ethnographic' [items], or whatever."

[Subject Specialist Network]

"For example, there's someone starting next working on something called the Blacker manuscripts, which are day books containing several thousand pages of manuscript material from the eighteenth century regarding the troubles in Ireland at the time. Nobody has comprehensively gone through the Blacker manuscripts, but it is one of those things that people will know, and there's a good harvest to be had in there if you're interested in that period of political history."

[Museum]

"In the past there was RASCAL [Research and Special Collections Available Locally], so I suppose our museum's greatest hits are available if you Google. There's a lot of hidden gems. A lot of that work was done in the 1990s as cataloguing has improved, and we've made new discoveries in the collection. There's a lot more relevant material that, unless you actually email, phone or write to me, you will not know about. I would feel very sorry for the average researcher, trying to negotiate their way around all these collections. I don't know what the answer is."

[Museum]

"Following the example of the Society of Museum Archaeologists, we looked at what capacity we have, what specialist knowledge we have. That, for us, is more important than any kind of mapping or getting collections data together. I would be keen for things like [writing] collection-level descriptors to be a skill that is developed across the UK."

[Subject Specialist Network]

How would you feel about sharing your collections data online?

"[Our collection] is 50 years old now and has been built up by volunteers with varying levels of skill. So we're not absolutely confident that our data is solid all the way, but personally I would really like to put it out there, because I think that's where we would be able to share and also occasionally show off. That would be a nice thing to do, but the technology we're using is very difficult for the audience trying to use it. By their very nature the people who do museums are not techies; usually they're historians, they're artists, they just don't do that stuff, so it has to be made as good as Google, which requires fairly powerful techniques.

I feel that the fears of 'wiki rubbish' are somewhat overstated. I've tried and tested it in the past, and it's difficult to corrupt it when enough people put their paddle in."

[Museum]

"We're all restricted on time to be able to [be part of something like TaNC]. I work 18 hours a week, I've got a whole list of jobs to do, and my archive volunteers are only here a couple of mornings a week. So whatever was to be set up would need to be something that could be done quick very quickly. I like the idea of having some sort of system which would highlight the key aspects of our collection, such as the work of Sir Goldsworthy Gurney, the Victorian engineer. He's of national and global importance. As well as the objects also having links to the people and stories behind the collection."

[Museum]

"Just getting money to do basic cataloguing is a nightmare but for us there's another issue terms of resourcing: while I would love to get more people to use our collections, I do not have the time - and the organisation does not have the resources - to employ someone to meet increased demand that would come from better knowledge of our collections. So if I put a load of stuff on Discovery that was driving people towards asking me questions, or wanting to come and visit, we couldn't fund that through the charity, especially as the charity doesn't have a purpose for archiving, education or history... So that's another big barrier in that we don't want to create demands that we can't fulfil."

[Archive]

"We've got some dodgy historical records, from 30 plus years ago, but we're gradually working through those and getting them up to standard. There's a good portion of them pretty well annotated, and we will publish those to an online catalogue, but only when we're happy with the records that are there."

[Museum]

"We're in a slightly unusual situation because basically our computerised records have started from scratch. The ones that we have are good, but we have next to none of them. We'd be really happy for them to go out, but they've not been prioritised by what's interesting, and at the moment it's just very obscure, paperwork documentation, archival stuff. So we'd be very happy to share all of that

data but it's not actually currently representative of the collection as a whole, and not actually enough of a proportion of the collection to actually be useful. Would it actually be perhaps a negative thing to reflect on institution if you put that forward as your contribution? People would look at it and go, 'well this isn't very impressive.'"

[Museum]

"I would say something we're debating within our team at the moment because we have very few things out there at the moment. We've got lots things on Art UK, and we've got a few things on our own website, maybe 200 objects, which for a collection 120,000 objects is very small. I think the general feeling amongst the staff is we have to be perfect. The reality is we're never going to get there, if we want to get it out there. I'm also part of the Scottish EMu user group and that's a discussion we've been having as well. 'Just get it out there and people will help us fix it' is certainly the approach of the special collections at the University of St Andrews. They have a 'can you help us improve this record?' button, so you can email them and say 'that's the wrong place' or whatever. We're in perfection mode at the moment but I don't think it's sustainable."

