TOWARDS A NATIONAL COLLECTION

UKRI Arts and Humanities Research Council

COMMISSIONED REPORT

User Research: UK Gallery, Library, Archive and Museum (GLAM) Digital Collections Infrastructure

Claire Bailey-Ross,
Emily Burgess,
Panagiotis Papageorgiou,
University of Portsmouth

July 2024

Table of Contents

Front cover image: Birmingham Museums Trust (1937). Le lecteur. [oil on canvas] Accessible: https://artuk.org/discover/artworks/le-lecteur-33993

Executive Summary

This report was commissioned by the Towards a National Collection (TaNC) Programme Directorate, to conduct a user consultation to identify researchers' needs and requirements, helping to define what should be included in a future UK digital collections infrastructure. This research was part of the TaNC programme, which aims to open up and democratise access to cultural heritage collections across the UK.

The main aim of the consultation was to gain a comprehensive understanding of the needs and requirements of different research users across academia and Independent Research Organisations (IROs), and what they would like to see included in a future UK digital collections infrastructure.

The consultation comprised three stages:

- Initial liaison with TaNC-funded projects and consultants gathered strategic insights, framing the user consultation approach.
- A mixed-method consultation, including focus groups, interviews, and a survey, to understand digital infrastructure needs across various research fields and career stages.
- Critical analysis of the collected data revealed user requirements, motivations, career levels, and behaviours. Insights were shared and refined with key stakeholders.

This approach was designed to gain a nuanced understanding of user perspectives, informing the development of a tailored and effective UK digital collections infrastructure.

Key Findings

This user research identified several opportunities and challenges for the future development of a digital collections infrastructure from the perspective of researchers. Key findings include:

Digitisation, Digital Preservation, and Collaboration:

- Researchers expressed a desire for more digitised materials but acknowledged that making more materials available online requires dedicated funding and support.
- There is a need to balance between shallow and deep digitisation approaches, which requires careful consideration of research needs and resource limitations.
- Researchers have concerns about the long-term availability of collections data, highlighting the importance of a robust digital preservation framework to safeguard digital materials for future use.
- Increased support, open access initiatives, and resources for staff training are necessary to sustainably advance digital cultural heritage initiatives.

Improved Search and Discovery:

- Researchers value comprehensive search functionalities, advanced filtering and sorting options, and user-friendly interfaces for exploring and interacting with collections. Discoverability and serendipitous discovery are also highlighted as important aspects.
- Metadata accuracy, completeness and potential enhancement were raised consistently. Standardised metadata practices are crucial for accurate description and discoverability.

Connection and Interoperability:

- There is strong support for connections to be built between and across collections and across institutions. But there is a need to address substantial differences in the structure, content, and coding of collections data presented from multiple sources across the cultural heritage sector.
- Researchers felt that standardisation is key to sustainability, and to enabling interoperability. Ultimately researchers want a digital collections infrastructure to have true interoperation.
- Researchers felt it is important to balance technological advancements with the preservation of human expertise and fostering community engagement in and across digital platforms.

Long-Term and Environmental Sustainability:

- Researchers want sustainable practices to be integrated into a future digital collections infrastructure.
- There is a need for a fundamental shift in culture to acknowledge and actively work towards environmentally sustainable practices.

Impact and Accountability:

- There is a perceived lack of impact assessment and accountability in digital cultural heritage projects.
- Concerns were raised about the repetitiveness of discussions and the limited lasting tangible results from previous digital collections and digital infrastructure projects. Knowledge sharing across the sector is desirable.

Collaborative efforts are essential to address challenges and leverage opportunities in developing a digital collections infrastructure. By prioritising sustainability, enhancing search and discovery, establishing standardised frameworks for interoperability, and fostering collaboration, we can create a more inclusive and interconnected digital cultural heritage research environment.

1. Introduction

This research user consultation report details the findings from a user study to identify arts and humanities researchers' needs and requirements, helping to define what should be included in a future UK digital collections infrastructure.

The user research is a component of the large-scale AHRC Towards a National Collection programme, which aims to have a transformative impact on digital search and cataloguing tools for collections, enhancing research capability, public access and public engagement with heritage. There are now vast amounts of digital heritage resources available to support research. These resources are changing the ways in which researchers work, offering convenient quick access to a wide selection of materials, particularly regarding cultural heritage content. These changes have affected how users interact with digital gallery, library, archive, and museum (GLAM) resources and the GLAM information environment.

Towards a National Collection (TaNC) is a five-year investment in the UK's world-renowned museums, archives, libraries, and galleries. Funding is provided through UK Research and Innovation's Strategic Priorities Fund and delivered by the Arts and Humanities Research Council (AHRC). The programme is taking the first steps towards creating a UK digital collections infrastructure by dissolving barriers between different collections – opening UK heritage to the world.

Key to the development of a UK digital collections infrastructure will be a solid understanding of what research users want from a future digital collections infrastructure. There is a need to investigate the ways in which research users value and use existing cultural heritage digital infrastructures and what motivations and priorities they have for a future UK digital infrastructure.

A UK digital collections infrastructure can be effective only if it addresses the abilities and needs of diverse audiences and the attributes of their environment. Engineering an infrastructure as a sustainable and effective ecosystem requires an understanding of the practices and needs of interdisciplinary scholars, technical specialists, research software engineers and other end users of cultural heritage knowledge production, reproduction, and dissemination processes.

This user research investigated the evolving landscape of digital heritage resources, focusing specifically on the requirements of research users for a prospective UK-wide digital collections infrastructure. As digital resources reshape scholarly work, particularly in the cultural heritage domain, the project aimed to explore user needs and requirements, ensuring that a future UK digital collection infrastructure aligns with those needs.

1.1 Structure of the Report

The overall purpose of this user research report is to present evidence from the research user consultation detailing the current practices and future needs and requirements of research users of digital collections for Gallery, Library, Archive and Museums (GLAM) sector. Following this Introduction, Section 2 discusses the approach to the user research; Section 3 describes the findings; finally, section 4 discusses the key findings and recommendations in the area of research users and a future UK digital collections infrastructure.

2. Approach

Predominantly qualitative methods were employed, incorporating three key approaches: interviews, focus groups, and a survey.

This commissioned user study aimed to deliver user-centred insights into what users want and need from a future digital collections infrastructure in the UK. The goal was to consult a diverse range of research users within higher education institutions and Independent Research Organisations (IROs), focusing on gathering priorities and needs related to a future UK digital collections infrastructure. This was based on their previous experience with digital collections, particularly for those involved in interdisciplinary research within fields such as digital humanities, history, arts, and social sciences.

Understanding the research community's insights and experiences is crucial for building a broad and inclusive picture of potential interest and engagement in a UK digital collections infrastructure.

The user consultation comprised three stages:

- The first stage involved liaison with relevant TaNC-funded projects and various appointed consultants working in other TaNC areas (including open standards). This study aimed to determine the strategic context and insights gained from TaNC projects on digital infrastructure needs. This initial contextual consultation with key research stakeholders provided framing to develop an approach for designing a user consultation to understand the needs and desires of different research users.
- The second stage involved shaping the understanding of digital infrastructure needs through a mixed-method approach to collecting user views and perspectives:
    - Focus groups were held across UK Research Institutions.
    - Interviews were conducted with researchers from interdisciplinary research in fields such as digital humanities, history, arts, and social sciences. These interviews covered a spectrum of career stages to understand needs and motivations across PhD, early career, mid-career, and senior academic research agendas.
    - A short survey was administered to contextualise the rich qualitative data provided by the focus groups and interviews.
- The final stage involved critical analysis of the collected data to identify user requirements, motivations, roles, and behaviours. Initial insights at this stage were shared and discussed with key stakeholders to establish meaning and refine the findings.

Finally, recommendations were produced based on the user consultation findings. Any significant gaps in the available data were pinpointed to inform any future TaNC consultancies on user needs.

Inclusiveness was a key aim, actively seeking views from all career stages, including senior decision-makers, early career researchers, curators, librarians, and research software engineers in the GLAM and arts and humanities disciplines.

2.1 Interviews

One -to-one interviews with researchers from across arts and humanities research disciplines, including gallery, library, archive and museum (GLAM) professionals, were conducted between 6th November 2023 and 29th January 2024. A total of 40 interviews were conducted.

Interviews were undertaken following the guidance and good practices laid out in Williamson and Bow (2002), Gillham (2005) and Rubin and Rubin (2005). Interviews allow for complex and complete responses and explanation, and clarification can be provided to the respondents as well as to the interviewer.

The participants invited to engage with this research were interviewed, using a list of prompt questions (Appendix A), to gain focused insights and a first-hand understanding about the participants':

- Research practices
- Interactions with digital infrastructure
- Features and tools of digital cultural resources
- Future directions of a digital collection research environment.

The interviews were carried out using recorded video interviews. The recorded video interviews took place online using an encrypted University of Portsmouth Zoom account. The audio was used for transcription purposes only. Zoom Audio transcription automatically transcribes the audio of a meeting that you record to the cloud. After this transcript is processed, it appears as a separate VTT file. Using the built-in recording and transcription tool minimised technical issues.

2.2 Focus Groups

The UK Digital Collections User Consultation Focus Groups ran between 14th November 2023 and 19th January 2024. Six focus groups were held across UK research institutions with a range of research users in higher education institutions and Independent Research Organisations (IROs). A total of 35 participants took part in the focus groups. The disciplines represented included art, archaeology, archives and records management, digital humanities, film studies, heritage science, history, information studies, media and communications, museum curatorial, and research software engineering.

The focus groups were undertaken following the guidance and good practices laid out in Williamson and Bow (2002). Focus groups are particularly appropriate as it allows the participants to respond and build on the reactions of other members in the group. The explicit use of group interaction to produce data and insights creates a 'synergistic effect' (Stewart et al. 2007).

The participants invited to take part in this research engaged with a group discussion with some ideation activities. A question guide (Appendix B) was used to elicit discussion from the participants to gain focused insights and a first-hand understanding of the participants' experiences of:

- Current practices
- Exploration of Needs
- User Interface and Experience
- Features and tools of digital cultural resources

- Digital Collections Research Infrastructure Ideation

The focus groups were audio recorded and facilitated by the authors.

2.3 Survey

The UK Digital Collections User Consultation Survey ran between 20th November 2023 and 31st January 2024. A total of 206 responses were received, seven of which were ineligible. A total of 199 responses were left for analysis.

The survey asked about views on current research processes and experiences with the use of galleries, libraries archives and museums (GLAM) digital collections. The participants were also asked for their views on the requirements and future direction of a GLAM digital collections infrastructure. It focused on the Arts and Humanities Research Council (AHRC) Research Community, both within academia and Independent Research Organisations, particularly those individuals who are connected to a UK-based institution. Full details of the survey can be found in Appendix C.

The survey was open from 20th November 2023 until 26th January 2024. The analysis of the findings took place from January-February 2024.

2.4 Note on Ethics

Ethical approval was received for this project by the University of Portsmouth's Faculty of Creative and Cultural Industries Ethics Committee. Reference Number CCI-FEthC 2023-026. To preserve participant anonymity information containing personal identifiers has been removed and replaced with numerical participant IDs to ensure it is impossible to identify the original source as much as is feasible. Institutional affiliation has been included in order to demonstrate the coverage in terms of representation of the UK research landscape.

3. Findings

3.1 Introduction

Drawing on insights from researcher interviews, focus groups, and surveys, this section examines how researchers currently leverage digital collections in their work. It provides detail on the typical data formats researchers require from GLAM digital collections, highlights challenges researchers face when accessing or using digital library, archive, museum, or gallery collections in their research and goes on to identify valuable features and functionality, and finally future requirements for a digital collections infrastructure essential for supporting effective research workflows.

3.2 Interviews and Focus Group Findings

3.2.1 Current Research Practices with Digital Collections

The findings from the qualitative interviews and focus groups provide insightful perspectives on participants' integration of digital cultural heritage collections into their research practices, shedding light on their use behaviours with GLAM digital collections and typical research workflows. The findings echo previous user research findings on current research practice with digital collections (Woodley and Towell 2022).

The Arts and Humanities researchers who participated in the interviews and focus group had diverse and varied approaches to their research workflows with digital collections. It is not possible to identify a single workflow of activities that describes all arts and humanities and GLAM digital collections-related research – it is neither attainable nor desirable to prescribe a standard model for humanistic research. But it is possible to provide a high-level overview of some of the common ways of approaching research using GLAM digital collections.

Specific collections vs. open online search

From the interviews several researchers prefer to use specific collections and institutions that are known for their holdings in a particular area of study.

"I started with the big organisations that I already knew." (TaNCINT24)

Certain disciplines gravitate toward specific archives known for their collections in their field, either starting with direct communication to archivists or curators or to the digital collection. For example, some of the historians interviewed refer specifically to the National Archives, literary scholars mentioned Early English Books Online (EBBO), and English Short Title Catalogue (ESTC), while art historians cite the exploration of museum digital collections including the V&A and the National Maritime Museum.

Researchers value archives with curated content and expert selection, as doing so saves time and ensures relevance. Well-established archives often provide contextual information and user guidance, further enhancing the research experience.

Other researchers value Google and online search for its ability to facilitate discovery. For example:

"I'll start with an online search and then just follow what comes through from the initial page or to the online search. I'd love to say that I generally start with library catalogues, but I don't. To be honest, I find I get a broader swathe of more relevant hits if I just do basic online searching." (TaNCINT39)

Keyword search

"I'm not usually browsing for stuff. I'm usually looking for something specific. Partly I would say I guess that I mean that stage of my career where I'm old enough and we've been around long enough that I have a fair idea of what's out there and therefore I can look for something specific." (TaNCINT06)

Researchers value comprehensive search functionalities with keyword support, topic filtering, and various refinement options, the British Newspaper Archive is cited several times as a good example, as is the V&A which provides filtered search by context type (e.g. period, place, themes, materials, makers and artist). Many researchers use keyword search to find specific information in GLAM digital collections. This approach is especially helpful for large datasets or when researchers have a clear information seeking need, and know exactly what they are looking for. While keyword search is common, researchers adopt different strategies based on their needs. Some use simple terms for quick results, while others employ multiterm searches incorporating period-specific language and synonyms for precision. The effectiveness of keyword searches depends on metadata quality, and relevant findings can be missed due to ambiguous terms or lack of comprehensive indexing. Researchers highlight the need for advanced search options such as faceted browsing and proximity searching for refined results.

