

**Digital Humanities and Science Museum Group:
A Landscape Study**

Rhiannon Lewis
October 2022

Table of Contents

Introduction	3
Methods.....	4
Initial Conclusions.....	4
SWOT analysis.....	7
Strengths	7
Weaknesses.....	8
Opportunities	11
Threats	13
Acknowledgements	15
Appendix I - Annotated glossary of DH methods for heritage collections.....	16
3D scanning / 3D imaging.....	16
Application programming interface (API).....	17
Augmented reality (AR)	18
Computer Vision	20
Crowdsourcing.....	21
Data Visualization	22
Emulation.....	23
Geographical Information System (GIS)	24
Handwritten text recognition (HTR).....	25
Knowledge Graph	26
Linked Open Data (LOD).....	26
Natural Language Processing (NLP).....	27
Network analysis	28
Optical character recognition (OCR).....	30
Persistent identifiers (PIDs)	31
Virtual reality (VR)	32
Input/output type table.....	33
Appendix II – Existing SMG DH projects	34
Appendix III - Ideas for DH projects in SMG.....	39

Introduction

The Science Museum Group (SMG) has a long history of leading and participating in research and holds UK Research and Innovation (UKRI) Independent Research Organisation (IRO) status. SMG staff are principal investigators or co-investigators in numerous funded research projects, SMG publishes the peer reviewed *Science Museum Group Journal*, and it hosts Arts and Humanities Research Council (AHRC) Collaborative Doctoral Partnership awards that have seen 66 studentships awarded since 2013.

Digital Humanities (DH) has emerged in recent years as a set of approaches that combine computational methods with humanities content and questioning. These approaches are highly diverse and address a wide range of research questions. There are growing grant calls from funding bodies – notably the AHRC and UKRI more generally – to support DH projects in higher education institutions (HEI) and IROs. There is therefore an emerging opportunity for SMG to undertake more work in this area.

Since DH is a growing area in the cultural heritage sector, this six-month landscape study hopes to shape SMG's approach to it. This landscape study aims to enable SMG to take a *strategic* approach to funding applications and areas of research interest going forward. It presents the most relevant approaches to cultural heritage from the wide range of ever-developing digital humanities methods available, while considering the affordances of SMG's collection and existing resources, both digital and analogue.

The key research question was: *how can digital humanities methods and practice address SMG's research and public engagement ambitions around its collection?* As part of this, the report also sought to explore:

- *What, if any, action does existing digitised collection material require so that it can be preserved and accessible?*
- *How can digital or digitised assets be best matched with current and/or evolving DH methods?*
- *What curatorial research interests can be explored/realized through DH methods that otherwise would not have been possible?*

Methods

This landscape study uses qualitative methods including interviews and workshops, along with desk research. The interviews and workshops were with SMG staff, as well as academic researchers affiliated with the Group.¹ Secondary desk research was used to situate digital humanities practice within current projects and literature. The report employs SWOT (Strength, Weakness, Opportunity and Threat) framework analysis – a tool for strategic planning – to evaluate SMG’s current position, to understand how successfully Science Museum Group (SMG) are enabling digital humanities methods and practice, and to address their research and public engagement ambitions around its collection.

Initial Conclusions

There are two main ways to enable DH research in a cultural heritage museum setting. The first is making collections catalogue data open access and as interoperable as possible. SMG can continue to take an “if you build it, they will come” approach, employing open data, making the collections digitally accessible for research. However, if taking this approach, it is important to be strategic. Technology and digitisation evolve quickly, and it is therefore advisable to anticipate where digitisation will be going and work to that, rather than building infrastructure and

¹ Dr Alexander Butterworth, Digital Humanities and SMG interview, 2022

Dr Rachel Boon, Ben Russell, Jessica Bradford, and John Stack, Workshop: Digital Humanities methods for SMG collection, 2022

Jack Kirby, Geoffrey Belknap, Ed Bartholomew, and Dr Tilly Blyth, Workshop: Digital Humanities methods for SMG collection, 2022

Professor Ross Parry, Digital Humanities and SMG interview, 2022

Somaya Langley, Digital Humanities research project - interview, 2022

producing assets for where it currently is.² The second is undertaking DH research through a partnership model, allowing for sharing of specialist knowledge and putting funded projects together. These partnerships are often built around a specific part of the collection, or research question. Although differing, these approaches are not mutually exclusive.