[Museum]

"I think the records, before they go online, just need to be checked for all the reasons that have been given, like spelling mistakes or the image isn't good enough quality. We have 27,000 records so we're never going to be in a position to put all those records online and for them to be perfect. Personally, I would prefer to check through each individual record before they go online, even if it's just so there's a basic provision of information there for people."

[Museum]

What are some of the barriers to sharing your collections online?

"We used to have a website, but now we don't. It's been absorbed into the bigger council one. Our problem really is access. The art collection is quite small and it's relatively well documented, but it's not accessible, unless people are in the building. How can we do that in a council context, working with IT guys who are not the most flexible in the world?"

[Museum]

"It might be just as simple as not having the bodies within the organisation itself to do the cataloguing to make the stuff available. It's that straightforward. I don't think it's even a technical [problem] because getting the stuff into Discovery isn't that difficult if your system can export information. Our cataloguing system can't do a handy export like that so I had to tinker around a bit but, in theory, it's fairly straightforward. But getting the stuff online is down to having someone in the archives actually cataloguing stuff."

[Archive]

"I'm sure everyone here has a huge cataloguing backlog and that's down to the staff time it takes, to devote that time and attention to cataloguing the collections. Once material is catalogued there are these portals and resources out there to share it. There is Discovery, there's the Archives Hub (which quite a lot of us use) and, in Scotland, the Scottish Archives Council are currently working on a brand new portal for all Scottish archives. So if you have catalogued collections the means are there to get them seen. It's [about] taking a step back to the collections that haven't even been catalogued."

[Archive]

"We are a library, and we have an archive and manuscript material, and also museum objects. So we straddle those different three areas, and they all have different cataloguing and documentation standards. So within museum we're working to Spectrum, in the library MARK, within the archives ISAD(G). Our catalogue and so on has been structured to the museum Spectrum standard, but that does sometimes create issues with us if we're looking to get our data into other larger databases (say, The National Archives' Discovery catalogue or what used to be called Copac). So people need to know about us to look on our website and search our catalogue. Whereas if it was interoperable and able to easily be compatible with some of those other database, that would be much more useful. Not a silo within your own organisation, but they can talk to each other in a way that we can't currently do. That certainly would be the one thing that I would be looking for a project to look into."

[Library, archive and museum]

"Being a military museum that was predominantly active throughout the Victorian period we do have some choice language in some of our records. So at the moment we are - well obviously not covering up that language - but we are looking at trying to make sure we're using appropriate language. If there is offensive language in there we want to reference that into the record as well, so that's another reason why we don't want to just push everything out once."

[Museum]

"The Scottish Transport Collections Knowledge Network machine tools projects took all the machine tools in public collections in Scotland and put them into a catalogue. It's still a very useful reference document, but within five years the website that hosted it, and all the HTML advantages of having it on a website, were gone and now it's just exists as a PDF because the money wasn't there to republish it and reconvert it for a new website."

[Museum]

"I think it's about what format the data would have to be presented in because I think most of the museums in question here have very limited curatorial staff, so time vs reward. I think it's great, putting data out there. Some of my curatorial colleagues are slightly twitchier about perfection, but I think putting data out there and getting involved in things would be brilliant. But the time required may be a sticking point for some."

[Museum]

"We are involved in the Heritage Information Access Strategy, and some of the datasets that Historic England have been looking at. I wanted to reference the rebirth of Heritage Gateway in a different format. What it does is harvest from other databases and provides access to records that way. When we first started engaging with the project the obvious question was "is there one point of access for all information about archaeological collections held in museums?" And the answer is there is not. There is some information that it held within Heritage Gateway about archaeological collections, mainly because they've been excavated and they appear on historic environment records, so HERs are obviously held by local authorities as part of the planning process in the sense that they need to keep that information. I can tell you that we've got very expert people running those systems and those systems are still falling short in terms of whether they are accurate or not – and accuracy is an issue."