Physical archives vs. digital archives

Some researchers rely heavily on physical archives, while others prefer digital archives. The choice between the two often depends on the nature of the research and the availability of materials. The choice between physical and digital archives is rarely exclusive. Researchers often use both strategies, leveraging the strengths of each format. Physical archives are still irreplaceable for unique or fragile materials, while digital collections offer broader accessibility, searchability, and ease of comparison. With increasing digitisation initiatives, researchers have adapted their approaches. Those accustomed to physical archives might initially find digital tools unfamiliar, but as accessibility and user interfaces improve, digital collections become more integrated into research workflows.

One of the foremost uses of a digital infrastructure, recognised by interviewees, would be its ability to allow less established academics to research while avoiding the anxieties that come with physically visiting an institution. Incidentally, this coincided with discussions about the benefits digital infrastructures offer to people who are neurodivergent and struggle in social settings (TaNCINT10, TaNCINT01). There was an overwhelming acknowledgement of time constraints and particularly for early career researchers, there is pressure and strain on work-life balance; therefore, the likelihood of being able to afford or having the time to travel to research institutions was minimal. The interviewees recognised that a digital infrastructure could and does alleviate this significantly.

"The time and money resources that you'd need to put into actually going to the Museum or Archive when I can just go to the library here or sit on my desk and access it just as easily and use search functions and stuff that you can't do in real life. I find that it kind of increases your productivity a little bit." (TaNCINT02)

Interview participants specifically within the early career demographic prioritised ease of access of digital collections, which resulted in productivity and time savings. These were aspects of great importance, especially among many interviewees who were balancing paid employment in junior research roles—such as research assistants and research officers—with doctoral degrees. Due to the lack of connections provided by an established career, there was a widespread reliance from this demographic on materials that were publicly available online, with many museums and open access research bodies proving useful.

International digital initiatives

Several interview and focus group participants spoke about their use of international digital initiatives and infrastructure projects that are multi sectoral, multi collection, and that connect cultural objects and records at a national scale.

There was much discussion about the advanced infrastructure of digital resources outside of Britain, for example, in Canada, the U.S. and Sweden (TaNCINT04), Gale Historical Newspapers (TaNCINT23), the Australian National Bibliographic Database (ANDB) (TaNCINT19), Digital New Zealand (TaNCINT27), and WorldCat (TaNCFG6). In these cases, overseas institutions were regarded as more approachable at answering queries and sending digitised copies of documents. British institutions, on the other hand, were cited as being much harder to access and less likely to share information without financial compensation. The need to physically visit an archive was raised by several participants, specifically the process of being approved by an archivist and having to book an appointment. In these cases, it was recognised by interviewees that a clear and concise research database would save time and solve accessibility concerns.

3.2.2 Type of Data Formats for Research

Textual data hold primary importance

Digitised text from documents, newspapers, books, and manuscripts remains the most utilised format across various research areas from the interviews and focus groups. This highlights the crucial role of accurate transcription and text-based search functionalities in digital collections. Several researchers discussed the use of optical character recognition (OCR) within digital collections and highlighted that transcripts and OCR can be challenging and noisy (for example errors, distortion or unwanted variation which obscures the text) due to imperfect output from OCR or because handwriting or the structure of the document produce imperfect results. Visual and multimedia data are gaining ground, images, including high-resolution, are increasingly used for visual confirmation and study. From the interviews and focus groups it is not clear if it is solely flat 2D static images that are considered more appropriate for research. Audio and video sources were used, but not as much as text.

Diverse formats are essential

Researchers draw upon a wide range of data formats depending on their research question and the content of the collections. These include text (newspapers, magazines, correspondence, books), images (photographs, paintings, artefacts), audio (interviews), video (film footage), and numerical data (census data, museum collection records). Features offered by digital collections, such as high-resolution images, zoom functions, and metadata, are highly valued by researchers. Additionally, they appreciate the ability to refine search results and explore broader contexts within digital collections. Within accessible digital collections, there was a focus on the importance of clear and concise imagery, readability, and additional descriptions for managing the data processing.

Some researchers highlighted that they don't really have a preference over type of data available:

"I don't really care about sort of whether it's a tiff or whatever. I just want a decent quality, but I do care about if there are formats that are not exposed to me, which I know are there. And I do care about, the turning of everything into text because the web works through text. That seems like a really impoverished way of thinking about digital collections." (TaNCFG6)

While some researchers expressed no particular preference for data formats, others emphasised the importance of accessibility and the availability of formats beyond text. They underscored the limitations of relying solely on text-based formats for digital collections.

"It depends on what the research questions are because if I'm a researcher for multi-spectral imaging, well then I want the highest quality, the supplier, and everything else about where that image was taken and what it's taken on if it's possible to emulate it in a different environment.

I want all of that information. But then I spoke to an accounting historian and he was like, I just want a picture of an accounting book. I don't need to be high resolution. I just want to visually confirm that this is the thing that we should be looking at, and then I'll go see the real thing if I can. So it again, it just shows that there's lots of different needs." (TaNCFG05)

Complete and detailed metadata

"Metadata is life." (TaNCINT11).

The importance of comprehensive, accurate, and detailed descriptive metadata alongside collection assets was emphasised in interviews and focus groups. Researchers expressed a need for additional contextual information beyond standard descriptive metadata elements. This includes contextual details, provenance information, and descriptive tags that enhance the discovery, understanding, and analysis of research data. Mentioned metadata standards include horizontal standards like Dublin Core and METS, as well as vertical standards tailored to specific data and communities of practice.

Interoperability

Researchers emphasise the need for long-term accessibility and standardised formats such as IIIF for images and MARC 21 for cataloguing to ensure data compatibility and reusability across platforms.

3.2.3 Challenges when using Digital Collections

Challenges stemming from complex interfaces, digitised material quality, inconsistent metadata, and inaccurate descriptions pose significant hindrances to research practices. Additionally, ensuring the quality and long-term accessibility of digitised materials, along with navigating copyright restrictions, remains a persistent challenge. Addressing these issues is crucial for fully realising the potential of digital cultural heritage collections for arts and humanities researchers.

Search functionality and user experience

Researchers struggle with complex interfaces, clunky navigation, and limited searchability across multiple collections.

"You don't want to take a whole training course just to be able to do a search on a website. You really want to be able to just go in, find what you need to find and use it. You don't really want to have to go. Okay. So now, this is another package, another platform that I need to work out how they're making it accessible to me." (TaNCINT01)

Several researchers felt frustrated that they were unable to find what they were looking for or lacked the necessary expertise on a specific platform to utilise digital collections effectively. They highlight the importance of being able to easily navigate and search for information without the need for extensive training or adaptation to different platforms. Many researchers placed emphasis on the importance of search functionality.

"*I mean so much of the experience comes down to the search function and how reliable that seems to be and I'm not a techy person so I don't know what things make a difference there but I just know that there's a lot of variation in the user experience with search functions, some digital archives seem to come up, you know, really easily. You can find the things you want and then others. It's like so hit and miss. And as we said previously, you try and narrow down your search terms and it actually doesn't return anything sometimes. So. The reliability of the search function, I guess, is really, really important for me." (TaNCINT23)*

"knowing what is in collections and whether or not they [have been] digitised in the first place and then it leads on to well what information are there about those things in those collections…But you need to be able to know in the first place where the thing that you're potentially looking for actually is." (TaNCFG01)

"I remember this catalogue and the result there just seemed to be no rhyme or reason with how they'd can categorise the digital collections so. I was searching for print collections. And they weren't coming up where I expected them to come up. And then I find that they've somehow been. Plugged into the fashion collections. So things like that." (TaNCINT05)

While usability and accessibility issues concern the ease of use for end-users, technical limitations and design issues pertain to the inherent characteristics of the collections themselves. One interviewee discussed the technical limitations and design issues inherent in many GLAM digital collections that were not originally intended for machine access or analytical use.

"Well, the major thing is most of them are bespoke, so they're very difficult to work with, a lot of the historical collections were never designed for machine access. They're designed purely to be looked at. That really limits your ability to use analytical tools or anything with those" (TaNCINT06)

The contrast between the excerpts suggests that addressing these challenges requires a comprehensive approach that considers both user experience design and technical infrastructure development. Additionally, it underscores the importance of adapting GLAM collections for contemporary digital contexts while preserving their integrity and authenticity.

Data quality and structure

One participant (TaNCINT19) uses the "holes in the cheese" analogy to describe how a UK digital cultural heritage collection would look like at the moment since there are many incomplete datasets (e.g. not everything is digitised and available online or significant metadata information is missing).

"The pitfalls of looking for digital content, especially digital content which isn't well managed. There isn't well ordered structure and I mean for example, right now if we're looking for a collection for records or data in our collection, we're just doing a controlled find rather than anything that is more sophisticated than that. I think in due course we will move towards a much more of a unique ID Object Management based system for finding individual records… We don't have those [Persistent Identifiers] right now. I think we increasingly realise we need them because our collections are getting bigger and I think what's also driving that change for us as well is that we're having to make different copies of the same thing." (TaNCINT14)

Archive descriptions are sometimes dated or inaccurate, so one interviewee had wasted time travelling to view a document, only for them to be irrelevant.

"Unfortunately, there have been occasions why I've gone and done that [to visit the physical archive] and found very little because the description of the document I'm looking at isn't quite actually what I wanted to find or doesn't have enough depth to it that explains it so I have had to go be disappointed on occasion." (TaNCINT04)

"We've recently been going through a huge digitisation process funded by the [Welsh] government where we've digitised over 25,000 artworks in the museum's collection. So that's been a huge discovery process where you recognise that there are gaps in the database and a way we can start to fill those gaps and have that associated material. And I think one of the things that makes it really tricky is human error. So, you know, over decades of people inputting data into databases, you know, there are instances of error where somebody misread a title and the title has been typed incorrectly or they've been misspellings and that kind of thing. So it's kind of that's what we're trying to look at." (TaNCFG01)

Researchers spoke of incomplete or inaccurate metadata, with poor tagging and inadequate descriptions that impede the discoverability and understanding of collection materials.

"it's impossible sometimes to find things that you know that are there." (TaNCINT27)

Researchers also raised challenges to their research practice due to errors and inconsistencies in digitised quality of collections: whether that be scan quality of an image, or the time required to wait for a higher quality image, handwriting readability, and transcription errors. These factors introduce noise, particularly stemming from errors in the transcription process. These errors can range from simple inaccuracies in punctuation to more serious issues like mistyped or misinterpreted words, which can alter the entire meaning of a sentence. Such errors increase inaccuracies and pose a greater risk of misinterpreting the data during the analysis stage, thereby hindering research efforts.

“Sometimes if the image that's online is only a very small thumbnail and you make a request for a, um, a higher quality image. The turnaround time for that can be very slow and. That, yeah, in those sort of cases and there's maybe a charge for that when if you're only doing a search, if you only want it to kind of verify whether or not it's useful for your research, that can be a barrier.” (*TaNCINT05)*

“Quality in terms of being able to read them because the handwriting is not the problem, it's the contrast, the light, the way it's been taken isn't the best… I mean, I'm very thankful for it because otherwise I'd be having to go to London to delve into the the archives…but the quality of some of the documents are not the best, especially you know, some of them are kind of and some of the handwriting is not the best. So it's a bit of a challenge sometimes.” (TaNCINT08)

Several responses mentioned challenges related to ensuring the quality and long-term accessibility of digitised materials. This highlights the importance of robust digitisation processes and sustainable preservation strategies for digital cultural heritage collections.

Digitisation

Many of the interview and focus group participants mentioned the ongoing work to digitise cultural heritage collections. This was largely framed as an endeavour which could be completed, but was subject to funding and resources. Frustrations were expressed by members of the sector, and a feeling that small organisations especially are facing particular barriers.

“There's a huge amount of manuscript material that hasn't got anywhere near being digitised. It’s tucked away in all sorts of collections and sometimes it's not even catalogued.” (TaNCINT17)

“I would say one of the biggest challenges right across collections in the UK and internationally is the fact that it's very often we do not have adequate basic level two-dimensional imagery which is for research. So before we even think about 3D, there is a huge challenge with basic.” (TaNCFG04)

A challenge which featured across the interviews and focus groups highlighted what within collections has been digitised and what has not. There is a perceived absence of coherent documentation on how many digital collections exist, as well what level of digitisation has occurred. It was highlighted that documenting and analysing the availability of digitised collections would help researchers identify gaps enabling new avenues for research.

“stuff is what people want, but it's the right stuff. If It's not the right stuff. It might as well not be digitised.” (TaNCFG04)

“there might be a risk that the non-digitised stuff is missed out of the picture perhaps.” (TaNCINT17)

“There's a lot to digitise, but like the biases of the archive that's kind of come into that…but we don't want all the same kind of narratives being told again and again through it [a digital collection] and ignoring the kind of aspects of colonialism agenda.” (TaNCINT01)

Concerns were raised regarding digitised information, which was also cited as regressive and sometimes uncomfortable. For example, an instance was given where the researcher was focused on a specific woman; however, her research was recorded in the files of her husband, brother or male publisher (TaNCINT08). The interviewee made it clear that this resulted in the woman being harder to research, prioritising the scientific subject matter based on gender. There was similar discussion about the lack of warning or context on many British archives regarding offensive language and topics, with links to empire and stolen material often not

being mentioned, constituting 'knowledge appropriation'. Overall, there was a consensus that these topics should not be sanitised but that there should be developed digital infrastructure in place to allow researchers to gain awareness of the source's origin and context, which can also aid the research process.

Several researchers raised concerns about bias in digitisation efforts, potentially skewing the historical record toward certain narratives or privileged collections. This includes decisions about what is digitised and who can access it, raising ethical questions and concerns about transparency within collections. There is a concern that the process may inadvertently reinforce certain narratives while overlooking others, particularly those related to colonialism or marginalised perspectives. This raises questions about the ethical implications of digitisation practices and the need for a more inclusive approach to representing diverse historical narratives.