SMG staff are skilled at communicating collections narratives and are already engaging with DH tools and methods to do so. Appendix II details a non-comprehensive list of digital humanities projects produced by, or in collaboration with, SMG. As a result of these projects, there is a breadth of knowledge deriving from working with DH methods on collections research that has already been a practice within the Group for some time.

DH is an evolving field. There is ambiguity about what is meant by “Digital Humanities” and what its methods are, as well as what they involve. A conclusive definition of DH is outside of the scope of this report. However, the fact that people conceptualize DH in different ways means it is difficult to be able to communicate what could be possible. A non-exhaustive list of DH methods is introduced in the Annotated Glossary (Appendix I). A distinction has been made between collection digitisation (cataloguing and imaging of collection objects) and DH methods. Once digitisation has happened to collections and sources, it is possible to use these digitised materials for DH research. Given recent and ongoing digitisation of collections (approx. 318,000 objects online and 150,000 objects with images at the time of writing), SMG has the potential to capitalize on this investment and use these digitised collections as the foundation for more ambitious and varied DH research.

This landscape study sought to understand where DH sits in the workings of the Group, before moving forward. An in-depth understanding of digital labour more generally throughout the Group is beyond the scope of this report but can be addressed via the work of Dr Sophie Frost as

² Hughes, Lorna M., *Digitizing Collections: Strategic Issues for the Information Manager*, 1st edition (London: Facet Publishing, 2003). 285

part of the One by One: Project #3 research project³. However, primary research has suggested that digital research is already happening across digital, collections and research teams. There is much room for further DH preservation infrastructure, and this is incorporated in the wider and in-progress digital preservation work already being undertaken, although the preservation of DH outputs has been identified as being more complex in nature than the wider digital preservation needs of SMG.⁴ If the Group hopes to expand further into DH collections research, it will have to continue to work with specialist partners in the HEI sector and hire or train staff with DH skills.

The current culture of digital collections research at SMG is driven largely from the curatorial department with a handful of projects also initiated by the Research and Digital departments.⁵ There is not one set definition of the curatorial role. It would be inappropriate to attempt such a definition here, though we may note that the skills that curators have and the technologies they employ have always been in a process of change and evolution. SMG will need to enable curators to gain further DH skills and experience, potentially by encouraging curators to act as secondary partners on projects. Given that DH research will be new to many curatorial and other staff, they will need to have the time within their working schedules to work on such projects and this maybe a greater proportion of time than is usual for more traditional research projects. SMG staff with specialist DH knowledge could be given time to experiment with tools and methods, as there is the potential to produce more innovative results if they are allowed space to develop. If SMG desires to be the lead in more DH projects rather than acting as a secondary partner (e.g. PI rather than CI roles) or content provider,⁶ staff will need to grow or build upon their DH skills.

³ Forthcoming six-part podcast series *Hidden Constellations*, released Autumn 2022, identifies five key kinds of digital labour practice at SMG: new, hidden, distributed, legacy and collective. Dr Sophie Frost, 'The Hidden Constellation' 'Project #3 (2021-2022) Modelling New Digital Leadership in Museums', One by One, 2021 <<https://one-by-one.uk/2021/08/11/project-3-summary>> [accessed 4 August 2022]

⁴ Suggestion by John Stack, Digital Director SMG

⁵ Suggestion by John Stack, Digital Director SMG

⁶ Suggestion by John Stack, Digital Director SMG

SWOT analysis

Strengths

These are of internal origin, they are attributes of the organization, and are helpful to achieving the objective. What is currently being done well by SMG to ensure DH methods and practice are in place to address the SMG's research and public engagement ambitions around its collection? What unique resources can SMG draw on for this? What are perceived as SMG's strengths in digital humanities by the workshop participants?