[Subject Specialist Network]

How do you capture information arising from the use of your collections?

"People just walk into the museum and volunteer backstories, particularly about things like evacuees, but at the moment whoever is the current steward makes a note of it and refers it to the curator for consideration. There's no easy mechanism at all to capture information coming in in that way.

We do have what we call object history files, so anything related to a particular object (additional paperwork or anything comes in with the object or anything that is found out about it afterwards) does go in there. But it's all paper-based; nothing then gets transferred onto Modes, for example. One of my early challenges was I made the assumption that Modes was the primary database. It's not, it's the catalogue. The primary database is a paper written set of files with hand signatures. The idea of a digital signature is something that needs to be tackled at a national level. We've got to move on from the quill to at least the keyboard, perhaps the digital camera eventually."

[Museum]

"We have two sets of documents [for capturing knowledge about the collections]. The volunteer stewards make notes of anything coming in. These are physically handwritten notes, useful but impossible to handle. They really just move from the [front] desk to the team that tries to do things and then to the archive. They don't really get shared.

On the positive side we do have a biannual magazine [with a] focus very much on what's coming recently or what's a local topic of interest. I put a PDF on to the internal system, which could as easily then go on to a global system if that was practical. So we do have the basics of a searchable system of journals and magazines, which is our probably our sole tool for gathering good stories and then presenting them to the world."

[Museum]

"Because they're more paper-based, we don't really link anything back digitally, but we do try and capture something [from our exhibitions]. [Pre-pandemic] every year we'd have a very small archive exhibition on a local interest topic that would last for a couple of weeks, and every time we did try and encourage people to write down their memories. One was about the railway, for example, and we've got lots of lovely comments and memories and things about the railway but we haven't then digitalised it, it's all still in the book on the shelf."

[Museum]

"We have a module in our collection monitor called exhibitions, so I can take a look back over a dozen years and tell you what the objects were an exhibition, what the level text was for each object and attach PDFs, to the record as well, so you can have little exhibitions living in Mimsy you can draw and recreate."

[Museum]

"Within our media collections management system you can add in labels with whatever exhibition an artifact was on display. I would try my best to add in the label, but before my time that wasn't done, because people didn't have time, etc."

[Museum]

"It was my job to be sharing images of artifacts on social media throughout this lockdown, sometimes as mystery objects, sometimes just something that's interesting history. It was such a strength to go in [the database] and find the caption from an exhibition, compared to trying to wind

your way through descriptive text, which didn't give you that information in a nutshell. So it was priceless."

[Museum]

What is your approach to digital preservation?

"We've got a digital preservation system. It's very early days and it's basically just me doing it, but one of the concerns I have with mass digitisation is the perception from trustees that digitalisation is just popping something on a scanner and then you've got it, haven't you? It would be useful to have some support advocating for all the metadata that's needed with digitised assets, because we've had digitisation projects but we've ended up with a basic file list. I'm having to retrospectively put metadata together for projects that we've had done, which is rather time-consuming. You can't just give us a USB stick with a load of images on.

The [digitisation] we have done we hold on our own service. That's not very sustainable in the longer term. If there was a nationally-run Preservica which we didn't have to pay a small fortune for that would be massively important for us, because it's a cost that as a charity we find difficult to [afford]. I'm now running something like a terabyte's worth of information, which is more than the entire rest of the organisation has in digital assets, so that's an issue for us … There are no checksums going on or anything like that; it's just a server sitting there and there's a backup tape that runs every day and, once a month, the tape comes to my house with me. The IT department within the organisation don't really understand most of what I do, and within the rest of the organisation there is a very limited amount of digital asset management."