A few of the researchers interviewed also worked at smaller local cultural heritage organisations, who highlighted challenges with backlog of uncatalogued items and a lack of dedicated staff and funding for digitisation, having to rely on volunteers with limited training and expertise.

"We do have a massive backlog, with our collections." (TaNCINT12)

"We do really want to get much more of our [collection] digitised it's a funding issue more than anything" (TaNCINT03)

"I've got a team of about 8 [volunteers] who do varying levels of work when it comes to documentation. If I was to put it on a scale of 0 to 10 before I train them. It's 0 to 1…some of whom also have a go at a bit of cataloguing but general knowledge of digital preservation is very, very low." (TaNCINT16)

Resourcing is a primary concern, with the lack of digitisation causing these institutions to fall behind and attain limited functionality in the aftermath of the COVID-19 pandemic. Therefore, there was an underdevelopment of physical record archiving, combined with a lack of resources and training. In the case of researchers in smaller heritage institutions, digital infrastructure seemed to be a utopian ideal that was too far away to be obtained even though it would be greatly welcomed. Digitisation provides these institutions with a way to protect and manage collections while also allowing significant outreach to other researchers.

Challenges and barriers were also raised regarding use of restricted access digital collections associated with specific objects or assets. This can be problematic because it is not always convenient for a researcher to pay an in situ visit to the memory institution hosting the requested object(s).

"Obviously the problem with digital archives is often the paywall. Which is a faff, but in terms of once you actually have access to the sources." (TaNCINT18)

"There's limits with these private companies that have digitised a lot of the materials and you've got to pay a subscription…I think the financial implications in lots of ways play into what people have access to and how much they can find here." (TaNCINT23)

Several responses mentioned challenges related to ensuring the quality and long-term accessibility of digitised materials. This highlights the importance of robust digitisation processes and sustainable preservation strategies for digital cultural heritage collections.

Summary

Researchers face a range of challenges when working with current digital collections. These challenges include various aspects, from search functionality, data quality, digitisation completeness, to access restrictions. Researchers encounter obstacles such as complex interfaces, limited searchability, and inconsistencies across digital platforms, impacting their ability to efficiently navigate collections for their research. Additionally, incomplete metadata, inaccurate descriptions, and poor tagging hinder the discoverability and understanding of materials, highlighting the critical need for improved data management. While digitisation efforts aim to broaden access, researchers have identified issues related to selective digitisation, biases, and resourcing constraints, particularly in smaller GLAM institutions. Access restrictions, such as paywalls and limitations imposed by private entities, further impede researchers' ability to access and utilise digital collections fully. Addressing these challenges is essential for enhancing accessibility, usability, and long-term preservation, ultimately enabling more inclusive and comprehensive research practices with digital collections.

3.2.4 Future Requirements

The landscape of digital collections is marked by significant challenges and opportunities. As articulated by various stakeholders, there exists a pressing need for comprehensive digital infrastructure that facilitates preservation, accessibility, and advanced research capabilities. This section outlines key future requirements identified through interviews and focus groups discussions with researchers across academia and independent research organisations across the cultural heritage sector.

Preservation infrastructure

Several researchers spoke of the importance of involving a multidisciplinary team of researchers in the development of the digital collection infrastructure, as well as a digital preservation infrastructure. Challenges related to ensuring the quality and long-term accessibility of digitised collections materials. In an effort to recognise the future research potential, a digital preservation infrastructure should be central to a UK digital collections infrastructure.

One interview participant in a senior position at an IRO stated that;

“No one's got the infrastructure there that they need yet…And we're still so far off, and we really shouldn't be. But there's been no investment in this, right. You know, preservation. Solid, good preservation infrastructure that you're going to be able to recover from and then to be able to build Infrastructure worthy of proper digital humanities research. That's more than just text data mining. It needs real investment, and it needs big investment too.” (TaNCINT19)

Preservation and knowledge-based resources, like the Archaeology Data Service[1], the UK’s openly accessible repository of heritage data, was cited as a good example for preservation practices.

[1] Archaeology Data Service

Challenge: The current state of digital cultural heritage collections is akin to "holes in the cheese," (TaNCINT19) signifying incomplete datasets and inadequate preservation infrastructure.

Requirement: Substantial investment is needed to establish robust preservation infrastructure capable of safeguarding digital materials and enabling future research activities. This infrastructure should support effective recovery and long-term accessibility.

Digitisation

On the whole researchers wanted access to more digitised materials but acknowledge that making more materials available online requires dedicated funding and support.

"The priority would be getting all of our collection, all our records of our collections onto a digital system. So at the moment we are in between that. So we've got Stay Books, which is a very old style of recording. We've got index cards again, very old style. Some of that is on this online collection on MODES, some isn't. But getting all of that on to MODES and having that hard copy, having that digital copy. Would be totally ideal here and then it opens up so many, so many things after that." (TaNCINT12)

"I think underpinning this is the fact that there's almost no money available for digitisation anywhere and that's kind of the elephant in the room." (TaNCFG02)

"Ideally what you want is the ability to either have everything digitised already or do On-Demand digitisation, where the researcher goes to colleagues here [GLAM organisation] and says, Look, can you digitise this stuff for me and they can because they've got the money and the kit. Now that's kind of a bit of a dream space." (TaNCFG04)

Researchers emphasise the significance of digested collections for research purposes. There were several discussions around what would be preferable; all possible collections digitised to a baseline standard (shallow digitisation), or key things, or indeed, pre-existing digitised collections, enhanced to the best standard (narrow and deep digitisation). The answers were very mixed.

"There is a horse and cart issue. That research will not occur if people do not know that there is material to research. And so until you have accessible digitised collections, you cannot actually understand what level of research demand is." (TaNCFG04)

Challenge: Limited resources, both financial and temporal, pose significant obstacles to the comprehensive digitisation and online accessibility of cultural heritage materials.

Requirement: There is a need to accelerate digitisation efforts, including the adoption of on-demand digitisation approaches to meet specific research needs efficiently.

Improved search and discovery

There was a consistent emphasis on user friendliness, improved search functions, advanced filtering and sorting options, and user-friendly interfaces for exploring and interacting with collections. Discoverability and serendipitous discovery are also highlighted as important aspects. There was a call for improved metadata standards and practices, as well as standardised data entry with human oversight for quality control. This approach is crucial for ensuring accurate description and discoverability of digital materials.

"Accessibility and ease of being able to use it and having being able to go in in one place. and that they're knowing that things that they're in that one place are quite interconnected, so that once you start putting you might not always get. I don't expect to get the key word straight off. But once you've tried a few key words, then you do start to get access to things that are useful to you. It is that interconnectedness, and we can only hope, I guess, that AI is going to solve that." (TaNCINT01)

The potential of Artificial intelligence (AI) was discussed multiple times, as a way to enhance collections data and search, whether that be through using the AI to create automatic keywords, particularly for image search, or to enhance descriptive metadata, to offering entirely new ways of accessing, understanding and researching collections at scale. Concerns were also raised around the ethics of AI and the potential for algorithmic bias. Some researchers were concerned about how the use of AI with collections could be made more transparent and clear for researchers.

"I'm looking at algorithmic bias that shapes what we see online, and especially how historians and our historians specifically are influenced by that in ways they might not understand, and how we can make that more transparent and clear to the user." (TaNCINT15)

Translation was also raised with the belief that consistent data and translation options would enhance the searchability and access. Both the National Museum of Wales[2] and the Digital Repository of Ireland[3] were cited as good examples in this regard.

"I think there's some practices that are happening in nation states within the UK that could be implemented on a larger scale and arguably should be in the larger scale, and then that could be scaled up to international." (TaNCFG02)

One focus group participant from a Welsh institution did raise a concern about reliance on AI for implementing multilingualism.

"I think there's an element of… about machine learning, that kind of thing, isn't quite as advanced for the Welsh language as it is for English and so, you know, you've got that potential shift in balance if you're not careful when you're machine learning can advance things really quickly producing English language content, but Welsh language content, they can't." (TaNCFG01)

While concerns about algorithmic bias and the ability of AI exist, researchers on the whole see AI as a valuable tool for enhancing search, analysis, and interpretation.

Challenge: Existing search functionalities often lack user-friendliness and fail to facilitate accurate and comprehensive discovery across diverse collections.

Requirement: Enhanced search capabilities, including networked digitised collections and AI-powered tools, to enable efficient access to cultural heritage materials. User-friendly interfaces are critical for improving discoverability.

[2] Amgueddfa Cymru
[3] Digital Repository of Ireland

Metadata completeness and accuracy

"The ideal environment. I think, uh [on], closer reflection of the archives that we have. That rich kind of rich metadata and keywords to kind of link cross collections." (TaNCINT05)

Metadata accuracy, completeness and potential enhancement was raised consistently across the researcher interviews and focus group discussions. Points were raised around descriptive and contextual or paradata and the importance of providing relevant information about the decision-making process with regards to digitisation will also enable researchers to make informed decisions when using digital collections content and gain necessary details for the purposes of their work. However, concerns were also raised about how much metadata are appropriate for each item in a collection, as the desire to add information for each research discipline may not be feasible as standard, and when does metadata enhancement become modification (TaNCFG01).

"the main point would be to link all these different collections, which are not all of them available digitally online. Some of them are just that you have to mention, say, like if you want to see that, you've got to go there in the actual. From this perspective the digital collections research infrastructure is about not just access to digital content, but access to metadata about content as well." (TaNCINT39)

The British Library service, EThOS[4], a database of UK theses was cited as a good example for thinking about metadata and digital research infrastructure;

"you might think about ethos from that perspective as well…All of the metadata about theses that are published in UK universities and on many occasions also copies of those pieces either born digital or digitised. Now those artefacts are available in most cases from the holding institutions, from their own institutional repositories, but the ethos is access and almost an aggregator if you like." (TaNCINT39)

Challenge: Inconsistent metadata standards and practices hinder effective description and discoverability of digital materials for research purposes.

Requirement: Standardised metadata practices and user-friendly interfaces are critical for improving discoverability. Enriching metadata completeness and accuracy for scholars can access relevant contextual information and make informed decisions about using digital collections in their research.

Interoperability and standards

Breaking down disciplinary silos and connecting resources across institutions are top priorities for researchers. The need for tools that support interdisciplinary research and collaboration is highlighted multiple times. This includes cross-disciplinary tagging, interoperability, and mechanisms for interdisciplinary connections and knowledge sharing.

"Maybe if there was a way of doing a digitised network where you could connect up with the archives of the newspaper, archives, different type source types and be able to do it in a way where people could maybe go through one avenue and then access everything and be able to say access to National Archives, X is the British Museum, those types, and be able to find sources more accurately. Maybe that would be a good idea in the long run." (TaNCInt23)

[4] Ethos

There was much discussion about connecting collections, interoperability and open standards. Researchers felt that standardisation is key to sustainability, and to enabling interoperability. Standards adoption goes beyond issues of metadata standards or standard file formats, but covers all aspects of a future digital collections infrastructure, including the organisational structure, technical architecture, and the persistent identifiers.

"So 90% of what I do is bringing together data from sources that haven't been brought together before in formats that have no relation to each other. But they're completely incompatible. So the idea of bringing everything together in one place in one standard format will basically lower the barrier to entry, because a lot of the very difficult part of the data analysis is analysing things and shifting them in such a way and shifting the data structures in such a way that they then make sense together." (TaNCInt15)

A range of frameworks and standards were discussed by researchers, including;

- International Image Interoperability Framework (IIIF)
- Oxford Common File Layout (OCFL) an application-independent approach to the storage of digital information in a structured, transparent, and predictable manner. It is designed to promote long-term object management best practices within digital repositories.
- Open Archival Information System (OAIS) standard
- Records in Contexts–Conceptual Model (RiC-CM).
- General International Standard Archival Description ISAD(G) although one interviewee believed ISAD(G) is not scalable enough, even though it is extensively used by some National Records Departments.

Some of the researchers go on to recommend that standardised application programming interfaces (APIs), as well as the use of Persistent Identifiers (PIDs) and a shared contextual metadata framework should all be basic components of a UK digital infrastructure.

Challenge: Integrating diverse collections from various sources with incompatible formats into a unified, interoperable digital network.

Requirement: Implementing standardised frameworks and a shared contextual metadata framework, to facilitate seamless access and interoperability across different digital collections.

Human-centred approach

Integrating human expertise and fostering community engagement remain vital alongside technological advancements. The early career researcher demographic, in particular, placed importance on human connections and collection management and an emphasis on a balance between the physical and the digital to create a rich and diverse research environment. Emphasis was placed on the importance of tools that support contextual understanding and highlight the need for human expertise and guidance within digital platforms. This includes mechanisms for connecting users with individuals who have intimate knowledge of collections.

"But every object, story, archival components, I have worked within any collection has always been mediated by a human who's daily work is in that collection. So in the context of a national digital infrastructure, I would still want to see. Well, who are those facilitators, those people and it might not be that it's a you know, one to one in person thing. a few universities do a version of that now, you go to the library website, and there's a

chat box where you can still chat to a librarian and even if you're looking to find an online resource. So even in a digital context for me, that would be the thing that I'd want." (TaNCINT09).

"I think that the key thing that often gets missed when you talking about physical infrastructure is not only people can use it, but actually the people who work on the thing who create, maintain and look after the things." (TaNCFG05)

Challenge: Balancing technological advancements with the preservation of human expertise and fostering community engagement in digital platforms.

Requirement: The need for tools and mechanisms within digital platforms that facilitate human connections and enable users to access expertise from individuals who have intimate knowledge of collections.

Sustainability

Environmental sustainability was a frequent topic mentioned in the interviews and focus groups. There was discussion about decarbonisation. Mainly, the subject centred on the benefits of digital infrastructure in preventing unnecessary travel and carbon emissions while also providing access to people with disabilities. Algorithmic bias and metrics were considered alongside functionality, transparency and usability (TaNCINT15).

"I think sustainability of digital infrastructures is growing in people's awareness and I think awareness is the first thing, right?... I think bringing greater awareness to that is a positive thing and will feed very much into institutional sustainability strategies and it's difficult when building a new digital thing to say that you're improving that sustainability picture of a data centre." (TaNCINT22)

"No one builds for climate. No one's building things that are slow deliberately. No one is challenging users to say, isn't this being always on a bad thing?