- **S.1 Collections**

- **S.1.1 Existing digital assets**

SMG has substantial digitisation of collection objects and most objects held in its collections are catalogued within a collection management system (CMS). SMG collections online, which draws on data from this CMS, has open-data principles applied to the collection,⁷ and SMG has already undertaken several digital humanities projects using existing digital assets⁸. There is the potential for further digitisation of physical collections building on this strength.

- **S.1.2 Discovery**

SMG staff are often finding collection items through collections online and in picture library.

- **S.1.3 Collection stories**

Conveying collection stories is already an embedded way of working.

- **S.2 Digital Infrastructure**

SMG is embarking on a five-year programme of transformation in digital preservation and digital collecting. In this moment of change, there is an opportunity to strategically plan.

- **S.3 Existing DH methods use**

Some DH tools and practices are already widely used within the Group. SMG is employing a few different DH strategies to explore object stories and convey

⁷ Open data available under a Creative Commons Zero (CC0) license
<https://www.sciencemuseumgroup.org.uk/about-us/collection/using-our-collection-api/> &
https://public.resource.org/8_principles.html

⁸ See Appendix II

collection narratives. See existing SMG DH projects in Appendix II for those completed/in progress.

- **S.4 Collaboration**

Existing DH networks. Partnerships with academic bodies, other IROs and industry (see SMG existing projects Appendix II for collaborators). SMG is working with users and audience on DH methods such as crowdsourcing, to produce research outcomes together.

- **S.5 Engaging audiences**

Engaging audiences through co-production or in the process of research, leading to the enrichment of research processes, representation and findings.

- **S.6 In-house specialists**

They can scope and support DH projects and can be found across SMG teams in collections services (Digital Preservation Manager), Research and Public History, Digital, and Collections (Curator role: subject specialist, programmer, facilitator, instigator).

- **S.7 Intellectual property (IP)**

- **S.7.1** Collections catalogue data is available as open data under an open access Creative Commons Zero (CC0) license.
- **S.7.2** Outputs using open-source code are ideal, meaning others can build on the work produced. Projects like *Heritage Connector* have produced outputs with open-source code.

Weaknesses

These are of internal origin, they are attributes of the organization, and are harmful to achieving the objective. What is it that SMG are doing that is preventing or hindering DH collections research and public engagement? What needs to be improved? Where does SMG have fewer resources for DH than comparative organizations? What did the workshop participants see as weaknesses?

- **W1 Resources**

- **W.1.1 Funding**

Currently, additional funding is needed for DH projects; this takes time and if projects are one-off or discrete this must be done individually. Where would

budget come from if it were not being fundraised for? Where does this sit within the Group structure?

- **W.1.2 Limited staff time**

Staff working schedules are at, or near, capacity. If DH projects were additional to existing workload, staff would not have time to plan and implement other projects or keep abreast of evolving DH discourse. When at capacity, staff have time only for the most pressing tasks and maintain existing systems. Investment of time is needed if the goal becomes to upskill existing workforce in DH techniques and produce innovative research outputs.

- **W.1.3 Staggered digitisation**

Museum collections and archives get digitised to varying specifications – some objects have photogrammetry, some have images, some only have catalogue metadata – depending on the time and the reason(s) why they were being digitised. SMG has a large collection with staggered digitisation and documentation of its collection, resulting in an imperfect data set.

- **W.1.4 Longevity**

Some but not all projects employ open-source code and therefore can be reused by others in their own DH work. SMG has a corporate records preservation policy, but infrastructure to support DH preservation still needs to be implemented, therefore project outputs and processes could be lost.