[Archive]

"We had an uphill struggle to persuade our IT people of the need for digital preservation beyond what they were they thought they were doing. We won that battle and we signed up with Preservica but we're just at the beginning of passing material to them and very wary of the cost involved in storing anything with them."

[Archive]

"We're really feeling our way and at the very beginning of taking steps towards digital preservation in my organisation. We are in talks with companies like Preservica and these conversations have actually been happening for a few years now. We might actually move towards putting the money in place and doing something in the next two years. I can't stress enough how slow everything is … I've got some pretty engaged colleagues but it's going to take a long time. It took seven years for us to finally get our library collections on Library Hub Discover, which I'm delighted about, and that's because I was able to delegate that to somebody working two days a week. I'm not a qualified archivist and the data is not as comprehensive as I'd like it to be even though we are on Discovery."

[Archive]

"Our [digital preservation] situation is actually quite good. We have Preservica at the moment because I'm working with our records management manager and our research data manager, so we made it into a project that's university-wide rather than just speaking for archives, and so we've managed to get money for that. Having said that it's mostly me and the records manager working on digital preservation and it's obviously only a tiny part of our jobs. We don't really have much IT support, so the IT department doesn't really have anything to do with these systems. That's something we'd like to change, but they only have so much capacity. We also have an online

catalogue for the archives and we have another research data management system, so now we're talking about how to streamline all of these how to make them work together."

[Archive]

"In a very small institution, serving a lot of different people, for me [digital preservation] boils down to two ways of doing it: DIY or get it funded. If it's DIY there's the difficulty of finding time, I have to re-learn how to do it all over again because I've forgotten since the last time I tried it, so it takes me longer than someone who really knows what they're doing."

[Archive]

"In recent years [digital preservation] has got better because we've started working with an IT company rather than trying to do it all in house, so there's backups done externally through this IT company - which obviously comes at a cost, but does give a bit more peace of mind. But this came up recently for me because we've had a large donation of about 5000 digital photographs, and they were handed to me on a hard drive by the donor. I thought to myself, I actually don't know what to do with it, like where am I going to store it? I just accessioned the hard drive like it's an object itself. We didn't really have any procedure for dealing with something like that. So I think in smaller museums that can be quite difficult, because it's quite a specialist area of collections ... It's something we're thinking about seriously, as we have other collections as well, like oral archive material, either on old-school tapes or VHS. Why do we still have this? It's making me panic a little bit."

[Museum]

"I'm sure any of us that have worked in smaller museums have actually had to Google our own things to find pictures of them, as opposed to looking at our own records, usually turning up on places like SCRAN, photos that were taken 20 years ago and you haven't got them anymore. What I am concerned about is that the only data that I would actually feel confident in its longevity at all, is stuff that's already based on 20-30 year-old standards which are just recognized by any device, such as tiff, jpg, raw text, that kind of thing. And what's happening is that we're moving, we've already seen it happen with Macromedia and Adobe Flash, which has disappeared off the face of the earth. It's happening now, with 3D data point clouds. There's not a standard way of storing that, and there's no guarantee that what we have now - which cost us hundreds of thousands of pounds to collect and collate - is going to be readable... It's not necessarily the confidence that I have in our ability to preserve those files, it's just the world moving on and nobody doing the digital equivalent of putting the videotapes through the digitiser … I think what would be useful is for some organisation to scan digital horizons for incoming new forms and types of data, or ones that are already few years old but there's not a set archival standard for keeping that data.

I feel like we've done quite a lot of scanning with images and things over the years, and I'm always slightly worried. Have I done it at the right DPI? In the right format? I've looked over the years for clear guidance and just never been able to find it, so something like that would just be hugely useful, so I don't have to sit and worry about that."

[Museum]

"The idea of the national collection is very appealing, because at the moment we're now moving into digital preservation. We have a lot of digital assets, and the intention is that the the digital images get loaded up on to Manage Your Collections, because they now have the facility to do that."

[Archive]

How easy or hard is it to get the resources you need for your work with collections data?