No one is looking at really good web design practices…but no one's looking at good web design practices that genuinely enable low resource access, which is more equitable and low energy use upon accessing." (TaNCFG06)

One interviewee suggested that it would be more environmentally friendly to offer on demand access to the high-quality version of a digital object, than having everything available online. Moreover, a consortium of HEIs responsible for the financial sustainability of the Towards a National Collection (TaNC) project can secure a "greener" administrative environment, since they already have the required research capacity (existing working groups).

Challenge: Finding ways to effectively reduce energy consumption and carbon emissions within digital infrastructures while maintaining or improving functionality, usability, and accessibility.

Requirement: Implementing sustainable practices within digital infrastructures, including considerations for environmental impact, such as reducing energy consumption and carbon emissions.

Addressing funding, resource and training Issues

Several participants mentioned the need for increased funding for digitisation projects, open access to materials, and resources for staff training. Funding and resources are seen as critical for expanding digitisation efforts and improving digital infrastructure.

"We spend an awful lot of money physically accommodating and looking after a large quantity of books. We probably spend a tiny fraction of that amount looking after our digital estate, which is probably a similar number of objects. But those objects need cataloguing and curating, because people are still convinced that the magic IT fairies do everything for free. And you got to realise that isn't happening. That's not real. But you know that is a mindset." (TaNCInt06)

"I think culture is maybe important, but also maybe it's more of an effect of funding, because also if you work in the different national libraries and I find the ones that are better funded they are often more open. So I also wonder if this is like, can you change your culture without changing the support for institutions?" (TaNCFG05)

"Look at some of the stuff the Mellon Foundation has funded in the US around bringing community archives into the fold and there's some other interesting initiatives like, even big university libraries sharing digital expertise so that they create a pool of resources which they share among the institutions, because that's a bit more sustainable than relying on every institution maintaining their own resource." (TaNCINT06)

Challenge: Insufficient funding and resources pose significant challenges to expanding digitisation efforts and improving digital infrastructure.

Requirement: Increased funding support and resources for staff training are necessary to sustainably advance digital cultural heritage initiatives.

Impact assessment and accountability

Several of the senior career researchers in both universities and Independent Research Organisations spoke of the perceived lack of impact and efficacy of digital cultural heritage projects. Questions were raised not only of the impact of Towards a National Collection projects, but also how other digital infrastructure projects in the UK and internationally were learning from each other. Examples included the 3D Data Service[5], E-RIHS the European research infrastructure for heritage science[6], The European Open Science Cloud[7], RICHeS, the Research Infrastructure for Conservation and Heritage Science[8] and the Museum Data Service[9]. There is a sense of frustration among some researchers regarding the outcomes of these initiatives, with concerns raised about the repetitiveness of discussions and the limited lasting tangible results.

"I think, you know, it's coming back to open communication really, and sharing best practice across different disciplines, not just within some varieties to understand what's already working. Yeah, we don't do enough of that. We don't have those conversations" (TaNCFG01)

[5] 3D Data Service supported by the UK's Arts and Humanities Research Council under the Scoping future data services for the arts and humanities scheme
[6] The European research infrastructure for heritage science
[7] European Open Science Cloud
[8] RICHeS, the Research Infrastructure for Conservation and Heritage Science
[9] Museum Data Service

“For example, a talk at the Internet Archive yesterday, and now you can upload an image and they generate all the metadata, IIIF, it does OCR on it anyone can upload to this, you know, platform, it's, you know, global technology is behind those. Why doesn't someone go speak to them and they've got all the code available on GitHub. Why isn't someone going to that repository and getting that code asking questions, raising issues?” (TaNCFG05)

“I wonder like do they actually have an impact or is it I guess this is a bit controversial but I see so well, I feel like you know the output these things they just become a Zenodo report and that's it. A lot of people in this room we got to contribute to one of these projects …I don't really know what impact it has and you know we stick it somewhere and that's it it's like déja vu. Every time I come to these events, it doesn't matter like what the time period is. So scroll back and it's the same conversation and it's all good. It's published and then on reads it, and then we have the same conversation. So I don't know practically how this sector moves on from that.” (TaNCFG05)

Challenge: Perceived lack of impact and efficacy of digital cultural heritage projects.

Requirement: Establish open communication channels for sharing best practices across different disciplines and projects, facilitating collaboration and learning from successful initiatives internationally.

Requirement: There is a need for more tangible and measurable outcomes from digital cultural heritage projects, along with mechanisms for tracking and assessing their impact effectively.

Conclusion

The future of digital collections infrastructure requires concerted efforts to address existing challenges and leverage emerging opportunities. Overall, the responses reflect a strong requirement for a future digital collections infrastructure to have tools and features that facilitate interconnectedness, interdisciplinary research, AI integration, improved accessibility, and user-friendly interfaces to support arts and humanities research.

By investing in preservation, digitisation, search capabilities, metadata standards, collaboration, accessibility, and funding, it will be possible to collectively shape a more inclusive, interconnected, and sustainable landscape for cultural heritage research.

3.2.5 Response from Society of Software Engineering

Following on from one of the focus groups which specifically targeted Research Software Engineers, the Society of Research Software Engineering[10], a registered charity with a large membership base, provided a formal response to the research user consultation. Their mission is to promote the recognition of software's critical role in research, including within the arts, humanities, and cultural heritage sectors.

Key findings and recommendations include prioritising standardised data formats like IIIF for image data and metadata for interoperability and API consistency. It also stresses the importance of developing computational pipelines for tasks like photogrammetry and text analysis, advocating for open-source tools. The Society advocates for programmatic access to large datasets through consistent APIs. They also propose supporting non-consumptive research methods, such as those used by the HathiTrust Digital Library[11]. Additionally, they highlight the necessity for accessible computational environments and RSE involvement in bridging the gap between infrastructure and research needs. The importance of providing accessible high-performance computing resources and enhancing digital skills in the arts, humanities, and cultural heritage sectors is also highlighted. It advocates for investments in RSE capabilities to democratise access to skills and foster professional collaboration. The Society's response further emphasises the potential of AI methods for information discovery while addressing biases and promoting representativeness in datasets. Finally, they advocate for prioritising analytical functionality, reuse and sustainability, engagement, and training in digital collections infrastructure. Overall, the Society of Research Software Engineering calls for enhanced collaboration between RSEs and the academic community to address challenges and improve the use of digital cultural heritage collections for research purposes.

[10] The Society of Research Software Engineering
[11] HathiTrust Digital Library

3.3 Survey Findings

3.3.1 Respondent's Profile – Research Discipline

Given that this user study focuses on arts and humanities disciplines, the survey maps respondents' research discipline using an adapted version of level one codes from the Primary Research Areas covered by AHRC discipline funding remit[12]. Respondents were asked to use the 'Other' option if their discipline was not listed. Given the nature of the topic, Digital Humanities and Heritage Science were also added as research discipline categories (Table 1).

Table 1: Survey Respondent's Research Discipline

Research Field/Area of Study	Responses
Archaeology	22
Classics	4
Creative Industries	4
Cultural and Museum Studies (Cultural)	16
Design	0
Digital Humanities (DH)	21
Divinity and Religion (Religion)	2
Drama and Theatre Studies (Drama)	1
Heritage Science	3
History	23
Information and Communication Technologies (ICT)	10
Languages and Literature (Languages)	12
Law and Legal Studies (Law)	0
Library and Information Studies (LIS)	12
Linguistics	0
Media	0
Music and Visual Arts (Music)	2
Philosophy	0
Political Science and International Studies (PolSci)	0
Theology	0
N/A — I do not do research	6
Other	8
Answered	146
Skipped	58

[12] AHRC Research Funding Guide - Section 7 Additional Information AHRC Disciplines

Participants were also asked which institution/organisation they work for, this was an optional question which received 130 responses. See Appendix D for details. The distribution of responses across institutions indicates that there is a diverse representation of participants from various universities and organisations. With 67 UK institutions represented, though representative, this does not show the whole picture. A complete mapping of the UK research landscape is beyond the scope of this survey, but the institutional profile of the participants and their responses discussed below reveal useful insights into the landscape.

3.3.2 Respondent's Profile - Career Stage

The career phases are based on the League of European Universities (LERU) research career stage definitions for the UK[13]. Modifications were made to include equivalent Research Software Engineering and GLAM research-related posts. The data indicate a fairly balanced distribution of respondents across different career stages, providing a varied set of perspectives and experiences (Table 2).

The Mid/Recognised career stage category has the highest percentage of responses, making up approximately 37% of the total. The early Career and Senior/Established/Experienced Career stage categories were also significantly represented, with approximately 23% and 31%, respectively. The distribution across different career stages provides a varied set of perspectives and experiences among the respondents. Thematic understanding will be categorised by career stage to understand any differences in needs and motivations across PhD, early career, mid-career and senior research agendas.

[13] Possible Research Career Paths in the United Kingdom

3.3.3 Type of Research Institution

Participants were asked which type of research institution they worked in, from the 144 who responded. The most frequent response was University (57.93%), followed by Archive (8.97%) and Library (7.59%) (Table 2).

Table 2: Type of Research Institution Represented in the Survey

Type of Research Institution	Response Percent	Responses
University	57.93%	84
Research Centre	2.76%	4
Library	7.59%	11
Archive	8.97%	13
Museum and Gallery	6.90%	10
Historic Environment/Heritage	3.45%	5
More than one of the type of organisation above	8.28%	12
None of the above	1.38%	2
Other	2.76%	4
	Answered	145
	Skipped	54
	Total	199

3.3.4 Use of GLAM Digital Collections

Participants were asked ‘Do you access or use any digital collections material held by galleries, libraries, archives, museums or other heritage-sector organisations in the UK? This was a mandatory question with two fixed choice answers ('Yes' or 'No'). 95% of respondents stated that they did access and use digital collections. Those who do not use GLAM digital collections stated numerous reasons for not using them, including a lack of awareness of relevant GLAM digital collections, concerns about copyright and licensing restrictions and a lack of relevance to their research field.

3.3.5 Types of Data Formats for Research

Survey participants were asked 'What GLAM data formats do you find the most useful in your research?' in order to understand the data types researchers are working with to better understand their research practices and digital collections needs. There were seven 'Other' responses, three which could be incorporated into pre-existing categories, the table below (Table 3) incorporates the 'Other' responses accordingly.

Table 3: Data Formats used by Researchers

GLAM Data Formats used by Researchers	Response Percent	Responses
Text (e.g. books, newspapers, encyclopaedias, archival documents and other historical sources)	87.7%	107
Numerical	19.7%	24
Digital Surrogate 2D material (e.g. Photographic images or videos)	76.2%	93
Digital Surrogate 3D material (e.g. 360 photos or videos, 3D digitisation mesh)	20.5%	25
Born Digital 2D material (e.g. computer aided design (CAD) data, tags, associations, texts)	18.9%	23
Born Digital 3D material (e.g. 3D modelling/sculpting and animations)	12.3%	15
4D material (e.g. immersive environments for VR/AR/MR, motion capture, sensor data for digital twins, motion capture datasets)	4.9%	6
Catalogues, databases	85.2%	104
Maps and Geospatial data (e.g. maps, 3D point clouds, digital terrain models)	48.4%	59
Audio data (e.g. music, voice recordings, oral histories)	32.8%	40
Other (respondents stated: concept of combined formats - combined together to form FAIR DOs; Research software; websites and visual timelines)	3.3%	4
	Answered	122
	Skipped	84

The data on GLAM data formats Figure 1: Types of Data Formats for ResearchFigure 1) reveals a range of insights into the preferences and practices of researchers. As you might expect for arts and humanities researchers, the majority of respondents, 87.7%, find text data formats such as books, newspapers, and archival documents to be the most useful in their research. This finding echoes the findings from the

interviews and focus groups. Closely followed by Catalogues and Databases with 85.2% of respondents expressing their usefulness. Catalogues and Databases were discussed within the interviews and focus groups, but not to this extent. A total of 76.2% of respondents reported using visual resources in the form of digital surrogate 2D materials, like photographic images or videos, as part of their research workflows. Additionally, 48.4% found maps and geospatial data valuable, indicating the importance of spatial information in GLAM research. Audio data, including music, voice recordings, and oral histories, were relevant to 32.8% of the respondents. Numerical data are used by 19.7% of respondents, demonstrating a quantitative dimension in GLAM research.