- **W.2 Gaps in existing collection data**

- **W.2.1 Thin data in SMG collection catalogue**

The SMG's is a very large collection,⁹ in which expansive catalogue entries are rare. Publicly accessible collections catalogues often quite properly exclude sensitive material and fields. Documentation has a legacy of differing approaches to cataloguing over the decades. SMG's collections online is built on the legacy of

⁹ '7.3 million items' 'Our Collection', Science Museum Group <<https://www.sciencemuseumgroup.org.uk/about-us/collection/>> [accessed 2 August 2022]

systems and processes that have gone before i.e., digitisation of card catalogue systems, and inconsistent field entry in the Mimsy collection management system.

- **W.2.2 Amplification of gaps in collection**

There are gaps in the Science Museum Group collection, and documentation of the collection. There is potential, through employing DH methods, to amplify these gaps.

- **W.3 Digital Infrastructure**

- There is not currently infrastructure in SMG to meaningfully support DH preservation. The ideal would be to embed DH preservation into the workflow. '[...] put foundations in place, so you've got infrastructure, to enable digital humanities.'¹⁰
- DH preservation needs not just to be something a single researcher has access to, but rather it needs to be a way of working across the organization.¹¹

- **W.4 Ways of working / mindset**

- **W.4.1 One-off projects**
Isolated DH projects, rather than a growing and evolving programme.
- **W.4.2 Large organization**
 - **W.4.2.1** SMG is a large organization with five museums, two collection storage sites and multiple teams within that. Therefore, there are multiple institutional structures and ways of working to navigated.
 - **W.4.2.2** If staff member or department has not instigated a DH project before, they may not be sure how to do so, or lack the experience to construct and realize opportunities.
- **W.4.3 Curator workload and role expectations**
Limited capacity for additional work, as they are stretched for time and have large collections to oversee. Curators, even within the Group, have different remits and

¹⁰ Somaya Langley, Digital Humanities research project - interview, 2022

¹¹ Somaya Langley, Digital Humanities research project - interview, 2022

skills. Where do the roles that are working on digital collections and DH research sit within the SMG structure? How does the group conceptualize their position?

- **W.4.4** What are digital humanities?

DH is a new academic area with diverse and evolving terms and methods therefore, people have their own perspective and assumptions of what it could be. Trends in DH are evolving and changeable. There is potential difficulty, keeping up if staff already at capacity with other work. (The approach to DH that this report takes is explained in the intro and explored in expanded glossary in Appendix I.)

Opportunities

These are of external origin, are attributes of the environment (rather than organization) and are helpful to achieving the objective. What helpful external attributes could enable SMG to realize its research and public engagement ambitions around its collection with DH? What of the SMG's strengths in DH could be turned into an opportunity? What trends can SMG take advantage of? What DH opportunities could be seized upon by SMG?

- **O.1 Partnerships**

It is possible to work with a variety of research partners – inclusive of but not limited to HEIs, other GLAM (galleries, libraries, archives and museums) organizations, IROs, SMEs, etc. There are skills-sharing opportunities between organizations through partnerships including digital skills. However, not all factors for facilitating DH *are* digital. Good partnerships can also be built on holistic factors like '[...] trust, respect and empathy'¹², on 'coherent institutional support' and willingness to translate specialist knowledge.¹³

- **O.2 Existing tools**

- **O.2.1** There are numerous existing off-the-shelf tools that do not require high levels of digital literacy to use.
- **O.2.2** There is the potential to build on open-source code or tools produced in existing DH projects. (See Appendix II for existing SMG DH projects).

¹² Professor Ross Parry, Digital Humanities & SMG interview, 2022.

¹³ Dr Alexander Butterworth, Digital Humanities & SMG interview, 2022

- **O.3 Scope for DH methods**

There are DH methods that have not yet been applied to the collection. (See Appendix I for project method examples.)

- **O.4 Funding**

There has been money for HEIs and IRO coming from funding bodies (e.g. AHRC via Towards a National Collection [TANC]).