"Pre-pandemic I was involved in a couple discussions about digital skills where people were drawn from across the country and from different organisations. The one thing that really came home to me is that GLAM only / cultural organisations (regardless of size) hadn't grasped that in LG you are way behind social care, bins, street lighting and pot-holes priorities. Our senior staff may or may not be interested in GLAM and certainly are unlikely to understand it. We can be moved from corporate team to team – tourism, communities. leisure etc and occasionally end up in a group of services that don't belong in any other easily defined group."

[Museum]

"We got a grant from the Business Archive Council to train us in cataloguing because we have years of business records of running the pier. And we've used [TNA's] Manage Your Collections from the get-go, putting in information that might be relevant to the general public, to researchers, but also the information we need to manage the collections within the Trust. And that's been incredibly useful, and we feel very confident with it. The National Archives are extremely generous with their training and support. Three of them came and spent three days with my team, so that we were completely trained in how to do it. I would thoroughly recommend that. One of the reasons for doing it was because it meant we would go on to Discovery and our collections could be accessed by people from anywhere in the world."

[Archive]

"In some ways our project has circumvented [the problem of resourcing] by outsourcing some of the cataloguing. We work with a number of community archive groups across Norfolk and we act as consultants, providing them with training in various archival procedures. Cataloguing is obviously one of the things they're keen to work on. If they can catalogue their own collections to a basic ISAD(G) standard with the mandatory descriptions and mandatory fields then - if they, their collections or their groups fold - they can donate it to us in a catalogue-ready format, or they can upload the information themselves."

[Archive]

"We've always been keen to try and get support for cataloguing from whatever means available. We've been lucky enough to have been involved in a couple of AHRC-funded research projects, where we managed to get cataloguing of particular collections into the bigger project; rather than just isolating the cataloguing as being a separate problem it became part of the bigger research project, and academics and the archive staff worked together on opening up the collection, and so supporting the research outcomes of the project."

[Archive]

"I've found it's really difficult to get funding for cataloguing or digitisation projects, so we're in a similar situation to many of you. We basically just have one permanent member of staff, which is me, so I'm trying to do everything. I've had some externally-funded projects, and had some collections catalogued that way, but otherwise I don't have the capacity to catalogue everything myself. There are very few places where you can apply for funding."

[Archive]

"Because we don't have history at our university, we don't have courses that use the archive on a regular basis, so I'm advocating all the time. I have to be really visible and sometimes putting things on a catalogue that people don't think they're going to want isn't visible enough; it has to be displays and going around and talking and banging on doors with armfuls of stuff. So the DIY side of it takes me longer and I'm not as good. If I want to get it professionally done I have to advocate for why it should be done. I have to find a sustainable way of doing it, and to get sustainable funding not just for this year but long-term I have to demonstrate how it's going to bring value for money for the students, because that's the key.

I also work with the community, but it can be really hard to put in for community funding when you're in a university, and for university funding if you're out in the community."

[Archive]

"Not all universities are equal in regard to supporting collections, and we have enormous backlogs. Even though it would appear [we're] better off than some small museums, we still don't get the opportunity necessarily to do the backbone work of what a museum or archive is - which is cataloguing. We also under pressure to speak to very different kinds of audiences and sometimes we can feel like we are forced to speak to researchers and students rather than the generic audience, and sometimes that's quite difficult when it comes to funding."

[Museum]

"The bottom line is that over years of austerity, sadly documentation is one of the last things that people get to do, it's always on their To Do List. They want to do it, they see the benefit of doing it, but if you're the curator plus the learning officer plus the person that's opening the door, and the person is cleaning the toilets and welcoming people to the museum, then actually working on your collection documentation can be the last thing on your list."

[Museum adviser]

As AHRC considers future infrastructure needs, what would be really useful to you?