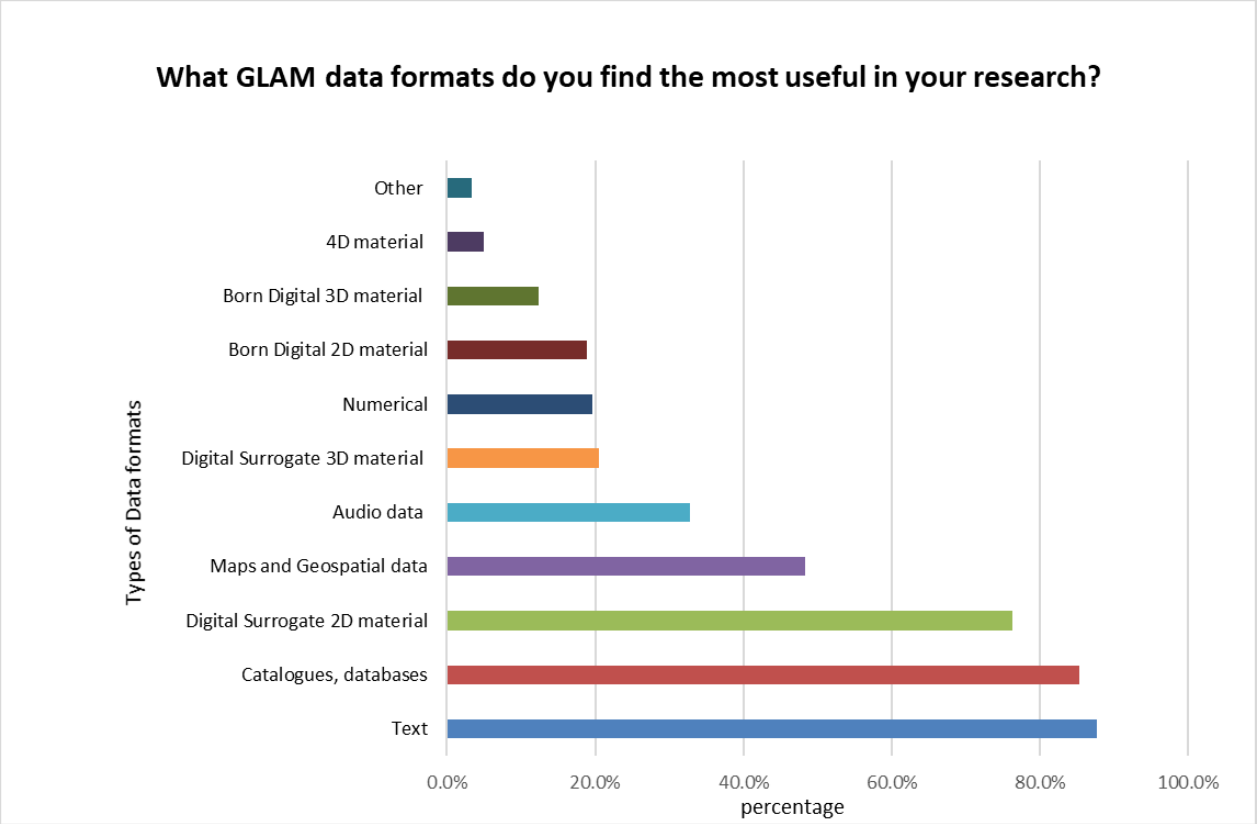

Figure 1: Types of Data Formats for Research

Born digital materials, both 2D (18.9%) and 3D (12.3%), such as computer-aided design (CAD) data and 3D modelling, are recognised in the research practices. Emerging trends include the interest in 4D material (4.9%), indicating a growing usage of immersive environments for virtual reality (VR), augmented reality (AR), mixed reality (MR), and motion capture and sensor data for digital twins. The 'Other' responses highlight (3.3%) a variety of primary data types that respondents are using, the concept of combined formats could indicate a need to accommodate evolving data formats and research practices into a future digital collections infrastructure. Overall, the data provide a comprehensive view of the diverse GLAM data formats utilised by arts and humanities researchers, emphasising the significance of textual, visual, spatial, and audio data in their work. This supports and adds further detail to the qualitative responses in section 3.2.2.

3.3.6 Challenges when Using Digital Collections

Participants were asked if they faced any particular challenges when accessing or using digital library, archive, museum, or gallery collections in their research; the responses indicate the multifaceted nature of challenges researchers face in accessing and using digital GLAM collections (Table 4).

Table 4: Challenges researchers face when accessing or using digital library, archive, museum, or gallery collections

Challenges	Response Percent	Responses
Limited search capabilities	67.54%	77
Limited interoperability	42.98%	49
Complex user interface	26.32%	30
Poor data quality and accuracy	41.23%	47
Incomplete metadata	56.14%	64
Lack of collaboration tools	22.81%	26
Inadequate customisation options	12.28%	14
Insufficient preservation of digital materials (e.g. inappropriate file formats)	21.05%	24
Absence of data analysis tools	17.54%	20
Lack of integration with research tools	22.81%	26
Inadequate user support and assistance	15.79%	18
Other	26.32%	30
	Answered	114
	Skipped	90

The highest reported challenge was limited search capabilities, with 67.83% of respondents facing difficulties. This indicates a need for improvements in search functionality within digital GLAM collections. A total of 55.65% of respondents reported that incomplete metadata are a notable challenge, underscoring the importance of comprehensive metadata for effective use of digital collections in arts and humanities research. This is alongside the reported challenges of poor data quality and accuracy, which 40.87% identify as a challenge. This emphasises the need for measures to ensure the reliability of digital materials within GLAM collections; echoing the findings from the interviews and focus groups.

A significant portion of respondents, 42.61%, highlighted limited interoperability as a challenge. This suggests that seamless integration with other collections and data sources is a crucial consideration for researchers. A considerable percentage (26.96%) cited a complex user interface as a challenge. One respondent expanded by stating navigation difficulties and the absence of comprehensive guides or instructional materials when navigating digital collections. This suggests that user-friendly interfaces are crucial for enhancing the accessibility and usability of digital GLAM collections. Inadequate user support and

assistance (15.65%) and insufficient customisation options (12.17%) suggest that providing user support and guidance and allowing users to tailor their experience are essential considerations for improving research user satisfaction. Concerns related to collaboration tools (22.61%) and lack of integration with research tools (22.61%) highlight the importance of collaborative and integrative features in digital GLAM platforms to support researchers in their work. Preservation challenges, including inappropriate file formats (20.87%), indicate a need for robust preservation strategies to ensure the longevity of digital materials within GLAM collections.

The cumulative effect of these challenges highlights the potential impact on research efficiency, effectiveness, and the overall user experience when working with digital GLAM collections. Addressing these challenges, such as enhancing search capabilities, improving interoperability, and ensuring comprehensive metadata, is crucial for optimising the utility of digital collections in research settings.

The 'Other' responses (26.09%) provide insights into the diverse challenges faced by researchers, encompassing issues related to catalogue and metadata information, digitisation processes, copyright, access, and financial considerations. Addressing these challenges is crucial for enhancing the accessibility and usability of digital collections in research.

Some of the 'Other' responses noted discrepancies between catalogue information and the actual availability of items, highlighting potential challenges in collection catalogue management. Metadata and cataloguing concerns were raised, including challenges with incomplete metadata, lack of cataloguing, and the expressed need for structured data downloads for interrogation and reuse, emphasising the crucial role of metadata standards in facilitating research. Challenges related to digitisation processes were also noted, including issues with incomplete digitisation, poor Optical Character Recognition (OCR) leading to bad or missing data, and concerns about low-resolution content. The availability and quality of images stand out as significant challenges, emphasising the importance of visual resources and the need for high-quality digital representations in research. Copyright and licensing complexities were raised, with concerns about unclear or incomplete copyright information, issues with licensing of digitised materials, and uncertainty regarding Intellectual Property (IP) status or copyright restrictions on reuse. Some participants noted access-related challenges including limited access to collections not subscribed to by universities, geographic inconsistencies in access to text sources, and uneven access to materials. These issues underscore the importance of broad and equitable access to digital collections.

Financial barriers, paywalls, and reliance on closed commercial providers are identified as significant challenges, underscoring the impact of financial considerations on access to digital resources and potential limitations associated with reliance on commercial platforms.

The survey responses provide complementary insights to the interviews and focus groups (section 3.2.3) into the challenges faced by researchers in accessing and using digital collections.

3.3.7 Features and Functionality

Survey respondents were asked about specific features or functionalities that they believe are currently missing from GLAM digital collections that could improve their research experience (Table 5). Advanced search capabilities including cross-collection searching is a feature which is highly desirable, with 72% of respondents indicating its importance. This suggests that users value the ability to conduct detailed and comprehensive searches across multiple collections simultaneously.

Table 5: Features or functionalities Researcher's feel are currently missing from GLAM digital collections

Features or functionalities	Response Percent	Responses
Advanced search capabilities including cross collection searching	**72%**	**83**
More user-friendly interface	**46%**	**53**
Improved data quality and accuracy	**54%**	**63**
Open data sets and collections	**61%**	**71**
Enhanced metadata completeness	**49%**	**57**
Better collaboration tools	**19%**	**22**
Increased customisation options	**17%**	**20**
Stronger preservation of digital materials	**27%**	**31**
Additional data analysis tools	**16%**	**19**
Improved integration with research tools	**24%**	**28**
Enhanced user support and assistance	**17%**	**20**
Other	**9%**	**10**
	Answered	**116**
	Skipped	**90**

A significant majority of respondents (61%) prioritised open datasets and collections, reflecting a desire for greater transparency and standardisation. This sentiment is closely aligned with the emphasis placed by nearly half of the respondents (49%) on the importance of comprehensive metadata, which plays a crucial role in facilitating the discovery and understanding of collection materials for research purposes. Additionally, over half of the respondents (54%) highlight the need for better data quality and accuracy, highlighting the essential role of reliable information within digital collections for research.

More user-friendly interfaces

Nearly half of the respondents (46%) expressed a want for more intuitive and easier-to-use interfaces. This indicates a need for improvements in usability and design to enhance the overall user experience. A relatively smaller percentage of respondents (17%) desired increased customisation options. This suggests that while customisation is valued, it may not be as critical as other features. Similar to customisation options, only 17% of respondents prioritised enhanced user support and assistance. This suggests that while support is valued, it may not be as critical as other features and functionality.

Stronger digital preservation of digital materials

Nearly a quarter of respondents (27%) expressed a desire for stronger digital preservation. This highlights the importance of ensuring the long-term accessibility and integrity of digital collections. Both the quantitative and qualitative findings highlight the need for stronger digital preservation practices to ensure the long-term accessibility, integrity, and research potential of digital collections.

Integration with research tools

Approximately a quarter of respondents (24%) indicated a desire for better integration with research tools. This emphasises the importance of seamless integration with existing research workflows. While only 19% of respondents prioritised collaboration tools, it still indicates a need for improved collaboration capabilities within digital platforms, suggesting that collaboration among users is valued but not as highly as other features. Only 16% of respondents prioritised additional data analysis tools. This suggests that while some users may find such tools useful, they are not universally considered essential.

Overall, the data suggest that respondents value features that enhance search capabilities, usability, data quality, openness, and preservation of digital materials. This supports the findings from the interviews and focus groups about future requirements for a digital collections infrastructure (section 3.2.4).

Survey respondents were also asked to rate the importance of a range of features in a digital collections infrastructure for their research (Table 7).

Table 6: Importance ratings for features of a digital collections infrastructure

Please rate the importance of the following features in a digital collections infrastructure for your research (1 = not important, 5 = extremely important):							
	Very Important	Important	Neutral	Less Important	Not Important	Total	Weighted Average
Search capabilities	93	17	0	2	0	112.0	1.21
User-friendly interface	62	40	7	4	0	113.0	1.58
Data quality and accuracy	90	22	3	0	0	115.0	1.24
Metadata completeness	53	45	11	2	1	112.0	1.69
Collaboration tools	9	30	47	20	4	110.0	2.82
Customisation options	6	35	42	19	5	107.0	2.83
Preservation of digital materials	58	39	9	3	1	110.0	1.64
Data analysis tools	10	36	37	20	4	107.0	2.74
Integration with pre-existing research tools (e.g., reference management software)	12	36	39	17	5	109.0	2.7
User support and assistance	25	53	26	7	1	112.0	2.16
Other						9.0	
						Answered	116
						Skipped	88

From this data, it is evident that the majority of respondents consider search capabilities, data quality and accuracy, user-friendly interface, and metadata completeness as very important aspects for a digital collections infrastructure for research purposes. On the other hand, aspects such as collaboration tools, customisation options, data analysis tools, and integration with pre-existing research tools are considered less important by a significant portion of respondents. However, there are variations in the degree of importance assigned to each aspect, with search capabilities and data quality and accuracy being rated slightly higher than user-friendly interface and metadata completeness.

3.3.8 Collaboration

Participants were asked about the importance of collaboration in research projects. The data indicate a strong consensus among respondents regarding the importance of collaboration to research projects, with the majority (70%) expressing either a very high or significant level of importance attributed to collaborative efforts. This aligns with the broader trend in academia towards interdisciplinary collaboration and team-based research approaches. A smaller percentage of respondents rated collaboration as 'Neutral' (15.57%), 'Less Important' (4.10%), or 'Not Important' (0.82%). While these percentages are lower, they still represent

a portion of researchers who may have varying perspectives on the importance of collaboration in their specific projects.

Survey participants were also asked what features or tools they like to see for maximising collaboration in a digital collections infrastructure. Despite 79% indicating that collaboration is important to research projects, the responses to collaborative tools in a digital collection infrastructure were limited. Forty-three open text responses were received. The open text responses highlight a range of needs and considerations for maximising collaboration in a digital collections infrastructure, including a request for Jupyter-type [14] tools for sharing analytical routines and tools for analysing and testing interactions with digital content. This reflects a need for robust analytical capabilities within a future digital collections infrastructure.

There is interest in tools for crowdsourcing and collaborative events, suggesting a want for platforms that could facilitate collaborative efforts among researchers and contributors. Other responses emphasise the importance of adherence to existing standards such as IIIF and Linked Art for representing and disseminating cultural heritage data. They seek seamless integration with established classification schemes and standards as well as open approaches to licensing and access across collections. Some advocated for tools that promote open access and facilitate collaboration across different entities.

Some respondents expressed uncertainty about the concept of collaboration as an infrastructure issue. A concern was also raised about security, particularly in light of recent cybersecurity issues in the cultural heritage sector. They questioned the need for a centralised portal for co-working and emphasise that collaborating researchers already have their own infrastructure or workflow for this task.

3.3.9 Copyright and Licensing

Survey respondents indicated that they did encounter issues related to copyright and licensing when using digital materials for research purposes. The survey data indicates that a significant proportion of respondents have encountered copyright or licensing issues in their research using GLAM digital collections. Specifically, nearly 60% of respondents reported that copyright restrictions limit the usage of materials, while approximately 44% noted that licensing issues hinder their research. This suggests that navigating copyright and licensing considerations is a common challenge faced by researchers. Additionally, a smaller percentage of respondents (20%) stated that they have not encountered such issues, while a minority (5%) mentioned that they exclusively use open access materials.

The data highlights the importance of addressing copyright and licensing complexities in a digital collections infrastructure, as these issues can impact access to and use of materials for research purposes. This suggests a need for strategies to mitigate these challenges and promote wider access to research materials while adhering to legal and ethical considerations.

[14] Jupyter notebooks

Table 7: Opinions on Inclusion of Open Access Content in Digital Collections Infrastructure

Should a digital collections infrastructure include only Open Access and equivalently licensed content?	Response Percent	Responses
Yes, only Open Access and equivalently licensed content should be included	**15.52%**	**18**
No, it should include restricted access materials and Open Access and equivalently licensed content	**62.93%**	**74**
Don't Know		**16**
Other		**9**
	Answered	**117**

Survey participants were also asked if a digital collections infrastructure should include only Open Access and equivalently licensed content (Table 7). The responses reflect a range of viewpoints regarding the inclusion of content within a digital collections infrastructure, underscoring the complexity of balancing accessibility, rights management, and other factors in its development. Approximately 15.52% (18 respondents) advocated for a model where only Open Access and equivalently licensed content are included. This perspective prioritises maximising accessibility and promoting the open dissemination of cultural heritage materials without restrictions.