- **O.5 Creative applications and outcomes of digital humanities research**

- **O.5.1** Potential for novel ways of conveying data findings (for example data visualization and data sonification).

- **O.5.2.** Exciting call to action for what DH could be...

Exciting DH/public history output from public perspective?

- **O.5.2.1** Museum not only as 'a holder of kind of artifacts and of knowledge in space' but as a space of presentation of research findings (virtual or augmented), of collaborative practice, and building communities of interest: '[...] the institution as a kind of nurturing center.'¹⁴

- **O.5.2.2** Public participation in process of research, including collaborative practice.

- **O.5.2.3** Humanities – inclusive of DH – 'constructively disruptive to the very notion of what a museum is.'¹⁵

- **O.6** DH not yet defined

Therefore, there is space for experimentation.

- **O.7** What is a curator?

The skills that curators have had historically have always been in a process of change and evolution, as well as being subject to different conceptualizations. Perhaps this is another moment of change, or a continuation of changing skills that curators possess.

¹⁴ Dr Alexander Butterworth, Digital Humanities & SMG interview, 2022

¹⁵ Professor Ross Parry, Digital Humanities & SMG interview, 2022

Threats

These are of external origin. They are attributes of the environment (rather than organization) and are harmful to achieving the objective. What are the threats that could harm SMG's DH ambitions? What stage are similar size cultural heritage institutions at in their deployment/utilization of DH methods for better understanding their collections? What of SMG's weaknesses are exposed by which threats?

- **T.1 Funding Cycles**

Given the size, scale and extent of grants, there is potential to lead to one-off or smaller projects, but can there be longevity maintaining project methods and outcomes in this funding model? What are the limitations and conditions of digital humanities work being produced with this funding model?

- **T.2 Skills**

- **T.2.1** High economic cost of DH skills.
- **T.2.2** Difficulty attracting talent of specialists with high-end coding skills and/or data scientists. They generally command much higher salaries in other industries where their skills are in demand.

- **T.3 Cutting edge nature of technology**

- **T.3.1** Constantly changing and getting updated.
- **T.3.2** The technology may be there, but the 'cutting edge' of technology may not be the best way to research and/or present collection narratives and research.¹⁶

- **T.4 Partnerships**

There is the potential for partners to not align with SMG needs and values. Management of partnerships can be time-consuming, and effort is required to find and build those partnerships initially. However, partnerships are often a requirement for securing funding grants.¹⁷

- **T.5 Digital humanities are not yet defined**

¹⁶Professor Ross Parry, Digital Humanities & SMG interview, 2022

¹⁷ Suggestion by John Stack, Digital Director SMG

Therefore, they are conceptualized in different ways, and people can come to DH expecting it to be and achieve different things.

- **T.6 Intellectual Property (IP)**
 - **T.6.1** Some – but not all – projects use open-source code and therefore reusable by others.
 - **T.6.2** IP is negotiated when projects are a collaboration, but where does the IP sit when the collaborative project ends?
 - **T.6.3** Waiting for primary legislation changes to be resolved with open data.

Acknowledgements

This work was supported by the Additional Student Development Fund (ASDF) which was provided by the AHRC for collaborative doctoral partnership (CDP) students.

Thank you to John Stack and Dr Tim Boon for their input and guidance.

Thank you also to Dr Gabriel Bodard for the opportunity to audit the *SunoikisisDC Spring 2022*:

Digital Approaches to Cultural Heritage module.

Appendix I - Annotated glossary of DH methods for heritage collections

The methods that are introduced here may work with existing digitised materials, or they may be a method through which collections and archives could be digitised. There are fifteen terms included in this annotated glossary, although this is not an exhaustive or conclusive list, it is meant as an introduction to current DH methods. An earlier version of this glossary was provided to workshop participants and some interviewees as an introduction to DH methods for cultural heritage (CH) collections.