"So there are all of these great resources out there, the trouble is most of that is collected from the large organisations … and they're all feeding into things like DiSSCo, but smaller organisations are still struggling to get properly acknowledged and mapped. It's something that a lot of smaller collections are really keen to get involved with because it does make the data much more accessible. Even if you see the value, and have a degree of buy-in to the idea of getting those collections online because you know that they will be used by researchers, we have a million competing priorities and it's one of those things that for smaller organisations just isn't on the radar. Any project which provides a mechanism for that digitisation (like the mollusc project) is a really valuable thing because it means that people are coming to you, to your collection, to do the research and digitisation, and collect the data. So it takes the onus away from the institution and puts it onto the project."

[Subject Specialist Network]

"I deal with a lot with copyright and it is unbelievably complicated, but there is guidance out there. It's just having the time and the energy to [do what the] guidance tells you is almost impossible … If there was a project which, for example, used AI to look at all copyright owners in all the archives across the country that would be a brilliant start, because just finding the [correct rights holder] can take me days sometimes, bringing up local records and finding out which newspaper bought this

newspaper, which bought that newspaper. If you could create a definitive list of copyright owners in the UK that would be fantastic. I guess there must be a fair amount of duplication, that other archives may have done that research and spent the days and days figuring out what the copyright position is regarding a particular individual, and another archive also has material relating to that same individual and they're also having to do their several days' worth of work as well. Often that background information never makes it onto a catalogue and so isn't easily searchable."

[Archive]

"We don't have the luxury to just [share data] for the greater good. I've got museums on their knees, looking at hundreds of thousands, or millions of pounds of deficit. But we don't have the luxury of going 'let's do this because it's the right thing to do'. We all know it's the right thing to do, but there's no business case. If it's possible to get funding to help museums that would be amazing, that to be what this is about."

[Museum]

"We've embarked on a process of rationalisation. It will remove our duplicates but we've actually identified holes in our collection, and finding other museums that have got collections that could fill those holes that is proving really tricky, as is actually getting an answer from any museums, when you identify somebody who might have something that you would like to acquire. Coming at it from a consumer point of view, we would pretty much like to know what's in other museums."

[Museum]

"Potentially something that might be very useful is using AI to go through records to help define controlled vocabularies and thesauruses."

[Museum]

"I think getting more voices involved and getting different viewpoints on our collection. Most local authority museums are of strange mix of, you know, the odd wealthy industrialist giving a load of stuff, and things that were collected in the 60s and 70s, and a great mix in between, not necessarily with great selection. So I think looking more broadly and making connections between other collections in other places would cheer up the gene pool."

[Museum]

"Is the model that Art UK used something that would be useful? They put out a call for all public institutions with artwork to be part of the public catalogue programme, and they paid for the digitisation, and paid for copyright clearance. They paid for photographers, and they paid for people to do the documentation, the text, the metadata creation. And then it was all put online in one central catalogue. Easy for us. The Art UK model works so well because it identified all the essential parts ... The strength of that model was that they gave the infrastructure. They stated what was needed, and then they provided it."

[Museum]

"Going back to 2004-05, [we] were looking at distributed national collections, identifying where gaps were, particularly with larger objects that we're not all able to collect. And there was a lot of work done in that emerging digital age, and that was drawn together for CultureGrid. But it's fair to say that since those mid-2000s what we've been doing more as an SSN is to focus our work on supporting curators who may not have the subject specialisms we had in the 1980s and '90s, to really engage their audiences with those collections rather than asking where example of those things are. Looking

forwards, the main focus of sharing information is probably more ad hoc; we use JISCMail as a way of our members being able to ask questions, and quite often it's where offers are made when collections are at risk, exploring where those rarer examples and ensuring we're collecting what we ought to be collecting as a distributed collection, so we're not all collecting the same example and missing out on those items that help complete the broader story. Anything that helps us as individual museums to be able to make those right decisions is positive from a research point of view."

[Subject Specialist Network]

Is there anything else you would like to say?