Conversely, a majority of respondents, comprising 62.93% (74 individuals), expressed the opinion that the infrastructure should encompass both restricted access materials and Open Access and equivalently licensed content. This viewpoint reflects a recognition of the value of diverse content types and suggests a willingness to balance accessibility concerns with the preservation of intellectual property rights or other considerations associated with restricted access materials.

A notable portion of respondents, accounting for 13.79% (16 individuals), indicated uncertainty by selecting "Don't Know," signifying a lack of clarity or conviction regarding the optimal approach for structuring the digital collections infrastructure.

3.3.10 Future Research Tools

In an effort to gauge future research needs, survey participants were asked about their interest in exploring potential computational research tools. The aim was to identify tools that could be beneficial for future research practice (Table 8). From the survey data, it is evident that text/data mining, data visualisation, database design, machine learning and artificial intelligence, as well as working with geospatial data are the most popular computational research tools or techniques among the respondents.

Table 8: Researcher interest in exploring computational tools

Computational Research Tool	Percentage	Number of Responses
Data Cleaning	37.86%	39
Text/Data Mining	58.25%	60
Data Visualisation (including 3D)	45.63%	47
Database Design	44.66%	46
Immersive Technologies - Virtual Reality (VR) / Augmented Reality (AR) / Mixed Reality (MR)	21.36%	22
Statistics	25.24%	26
Machine Learning / Artificial Intelligence / Generative Adversarial Networks	40.78%	42
Working with Geospatial data	40.78%	42
Natural Language Processing (NLP) / Textual Analysis / Sentiment Analysis	33.01%	34
Computer Vision	15.53%	16
Topic Modelling	12.62%	13
Semantic web	22.33%	23
Other	4.85%	5
	Total Respondents	103

However, tools such as topic modelling, computer vision, and immersive technologies (VR/XR) have relatively lower percentages of respondents indicating some interest or usage, suggesting either less familiarity with

these tools or a lower perceived relevance to the respondents' research areas. In terms of potential areas for further exploration or development, it could be beneficial to delve into the tools with lower percentages of interest, such as topic modelling or computer vision, to understand the specific needs or barriers that respondents perceive in these areas and to explore ways to address them. Additionally, considering the relatively low interest in immersive technologies, there may be opportunities to investigate how these technologies could be better integrated into research practices in fields such as arts and humanities.

3.3. 11 Open Data

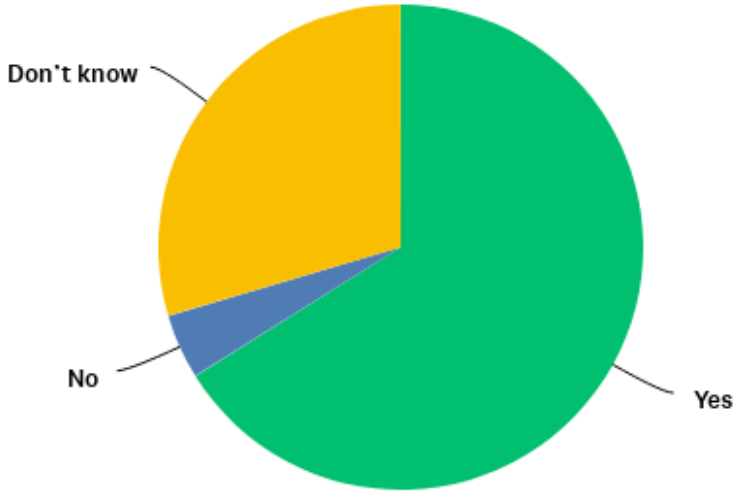

Figure 2: Chart showing respondents' views on whether the adoption of an open data approach, allowing third parties to build new products and services on top of a UK digital collections infrastructure, would be beneficial to their research

Survey participants were asked if the adoption of an open data approach, allowing third parties to build new products and services on top of a UK digital collections infrastructure, would be beneficial to their research. The data (Figure 2) suggests that a significant portion of respondents view the adoption of an open data approach, enabling third parties to develop new products and services on top of a UK digital collections infrastructure, as beneficial to their research. A total of 66.09% (76 respondents) indicated that they believe adopting an open data approach would be beneficial. This majority opinion suggests that many researchers perceive the potential value in opening up access to the UK digital collections infrastructure, allowing for innovation and the creation of new tools and resources by third-party developers. Five respondents expressed the opposite viewpoint, stating that they do not believe adopting an open data approach would be beneficial. While this percentage is relatively low, it still represents a minority perspective that may reflect concerns or reservations about the potential implications of opening up access to the infrastructure. 29.57% (34 respondents) were unsure. This uncertainty could be due to insufficient information about the

implications of an open data approach or an indication that the researchers who participated in this survey may not have the technical skills or confidence to work with open data.

3.3.12 Support to use a future Digital Collections Infrastructure

The data provided present insights from a survey aiming to discern the types of support deemed beneficial for individuals utilising a forthcoming digital collections infrastructure for their research (Table 9). Among the respondents, a significant majority, approximately 74.56%, expressed a strong desire for clear and easily comprehensible information presented in plain language. This indicates a notable need for transparent communication elucidating the advantages of the digital collections infrastructure and offering guidance on how to become a user. Following closely, about 61.40% of participants highlighted the importance of receiving tailored training sessions focusing on the specific functionalities and navigation of the digital collections infrastructure. This underscores the significance of hands-on guidance in enabling users to effectively leverage the resources and features of the infrastructure.

Table 9: Support types researchers would find useful for a future digital collections infrastructure

Types of Support	Percentage	Number of Responses
Clear information in plain language that explains the benefits of the digital collections infrastructure and how to become a user	**74.56%**	**85**
Training on how to use the digital collections infrastructure	**61.40%**	**70**
Active user community	**40.35%**	**46**
Free storage	**36.84%**	**42**
Training on digital skills	**35.96%**	**41**
IT support	**29.82%**	**34**
Appropriate IT equipment	**20.18%**	**23**
Other	**7.89%**	**9**
	Total Responses	**114**

Furthermore, approximately 40.35% of respondents emphasised the value of fostering an active user community, suggesting the importance of collaborative platforms and peer support networks to enhance engagement and user satisfaction. A substantial portion, around 36.84%, identified free storage as a valuable form of support, indicating the significance of addressing data management concerns and ensuring accessibility for users. Additionally, approximately 36% of participants indicated a need for training on digital skills, underscoring the recognition of the importance of enhancing one's proficiency in digital competencies to fully utilize the infrastructure.

While fewer respondents, approximately 29.82%, expressed a need for IT support, the data still highlights the relevance of having technical assistance available to address issues and provide troubleshooting solutions. Additionally, a smaller percentage, about 20.18%, identified the provision of appropriate IT equipment as a useful form of support, suggesting that while important, the focus may be less critical compared to other support services. Lastly, a minority of participants, around 7.89%, specified other forms of support they deemed useful, indicating the importance of considering diverse user needs and preferences in designing support services for a digital collections infrastructure.

Overall, the data highlights the necessity of offering comprehensive support services tailored to users' needs, preferences, and skill levels to facilitate effective adoption and utilisation of the forthcoming digital collections infrastructure in research contexts.

3.3.10 Future Requirements

The survey did not explicitly ask about future requirements but there were questions about which aspects of a digital collections infrastructure did they consider to be important. Participants could only select three, this was intended to encourage decisions about top priorities for their research needs. The data (Table 10) highlights the diverse priorities and considerations that researchers have regarding a digital collections infrastructure, emphasising the importance of addressing accessibility, discovery, preservation, sustainability, analytical functionality, and data reuse to meet the needs of diverse user communities effectively.

Table 10: Important aspects of a future digital collections infrastructure

Which of the following aspects are most important to you in a digital collections infrastructure? (select 3)	Response Percent	Responses
Discovery: Enable effective search and filtering features within the infrastructure, with specific criteria to be defined during the scoping process.	**69.57%**	**80**
Accessibility: Ensure that the digital collections infrastructure is designed inclusively to accommodate diverse audiences with varying abilities and expertise.	**66.96%**	**77**
Preservation: The digital collections infrastructure should ensure the long-term availability of collections data.	**52.17%**	**60**
Sustainability: The digital collections infrastructure should inspire confidence in the infrastructure's long-term sustainability.	**40.87%**	**47**
Reuse: The digital collections infrastructure should maintain accessibility to data for ongoing research and reproducibility.	**40.0%**	**46**
Analytical Functionality: The digital collections infrastructure should provide analytical tools to interact with the content, enhancing its utility for research.	**30.43%**	**35**
Compliance: The digital collections infrastructure should promote and facilitate best practices in managing digital collections.	**25.22%**	**29**
Equitability: Ensure that the infrastructure addresses accessibility and decolonization agendas, promoting fairness.	**25.22%**	**29**
Learning: The digital collections infrastructure should provide opportunities to enhance learning experiences through access to content.	**20.87%**	**24**
Public Engagement: Extend the reach of the infrastructure beyond primary audiences.	**20.87%**	**24**

Content Creation: The digital collections infrastructure should enable contributors to generate derivative products, subject to permissions.	**14.78%**	**17**
Engagement and Training: The digital collections infrastructure actively engage with diverse user communities, offering guidance, training, and best practices.	**14.78%**	**17**
Impact Metrics: The digital collections infrastructure should provide reporting tools for assessing the impact of collections.	**6.96%**	**8**
	Answered	**115**
	Skipped	**89**

Accessibility and discovery

The majority of respondents (66.96% and 69.57%, respectively) emphasised the importance of accessibility and effective search and filtering features within the infrastructure. This suggests a strong requirement for developing a digital collections infrastructure which is inclusive and easily discoverable to accommodate diverse audiences and enhance usability for researchers.

Challenge: To ensure that a digital collections infrastructure is accessible and that it provides effective search and discovery features.

Requirement: To develop a digital collections infrastructure that is inclusive and easily discoverable to accommodate diverse audiences and enhance usability for researchers.

Digital preservation

Over half of the respondents (52.17%) prioritise digital preservation, indicating a significant concern for ensuring the long-term availability of collections data. This highlights the importance of a robust digital preservation framework to safeguard digital materials for future use.

Challenge: To ensure the long-term availability of collections data

Requirement: To establish a robust digital preservation framework to safeguard digital collection materials for future research use.

Long term sustainability

A considerable portion of respondents (40.87%) felt sustainability was important, indicating a desire for confidence in the long-term viability and maintenance of a digital collections infrastructure. This finding suggests a recognition of the importance of ensuring the ongoing availability and functionality of digital collections for research.

Challenge: To ensure the long-term viability and maintenance of a digital collections infrastructure.

Requirement: To have access to a digital collections infrastructure that inspires confidence in its long-term sustainability, ensuring ongoing availability and functionality of digital collections for research purposes

Reuse

While not as high as accessibility, discovery or preservation, a significant portion of respondents (40.0%,) value the ability to reuse data for ongoing research and reproducibility. This suggests a recognition of the

importance of open licenses allowing researchers to imbed data reuse in research workflows. This is supported by the positive responses about the adoption of an open data approach, allowing third parties to build new products and services on top of a UK digital collections infrastructure, with 66.09% (Figure 2) respondents stating that it would be beneficial to their research. This suggests that the reframing of collections as data is beginning to gain traction. An open data or reuse framework could create an ideal moment for a reflection on how future researchers can interact with collections data. The challenge as highlighted in Section 3.3.9 is that a significant proportion of respondents have encountered copyright or licensing issues in their research using GLAM digital collections, which is create a barrier to maximising the value of digital collections in research.

Challenge: Researchers' uncertainty about copyright and licensing creates obstacles for reusing collections data for research purposes. This uncertainty hinders the integration of data reuse into research workflows.

Requirement: To adopt an open data approach to facilitate data reuse and allow researchers to embed data reuse in their research workflows, enabling them to interact with digital collections data effectively.

Conclusion

In summary, addressing these future requirements will be important for the development of a digital collections infrastructure that meets the diverse needs of researchers.

4. Key Findings and Recommendations

From this user consultation, there are numerous challenges and opportunities for a digital collections interface for research purposes. Moving forward, effectively managing expectations and prioritising requirements will be key. The main findings highlight the pressing need for advancements in digitisation practices, preservation, collaboration, search and discovery, interoperability, and possibly more importantly interoperation, as well as long-term environmental sustainability.

4.1 Digitisation, Digital Preservation and Collaboration

Researchers want more digitised materials, but limited resources hinder comprehensive digitisation efforts. Supporting an expanding digitisation programme, including exploring on-demand approaches, is essential to meet specific research needs efficiently. There were mixed feelings about either narrow and deep digitisation versus shallow digitisation.

"my sort of naive instinct is to do the [shallow digitisation] because once you've digitised that, you can do more sophisticated things later. I suppose there's an argument for that's not the best use of resources at digitising things that are buried in a salt mine somewhere. But, but then I think a huge part of the value of this is discovering what's in those things in the salt mine, which we can't do by traditional means because nobody's going to order them up and read them. And so I might be naive, but my feeling is probably shallow digitisation is better. I guess it's also actually less fragile because it's simple." (TaNCFG05)

It is also important to note, that physical collections play a significant role in academic research. The choice between physical and digital archives is rarely exclusive.

"Actually most of our audiences and our research is not interested in the digital, they're interested in the physical. And so this has to be secondary to what we are actually doing." (TaNCFG04)

Therefore, it is important to acknowledge that a digital collections infrastructure is a starting point for research. The concept of digitising collections solely based on a general perception of research value or interest poses challenges, especially considering the time, resources and environmental impacts involved. Exploring on demand digitisation, where specific items are digitised upon user request or when they are at risk from a preservation perspective is important, while also working to increase collection discoverability.