For this annotated glossary, each entry begins with the DH term and definition. It is *annotated* because term is then put into a galleries, libraries, archives and museums (GLAM) context, starting with potential applications, existing examples, and finally further resources for each method. Included at the end is an input and output type table, which sets out the nature of the sources that the DH methods use and/or produce.

These methods are defined separately but are often used together. For example *Heritage Connector*¹⁸ used named entity recognition (NER), a natural language processing method, with linked open data (LOD), to be able to link collections through a knowledge graph. The methods in the glossary are often one element in a method, this is alluded to in the 'see also' section.

3D scanning / 3D imaging

Also see AR & VR

3D scanning and imaging methods encompass several processes, including but not limited to, photogrammetry, computerised tomography (CT) scanning, structured light scanning, and laser (LIDAR) scanning.¹⁹ Photogrammetry is the process of digitally photographing and producing 3D objects that can

¹⁸ Heritage Connector: Transforming text into data to extract meaning and make connections
<<https://www.sciencemuseumgroup.org.uk/project/heritage-connector/>> [accessed 3 August 2022]

¹⁹ Urban, Richard, 'Collections Cubed: Into the Third Dimension | MW2016: Museums and the Web 2016' (presented at the MW2016: Museums and the Web 2016, Los Angeles, CA, USA, 2016)
<<https://mw2016.museumsandtheweb.com/paper/collections-cubed-into-the-third-dimension/index.html>> [accessed 28 June 2022]

have the same shape, colours and textures of the original object. To create the photographs are taken, then a software (such as [Agisoft Metashape](#)) is used to create a point cloud, subsequently a mesh of the cloud is made, on to this photographic textures can be applied. [Sketchfab](#) is currently a well-known platform on which these 3D models can be hosted and shared. While there are some limitations on the size and complexity of Sketchfab models, this is likely to improve rapidly in the years ahead.

GLAM and cultural heritage potentials

3D imaging of collection objects creates 3D representations of collection artefacts and can aid in their research. Representations of museum collection object in, for example, photogrammetry allows a move beyond 2D digital image representations of collection objects, where the subjectivity of the photographer may have meant that something of interest to the user could have been missed. It has the potential to allow for the representation at scale and texture in combination with other technologies such as augmented reality.

GLAM and cultural heritage examples

- [Portrait bust of James Watt](#) in SMG collection (more [here](#))
- [Rekrei](#) – digital preservation through crowdsourcing images of lost heritage

Further resources

- Cattel-Killick, Charlie, 'New Work: Photogrammetry for the New Medicine Galleries at the Science Museum!', *Museum in a Box*, 2020 <<https://museuminabox.org/new-work-photogrammetry-for-the-new-medicine-galleries-at-the-science-museum/>> [accessed 26 February 2022]
- Jeffs, Amy, Fern Insh, Erich Hatala Matthes, Tom Flynn, Daniel Pett, Tim Ayers, and others, 'Disciplining the Digital Virtual 3D Reproduction, Pilgrim Badges, and the Stuff of Art History', *British Art Studies*, 6, 2017 <<https://doi.org/10.17658/issn.2058-5462/issue-06/conversation>>
- Sunoikisis DC, *SunoikisisDC 2022: 3D Imaging in Museum Research*, 2022 <https://www.youtube.com/watch?v=YoR_ZPyS9sU> [accessed 23 March 2022]

Application programming interface (API)

See also Linked Open Data

APIs are a way of structuring data that makes it accessible and transmissible in a machine-readable way, allowing for communication between software programs. They act as intermediaries between

programs,²⁰ transferring raw data through their connection to a source server. How they are structured then allows this data to then be accessible through other apps and websites. An oft employed metaphor is that of the API as wait staff, bringing food between the kitchen and the customer, much like the API requesting data from source server and the app that it is being used in. APIs can be used as tool to access large volumes of data in defined ways.