"[TaNC] is such a big project, but actually it is quite difficult sometimes to understand where certain museums might fit within a broader picture. I don't understand what [TaNC] is there for, what is it trying to achieve. I come from a marketing background so will always go for audience first. Even having been involved in most of the meetings, so far, and the public meetings and everything else about TaNC, and talking to people who are sitting on the steering committee, I still can't get what it is."

[Museum adviser]

"About 30 years ago there were about a dozen collections that said "we are the national collections in photography" and set up the forerunner to our SSN. And it hit the rocks, it foundered because they could never get beyond those dozen organisations. And when we started of course we realised straight away that there are thousands of photographic collections in the UK with some sort of public remit. And I think many are nationally significant collections, so [TaNC] is really interesting to us. I think our best role model for how to proceed is actually the national plant collections, Plant Heritage, because of the way they run. Basically the plants can stay where they are: if you have the national collection of daffodils, you submit yourself to the database, that's acknowledged, you go through certain levels of quality and standards and so on, and it's a kind of virtual grid that's dropped over all of the plants in the UK to allow you to get yourself on to that national collection. The really fantastic advantage to that is that anyone can be involved and submit themselves and have a role. And that is so different to the 12 giant organisations who said "we are it."

[Subject Specialist Network]

So we're really at that point where, like, it's amazing to think about building to being part of a national collection, but at the moment that needs to be help at the very start and over the long term. A lot of the time it's 'yeah, we need to help these small locations', and there's help for six months or a year. And then that goes away. And then there's a period of stagnation as the people who were involved in it disappear and the institutional knowledge moves on. And then you come back to it in another few years and, 'hey, let's do it again' - and you're already back to square one again."

[Museum]

"However, it's got to be done in absolute collaboration and partnership, not some paternalistic, we-know-better relationship, especially if they're serious about local museum collections being seen as part of the sum of the parts of the national collection. Often these kinds of hugely funded projects hope to support the smaller, local sector. And very often, there's quite a lot of money, understandably, put into staffing for that particular moment in time. But it's on admin, training and putting some toolkits together, and all that stuff. The local sector is still struggling to find the capacity to use those toolkits or infrastructure. So if this is going to succeed at all, it's going to have to be driven from the bottom, not just doing some roundtables and some people on a steering group,

but really making sure that there is capacity, on a grassroots level, on a more-people-with-more-time level within the local sector.

One hesitancy is about projects like this is that the people who can best provide what is needed at the ground level are the people who are at ground level. And one of the concerns about a massive project, when you get to millions and millions of pounds, is what this project is going to look like. It ends up being top heavy and there's a lot of money wasted at meetings, when actually what we need is feet on the ground, whether more staff to do the collections or [other things to free up time]. I'm always hesitant to say that 'staff needs to be thrown at this' but I do genuinely think in this case that staff needs to be thrown at this."

[Museum adviser]

"I think academic researchers default to thinking that the significant collections are within national museums or large regional museums when actually there are significant collections in the tiniest of museums, tucked away. Some campaign or something needs to be done to chip away at that attitude. Anyway, I think the whole TaNC programme is almost perpetuating that, and there's potential for a missed opportunity to get the local significance of collections out there.

And I think part of the problem is where we are with this one-stop shop: who, and how, and how do we make sure that every single museum in the country puts that one-liner about what their collection is about? And then how do we make sure that academics or researchers know that that's the one-stop shop? If it's anything more than a one-liner then you've got the issues about different collections management systems talking to each other.

I also think there's something around 'significant' collections. It's just a term we all use. There's training needed about how you measure significance in a local collection: what's local significance, and when does that tip into regionally significant or nationally significant? A database of holding 'significant' collections is great, but it's not necessarily going to capture everything that's a significant collection. Maybe those looking after it don't realise its significance.

Something that is significant for specific research is not necessarily a nationally significant thing. It might will be very significant in pocket watches, for example, but we don't know that because we haven't got the information and we need researchers to tell us, which seems to be the cyclical what-came-first-chicken-or-egg, researchers-or-collections."

[Museum adviser]