Researchers also believe that there is a pressing need for robust digital preservation infrastructure to ensure the long-term availability of digital materials. Achieving this may require procedural changes, redefining how digital preservation tasks and activities are carried out, as well as organisational shifts, including new ways of providing existing offerings. Part of this must also address insufficient funding, resources and training. Questions were raised by researchers about how to support small, local and volunteer run cultural heritage organisations that face significant challenges to expand digitisation efforts and improving collections access. Increased support, open access initiatives, and resources for staff training are necessary for sustainable advancement. One of the building blocks of any digital research infrastructure system should include people: the users, and the experts who develop and maintain these resources. There is a need to focus on people, skills and collaboration, not solely about the technological infrastructure.

"I think I would love to see a move towards something a bit more national, it may be virtual in shape, but moving in that direction so people can find that collaboration and let's call it collaboration, not support." (TaNCFG05)

4.2 Improved Search and Discovery

Researchers value comprehensive search functionalities with keyword support, topic filtering, and various refinement options. Typically, researchers follow information seeking behaviours, and ultimately researchers want to find what they are looking for. Enhanced search capabilities, including AI tools and user-friendly interfaces, are critical for efficient research access to cultural heritage materials. Standardised metadata practices are necessary for accurate description and discoverability. This standardised information framework could point to persistent identifiers (PIDs) rather than manually cataloguing all this information. However, the technology behind PIDs is currently far too visible to researchers.

"at the moment, technology is far too visible. I think in the digital humanities. You know, people spend an awful lot of time wrangling with XML and you know TEI, it absorbs an enormous amount of time and effort, which you really shouldn't. You'd be better off actually considering interesting problems rather than solving XML coding issues, which is not humanities, exactly." (TaNCINT06)

"Most folk just want to get access to the materials. Basic digitisation on a searchable sustainable platform is all that is required." (additional comment from survey respondent)

4.3 Connection and Interoperability

Researchers would like connections to be built between and across collections and across institutions. Breaking down disciplinary silos and connecting resources and cross collection searching are research priorities. Implementing standardised frameworks and metadata practices is crucial for interoperability across digital collections. It is important to note that interoperability is a statement of potential and researchers want a digital collections infrastructure to have true interoperation. In order to achieve this there is a need to address substantial differences in the structure, content, and coding of collections data presented from multiple sources across the cultural heritage sector.

Integrating human expertise and fostering community engagement alongside technological advancements is important, whether this be the curators, librarians, and archivists who manage, preserve and catalogue collections or the research software engineers who build and maintain a digital collections infrastructure. Incorporating mechanisms for facilitating human connections and expertise within a digital collections infrastructure is necessary for a rich and diverse research environment.

4.4 Long Term and Environmental Sustainability

Digitisation, digitalisation and a digital collections infrastructure has a carbon footprint. Thinking critically about the technology we are using and the impact it has on the environment is increasingly important. Researchers want sustainable practices to be integrated; to move towards an environmentally sustainable digital collections infrastructure it is important to critically examine current practices and underlying

assumptions about current digitisation, digitalisation and preservation practices. Altering technology is not enough, a fundamental shift in culture and mindset is needed. From the perspective of researchers, a shift in thinking has already started. It may be more appropriate to prioritise discoverability of collections and collections records, rather than mass digitisation. Determining whether there is a demonstrated need for full digitisation of an entire collection or if baseline collection records are sufficient for research purposes would be beneficial. It will also be important for a digital collections infrastructure to recognise the sector's role in advocating for and following environmentally sustainable digital practices.

4.5 Impact and Accountability

There is a perceived lack of impact and effectiveness from digital cultural heritage projects, particularly large and ambitious digital projects. Concerns were raised regarding the repetitive nature of discussions and the limited lasting tangible results from past digital collections and infrastructure projects. There is a strong desire for knowledge sharing across the sector. Establishing open communication channels for sharing best practices and tangible outcomes, along with mechanisms for tracking and assessing impact, is essential.

Facilitating opportunities for researchers to share and learn fosters a community that supports the development, continuity, and evolution of knowledge, skills, and expertise across the sector. This contributes to higher levels of impact and accountability. Additionally, providing a clear future strategy that outlines the roadmap for the sector will act as a catalyst for digital transformation. Such a strategy should articulate the potential impact of any future implementation of a UK digital collections infrastructure.

4.6 Summary and Recommendations

This user research highlights the challenges faced by research users and identifies the opportunities for a more discoverable, inclusive, interconnected, and environmentally sustainable digital cultural heritage landscape. Based on the findings, the following recommendations are proposed:

Prioritise resources for digitisation and encourage collaboration to achieve this

- Document and analyse the availability of digitised collections across the cultural heritage sector.
- Plan digitisation efforts based on specific research needs, considering both shallow and deep digitisation.
- Foster collaboration between institutions to address resource limitations and enhance digitisation programmes.
- Encourage collaboration between research software engineers, cultural heritage organisations and the academic community to address challenges and improve the use of digital cultural heritage collections for research purposes.

Enhance search and discovery functionality

- Prioritise effective keyword search and filtering features.
- Implement standardised metadata practices to enhance discoverability, as well as standardised data entry with human oversight for quality control.
- Explore the potential of AI as a way to augment collections data and search capabilities.

Promote interoperability and connection

- Establish standardised frameworks for interoperability across digital collections.
- Encourage expertise exchange between stakeholders to foster a rich research environment.
- Develop frameworks and tools that support interdisciplinary research.

Embrace sustainable practices and culture

- Incorporate sustainable practices into digitisation, digitalisation and preservation efforts to embed them within the culture of the sector.
- Evaluate the necessity of full digitisation versus baseline collection records, prioritising environmental impact to minimise carbon-intensive storage and data processing.

Assess impact and improve accountability

- Establish open communication channels for sharing outcomes and lessons learned to avoid repeated or abortive initiatives.
- Develop a network dedicated to digital infrastructure projects in the UK and internationally, serving as a platform to disseminate best practices.
- Implement mechanisms for tracking and assessing the impact of digital cultural heritage projects.
- Articulate the impact of a future UK digital collections infrastructure though establishing a roadmap for the sector.

Appendix A - Interview Protocol

TaNC User Consultation Interview Protocol

User study of researcher use of digital cultural heritage collections

Objective: The interview aims to gain a comprehensive understanding of how academic researchers use digital cultural heritage collections and interact with digital infrastructure, including their challenges, preferences, and suggestions for improvements.

Duration: Approximately 45-60 minutes per interview

Participants: researchers with experience in using digital cultural heritage collections and digital infrastructure for their research.

Materials:

- For online interviews: Zoom with built in transcription
- Audio Recording equipment (with participant consent)
- Note taking materials

Interview Protocol:

Ensure that the interview process is flexible enough to allow participants to elaborate on topics of interest. Based on the participant's responses, ask follow-up questions to delve deeper into specific topics or to gain clarity.

1. Introduction (5 minutes)

- Greet participants.
- Introduce the purpose of the interview and the study's objectives (from information sheet)
- Verbally read the consent form (if they have not signed) and how data will be kept confidential)
- Explain the importance of their insights and how their feedback will be used.
- Introduction to Digital Infrastructure – What is a UK digital collections research infrastructure? The honest answer is we don't really know, but collecting your needs and requirements is part of shaping this. Towards a National Collection has framed the idea of a digital collections infrastructure as the following:
    - A UK digital collections research infrastructure will create new technologies and new knowledge
    - A UK digital collections research infrastructure will look to increase access to and connectivity across both existing and new digital cultural heritage collections.
    - Digital cultural heritage collections are made up of digitised versions of physical objects as well as born digital assets through images, metadata, and other visual and audio materials.
    - Materials come from museums, galleries, libraries, archives and heritage organisations, including content generated by communities.

        - A UK digital collections research infrastructure will create a common approach to technology, content and data for the cultural heritage sector, and build new technologies for exploring digital cultural heritage collections

2. Participant background (5 minutes)

- Ask the participant to briefly introduce themselves, including their field of research, career level (PhD, ECR, Mid-Career, Senior-Career) and their experience with digital cultural heritage collections and digital infrastructure.

3. Research practices (10 minutes)

- Explore how the participant incorporates digital cultural heritage collections into their research. Ask about examples of Digital Collections they use.
- Ask about their typical research workflow, from identifying resources to using them in their work.
- Discuss their preferred methods for discovering, accessing, and citing resources from digital collections.
- What types of content/data are they looking for from digital collections? Does their work involve metadata or rich datasets?
- Ask about search behaviours – e.g. Information Seeking or Browsing

4. Interactions with digital infrastructure (15 minutes)

- Discuss the participant's experience with various digital platforms (from the examples they have already stated), tools, and resources.
    - Inquire about usability, user interface, search capabilities, and any difficulties encountered.
    - Prompt the participant to share any challenges they've faced when using digital collections or infrastructure.
- Explore issues such as data availability, interoperability, accessibility, or technical limitations.

5. Features and tools (10 minutes)

- Ask about specific features or tools that the participant finds valuable when working with digital cultural resources.
- Inquire about any additional features they wish existed in their current practice to enhance their research process.

6. Suggestions and future directions (10 minutes)

- Encourage the participant to provide suggestions for improving digital cultural heritage collections and digital infrastructure.
- Ask them to share their vision for the ideal digital collection research environment.
- If time allows a discussion about sustainability and ethics of a digital collections infrastructure.

7. Closing remarks (5 minutes)

- Thank the participant for their time, insights, and contributions.
- Explain how their feedback will influence future developments and enhancements.
- Optional: Follow-up Questions (if needed)

Appendix B - Focus Group Protocol

TaNC User Consultation Focus Group Protocol

Focus group protocol for a user study of academic and IRO researcher use of digital cultural heritage collections designed to gather insights, opinions, and experiences from participants.

User study of academic researcher use of digital cultural heritage collections

Objective: The focus group aims to understand how academic researchers utilise digital cultural heritage collections, their needs, challenges, and suggestions for improvements.

Duration: Approximately 2.5hours

Participants: Academic/IRO researchers with experience in utilising digital cultural heritage collections for their research.

Materials:

- Presentation providing context: the purpose of the study, definition of a digital collections research infrastructure and the focus group process.
- Flipcharts or whiteboards and markers for visual note-taking.
- Padlet for online visual note-taking
- Audio recorders x2

Facilitator: Claire Bailey-Ross

Assistant: Research Assistant to help with logistics, note-taking, and technical issues.

Agenda:

1. Welcome and introduction (15 minutes)

- Greet participants.
- Introduce the purpose of the focus group and the study's objectives.
- Explain the importance of their insights and how their feedback will be used. - Introduction to Digital Infrastructure – What is a UK digital collections research infrastructure? - Space where participants should feel encouraged to share their honest opinions. Participants responses will be anonymised.
- Respect time and keep the discussion focused on the outlined topics. Record the discussions for accurate analysis and reporting.
- Information Sheet and Consent forms.

2. Participant introductions (10 minutes)

- Each participant provides a brief introduction, including their field of research, career level (PhD, ECR, Mid-Career, Senior-Career) and their experience with digital cultural heritage collections.
- Switch Audio Recording on after Introductions.

3. Discussion on current practices (20 minutes)

- Ask participants about their typical approaches to using digital cultural heritage collections in their research.
    - Discuss their preferred methods for discovering, accessing, and citing resources from digital collections.
    - Examples of Digital Collections they use.
    - Types of Content.
    - Data Types.
    - Search Behaviours.
    - Information Seeking or Browsing.
- Explore challenges participants face, such as search difficulties, usability issues, or limitations in the collections.
- Explore issues such as data availability, interoperability, accessibility, or technical limitations

4. Exploration of needs (20 minutes)

- Present hypothetical scenarios or real-life examples to prompt discussions about specific needs and pain points.
- Digital surrogates?
- What level of digitisation is needed? E.g. High-quality images? 3D objects? IIIF? - Raw data or cleaned data?
- Metadata.
- Rich datasets?
- Data aggregation?
- Is Linked Open Data a need?
- Data standards.
- Levels of Digital Literacy – training support for a digital collections research infrastructure? - Explore how participants collaborate with peers using digital cultural heritage resources. - Discuss any challenges they encounter when sharing or integrating resources into collaborative projects.

5. Break – networking and coffee (15 minutes)

6. User interface and experience (15 minutes)

- Show screenshots or examples of different digital cultural heritage platforms (from examples the participants have already stated they use).
- Discuss the usability, user interface, and overall experience of these platforms. - How/should user interface of a digital collections research infrastructure compare to individual digital cultural heritage collections?
- Gather feedback on what works well and what could be improved.

7. Features and tools (15 minutes)

- Discuss specific features or tools that participants would find valuable in digital collections research infrastructure.
- Inquire about any additional features they wish existed in their current practice to enhance their research process.

8. Digital collections research infrastructure ideation (20 minutes)

- Ask participants to brainstorm suggestions for developing a digital collections research infrastructure for academic research purposes.
- Encourage participants to think about new features, improved search capabilities, or better integration with pre-existing research tools.
- If time allows a discussion about sustainability and ethics of a digital collections infrastructure

9. Closing remarks (5 minutes)

- Thank participants for their valuable insights and contributions.
- Mention how their feedback will be used to shape future developments in TaNC.

10. Optional: Follow-up interviews (if needed)

- If certain topics require further exploration, offer the possibility of follow-up interviews for in-depth insights.

Appendix C - Copy of the Survey instrument

TaNC User Consultation Survey

UK Digital Collections User Consultation Introduction

If you use Digital Collections from Galleries, Libraries, Archives, and Museums as part of your research, we invite you to complete this survey.

Your feedback is important for shaping the future of UK digital collections infrastructure and building a case for its future funding. Historic Environment Scotland has commissioned the University of Portsmouth to conduct a user consultation to identify researchers' needs and requirements, helping to define what should be included in a future UK digital collections infrastructure.

This survey asks about your research process and thoughts on the use of Galleries, Libraries Archives and Museums (GLAM) digital collections. It also provides you with an opportunity to express your views on the requirements and future direction of a GLAM digital collections infrastructure. Survey results will be important for influencing future support from the Arts and Humanities Research Council (AHRC) in the GLAM sector.

The survey will take **no longer than 10 minutes to complete**, and your responses will remain entirely confidential. They will only be accessible to the University of Portsmouth research team. Once the survey concludes, we will analyse the findings and present overall trends to Historic Environment Scotland. No individual respondents will be identified in our reporting.

This research is part of the "Towards a National Collection" (TaNC) programme, which aims to open up and democratise access to cultural heritage collections across the UK. This user consultation will help to ensure that the future infrastructure meets the needs of the people who will be using it.

If you have any questions about this survey, please contact Dr Claire Bailey-Ross at claire.bailey-ross@port.ac.uk

We sincerely thank you in advance for your time and valuable input.

* 1. I have read and understood the terms of participation as outlined above, and I agree to proceed with taking the survey

☐ Yes

TaNC User Consultation Survey

About your role and organisation

Before we delve into the specifics, please tell us about your background as an arts and humanities researcher. This will help us understand your unique perspective and requirements. These will help with our analysis and will not be used to identify you in reporting.

* 2. In which type of research institution do you work?

- ◯ University
- ◯ Research Centre
- ◯ Library
- ◯ Archive
- ◯ Museum and Gallery
- ◯ Historic Environment/Heritage
- ◯ More than one of the type of organisation above
- ◯ None of the above
- ◯ Other - please describe

3. What is your Research Field/Area of Study:

4. Which of the following best describes your career stage?

- ◯ PhD Student, Junior Research Software Engineer
- ◯ Early Career (e.g. Research Assistant/Associate, Assistant Curator, first grant holder, Lecturer, Assistant Professor, Research Software Engineer)
- ◯ Mid/Recognised career (e.g. Senior Lecturer, Curator, Reader/Associate Professor, Senior Researcher, Senior Research Software Engineer, Research Software Group Leader)
- ◯ Senior/Established/Experienced Career (e.g. Professor, Senior Curator, Director of Research, Distinguished Engineer, Chief Data Scientist)

5. What is your job/role title?

6. Which institution/organisation do you work for?

TaNC User Consultation Survey
Digital Collections data from UK's Galleries Libraries Archives Museums and heritage

*7. For your research, do you access or use any digital collections material held by galleries, libraries, archives, museums or other heritage-sector organisations in the UK?

- ◯ Yes
- ◯ No

TaNC User Consultation Survey

8. If no, why do you currently not use digital collections held by galleries, libraries, archives, museums or other heritage-sector organisations (GLAM) as part of your research?

- [ ] I am not aware of relevant GLAM digital collections'
- [ ] I prefer to use physical collections Preference for physical collections
- [ ] Digital Collections have limited relevance to my research field
- [ ] I find it difficult to access GLAM digital collections
- [ ] I have concerns about copyright and licensing restrictions
- [ ] I find digital collections have inadequate search and discovery tools
- [ ] I find digital collections have poor data quality and accuracy
- [ ] Other (please specify)

Digital Collections data from UK's Galleries Libraries Archives Museums and heritage (GLAM)

9. What galleries, libraries, archives, museums or other heritage-sector organisations (GLAM) data formats do you find most useful in your research? (Select all that apply)

- [ ] Text (e.g. books, newspapers, encyclopaedias, archival documents and other historical sources)
- [ ] Numerical
- [ ] Digital Surrogate 2D material (e.g. Photographic images or videos)
- [ ] Digital Surrogate 3D material (e.g. 360 photos or videos, 3D digitisation mesh)
- [ ] Born Digital 2D material (e.g. computer aided design (CAD) data, tags, associations, texts)
- [ ] Born Digital 3D material (e.g. 3D modelling/sculpting and animations)
- [ ] 4D material (e.g. immersive environments for VR/AR/MR, motion capture, sensor data for digital twins, motion capture datasets)
- [ ] Catalogues, databases
- [ ] Maps and Geospatial data (e.g. maps, 3D point clouds, digital terrain models)
- [ ] Audio data (e.g. music, voice recordings, oral histories)
- [ ] Other (please specify)

10. Do you face any particular challenges when accessing or using digital library, archive, museum, or gallery collections in your research? (Select all that apply)

- [ ] Limited search capabilities
- [ ] Limited interoperability
- [ ] Complex user interface
- [ ] Poor data quality and accuracy
- [ ] Incomplete metadata
- [ ] Lack of collaboration tools
- [ ] Inadequate customisation options
- [ ] Insufficient preservation of digital materials (e.g. inappropriate file formats)
- [ ] Absence of data analysis tools
- [ ] Lack of integration with research tools
- [ ] Inadequate user support and assistance
- [ ] Other (please specify)

11. How important is collaboration in your research projects?

- ◯ Very Important
- ◯ Important
- ◯ Neutral
- ◯ Less Important
- ◯ Not Important

12. Are there specific features or functionalities that you believe are missing from existing GLAM digital collections that could improve your research experience? (Select all that apply)

- [ ] Advanced search capabilities including cross collection searching
- [ ] More user-friendly interface
- [ ] Improved data quality and accuracy
- [ ] Open data sets and collections
- [ ] Enhanced metadata completeness
- [ ] Better collaboration tools
- [ ] Increased customisation options
- [ ] Stronger preservation of digital materials
- [ ] Additional data analysis tools
- [ ] Improved integration with research tools
- [ ] Enhanced user support and assistance
- [ ] Other (please specify)

13. Do you encounter any issues related to copyright or licensing when using digital materials for research purposes? (Select all that apply)

- [ ] Yes, copyright restrictions limit usage
- [ ] Yes, licensing issues hinder research
- [ ] No, I haven't encountered copyright or licensing issues
- [ ] No, I only use open access materials
- [ ] Not applicable

14. What computational research tools are you using or would you be interested in exploring for your research? (Select all that apply)

- [ ] Data Cleaning
- [ ] Text/Data Mining
- [ ] Data Visualisation (including 3D)
- [ ] Database Design
- [ ] Immersive Technologies - Virtual Reality (VR) / Augmented Reality (AR) / Mixed Reality (MR)
- [ ] Statistics
- [ ] Machine Learning / Artificial Intelligence / Generative Adversarial Networks
- [ ] Working with Geospatial data
- [ ] Natural Language Processing (NLP)/ Textual Analysis /Sentiment Analysis
- [ ] Computer Vision
- [ ] Topic Modelling
- [ ] Semantic web
- [ ] Other (please specify)

15. To what extent do you agree with each of the following statements?

	Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree
I am confident in navigating and utilising digital collections relevant to my field of arts and humanities research.	○	○	○	○	○
I am familiar with and utilise open source tools that are specifically designed for arts and humanities research.	○	○	○	○	○

I employ technology to					
automate aspects of my research workflow, optimizing efficiency in data collection, analysis, and documentation.	○	○	○	○	○
I have a strong understanding of digital preservation methods and practices relevant to safeguarding cultural and historical materials.	○	○	○	○	○
I possess advanced programming skills, allowing me to develop customized solutions and tools to address specific research challenges.	○	○	○	○	○
I am adaptive to emerging technologies in the arts and humanities landscape, incorporating new tools and methodologies into my research practice.	○	○	○	○	○
I prefer to use physical collections than digital where possible (score inverted so that online preference scores are higher)	○	○	○	○	○

Future Digital Collections Infrastructure for UK's Galleries Libraries Archives Museums

16. Please rate the importance of the following features in a digital collections infrastructure for your research (1 = not important, 5 = extremely important):

	Very Important	Important	Neutral	Less Important	Not Important
Search capabilities	○	○	○	○	○
User-friendly interface	○	○	○	○	○
Data quality and accuracy	○	○	○	○	○
Metadata completeness	○	○	○	○	○
Collaboration tools	○	○	○	○	○
Customisation options	○	○	○	○	○
Preservation of digital materials	○	○	○	○	○
Data analysis tools	○	○	○	○	○
Integration with pre-existing research tools (e.g., reference management software)	○	○	○	○	○
User support and assistance	○	○	○	○	○

Other (please specify)

17. In your opinion, should a digital collections infrastructure include only Open Access and equivalently licensed content?

- ◯ Yes, only Open Access and equivalently licensed content should be included
- ◯ No, it should include restricted access materials and Open Access and equivalently licensed content
- ◯ Don't Know
- ◯ Other (please specify)

18. In your opinion, would the adoption of an open data approach, allowing third parties to build new products and services on top of a National Digital Collection infrastructure, be beneficial to your research?

- ◯ Yes
- ◯ No
- ◯ Don't know

19. What kind of support would you find useful to help you to use a future digital collections infrastructure in your research?

- ☐ clear information in plain language that explains the benefits of the digital collections infrastructure and how to become a user
- ☐ training on how to use the digital collections infrastructure
- ☐ IT support
- ☐ training on digital skills
- ☐ active user community
- ☐ appropriate IT equipment
- ☐ free storage
- ☐ Other (please specify)

20. Which of the following tools that a future digital collections infrastructure could eventually develop would be most useful for your research? (select 3)

- [ ] Tools for creating, sharing and re-using digital content
- [ ] Tools for data visualisation and exploration
- [ ] Tools for AI-assisted metadata enrichment (e.g., to assist in making cultural heritage content findable)
- [ ] Tools for advanced interaction with the digital content (e.g., enabling new ways of cross-mixing, remixing content)
- [ ] Tools for analysing, designing and testing interactions with visitors (e.g., to help design and test exhibitions)
- [ ] Tools for conservation and restoration based on the digital collections
- [ ] Tools for data cleaning and preparation
- [ ] Tools for text analysis and natural language processing
- [ ] Tools for citizen science or crowdsourcing
- [ ] Tools for semantic web and linked data applications
- [ ] Other (please describe)

21. What features or tools would you like to see for maximising collaboration in a digital collections infrastructure?

22. Which of the following aspects are most important to you in a digital collections infrastructure? (select 3)

- [ ] Accessibility: Ensure that the digital collections infrastructure is designed inclusively to accommodate diverse audiences with varying abilities and expertise.
- [ ] Analytical Functionality: The digital collections infrastructure should provide analytical tools to interact with the content, enhancing its utility for research.
- [ ] Compliance: The digital collections infrastructure should promote and facilitate best practices in managing digital collections.
- [ ] Content Creation: The digital collections infrastructure should enable contributors to generate derivative products, subject to permissions.
- [ ] Discovery: Enable effective search and filtering features within the infrastructure, with specific criteria to be defined during the scoping process.
- [ ] Engagement and Training: The digital collections infrastructure actively engage with diverse user communities, offering guidance, training, and best practices.
- [ ] Equitability: Ensure that the infrastructure addresses accessibility and decolonization agendas, promoting fairness.
- [ ] Impact Metrics: The digital collections infrastructure should provide reporting tools for assessing the impact of collections.
- [ ] Learning: The digital collections infrastructure should provide opportunities to enhance learning experiences through access to content.
- [ ] Sustainability: The digital collections infrastructure should inspire confidence in the infrastructure's long-term sustainability.
- [ ] Preservation: The digital collections infrastructure should ensure the long-term availability of collections data.
- [ ] Public Engagement: Extend the reach of the infrastructure beyond primary audiences.
- [ ] Reuse: The digital collections infrastructure should maintain accessibility to data for ongoing research and reproducibility.

23. What do you want a future digital collections infrastructure to do for Galleries Libraries Archives Museums and heritage? Please rank in order of importance (where 1 is most important and 4 is least important):

- ☐ Connect and cross-search data from different institutions
- ☐ Enable citizen science such as crowdsourcing research
- ☐ Open multiple collections to current computational research tools
- ☐ Support the decolonisation of collections

24. Please use this space to provide any additional comments, suggestions, or specific requirements you have regarding digital collections infrastructures and the use of digital materials in your research.

Appendix D - Institutional Affiliation

Interview and Focus Group Institutional Affiliation

Institutional Affiliation	Count
Aberystwyth University	1
Ashmolean Museum	1
Bodleian Libraries, University of Oxford	1
British Library	1
Cardiff University	4
Durham University	11
Irish Traditional Music Archive	1
Jamie's Computers	2
Lancaster University	2
Maynooth University	1
Museum of Farnham	1
Museum of Wales	1
National Records of Scotland	1
Newcastle University	1
Portsmouth History Centre	1
Queen's University Belfast	3
Science Museum Group	1
School of Advanced Study	1
The Alan Turing Institute	3
The National Archives	2
The Parliamentary Archives	1
The Primary Trustworthy Digital Repository Authorisation Body (PTAB)	1
Tyne & Wear Archives & Museums	1

University for the Creative Arts	**1**
University College London	**5**
University of Aberdeen	**1**
University of Bristol	**1**
University of Cambridge	**1**
University of Exeter	**5**
University of Portsmouth	**12**
University of Southampton	**3**
University of Sussex	**1**
University of the Arts London	**1**
V&A	**1**

Survey Institutional Affiliation

Survey participants were asked which institution/organisation they work for, this was an optional question which received 130 responses.

Below is a table of responses where there was more than one respondent:

Institutional Affiliation	Count
University of Edinburgh	7
University of Liverpool	6
University of Oxford	6
University of Nottingham	6
University of Exeter	5
Historic England	4
Birmingham City University	3
Durham University	3
King's College London	3
University College London (UCL)	3
Bangor University	2
Brunel University London	2
Goldsmiths, University of London	2
Historic Environment Scotland	2
Lancaster University	2
National Museum of the Royal Navy	2
The Open University	2
Royal Commission on the Ancient and Historical Monuments of Wales	2
University of Manchester	2
University of Reading	2
University of Warwick	2

There were institutions that provided a single response, including: Aberystwyth University, Bath Spa University, Bournemouth University, British Library, British Museum, Cambridge University Library, Courtauld institute of Art, English Heritage, Glasgow Museums, Newcastle University, Oxford Brookes University, Parliamentary Archives, Science Museum Group, The National Archives, The National Gallery, University of Glasgow, University of Bristol, University of Bristol, University of Hull, University of Kent, University of Leeds, University of Portsmouth, University of Sheffield, University of St Andrews, University of York, University of the Arts London.

The distribution of responses across institutions indicates that there is a diverse representation of participants from various universities and organisations. With 67 UK institutions represented, though representative, this does not show the whole picture. A complete mapping of the UK research landscape is beyond the scope of this survey, but the institutional profile of the participants offers several useful insights of the landscape.

The University of Edinburgh has the highest count among the listed institutions, suggesting a relatively strong presence or engagement of individuals from arts and humanities disciplines with an interest in digital collections.

References

Woodley, S., & Towell, P. (2022). User Research. Zenodo. https://doi.org/10.5281/zenodo.6684165