GLAM and cultural heritage potentials

APIs have been used for some museum collections for about a decade.²¹ Now part of good open GLAM practice, they make data freely available with Creative Commons licensing allowing for exploration/use/research of collections data. Future DH GLAM projects will likely need APIs to work with large data sets and large data dumps, or indeed making sure that the most up to date curatorial data is used when representing collection objects. There are also specific API standards developed for cultural heritage content, such as the Internal Image Interoperability Framework (IIIF).

Note: SMG has a public API which contains the content from the collection website and is detailed at <https://www.sciencemuseumgroup.org.uk/about-us/collection/using-our-collection-api/>

GLAM and cultural heritage examples

- Open GLAM survey – API column
- Never been seen – uses SMG API

Further resources

- Culbertson, Joy, 'Top 10 APIs for Museums', *Programmable Web*, 2020
<<https://www.programmableweb.com/news/top-10-apis-museums/brief/2020/12/26>>
[accessed 2 March 2022]
- Wolff, Angela, 'V&A Launches New Developer API • V&A Blog', *V&A Blog*, 2021
<<https://www.vam.ac.uk/blog/digital/va-launches-new-developer-api>> [accessed 2 March 2022]

Augmented reality (AR)

Also see VR and 3D scanning / 3D imaging

²⁰ RapidAPI, 'Have Some Difficulties Understanding What an API Is? Take a Look at This RapidAPI Comic. 📖👉 { 1 / 6 } <https://t.co/10tIALNpbH>', @Rapid_API, 2022 <https://twitter.com/Rapid_API/status/1502283028676198403> [accessed 23 March 2022]

²¹ Dearnley, L., Reprogramming The Museum. In J. Trant and D. Bearman (eds). *Museums and the Web 2011: Proceedings*. Toronto: Archives & Museum Informatics. Published March 31, 2011. Consulted http://conference.archimuse.com/mw2011/papers/reprogramming_the_museum

Images that can be seen within augmented reality are the product of layers of information. The visual created by AR is one of hybrid immersion, a combination of real and fake. In some cases, AR needs to be experienced in a specific location.²² It can be accessed through a smart phone, or through specialist hardware like using Google Glass, Microsoft HoloLens or Magic Leap, with rumours of similar hardware in development from Apple and others.

GLAM and cultural heritage potentials

AR can showcase scientifically or historically accurate digital surrogates, these can be produced by the museum but interacted with outside of the physical site. As AR visuals are a combination of 3D object, layered on real background, this means it is much better at communicating scale of collection objects than de-contextualized collection photography. AR has preservation potential as it removes the need to have objects physically present. It could also be used to create accurate reproductions of lost objects.

GLAM and cultural heritage Examples

- Coates, Charlotte, 'How Museums Are Using Augmented Reality', MuseumNext, 2021
<<https://www.museumnext.com/article/how-museums-are-using-augmented-reality/>>
[accessed 25 March 2022]
- Work is in progress on an AR app from the Science Museum Group to launch in summer 2022.

Further resources

- Pedersen, Isabel, Nathan Gale, Pejman Mirza-Babaei, and Samantha Reid, 'More than Meets the Eye: The Benefits of Augmented Reality and Holographic Displays for Digital Cultural Heritage', *Journal on Computing and Cultural Heritage*, 10.2 (2017), 1–15
<<https://doi.org/10.1145/3051480>>
- Sunoikisis DC, *SunoikisisDC 2022: Archaeology, VR and Other Gaming Technologies*, 2022
<<https://www.youtube.com/watch?v=IL1GavATsXM>> [accessed 21 March 2022]
- Vince, 'Moving Beyond "Engagement": Using Sketchfab's 3D Models In the High School History Classroom', *Sketchfab Community Blog*, 2022
<<https://sketchfab.com/blogs/community/moving-beyond-engagement-using-sketchfabs-3d-models-in-the-high-school-history-classroom/>> [accessed 22 March 2022]

²² Sunoikisis DC, *SunoikisisDC 2022: Archaeology, VR and Other Gaming Technologies*, 2022
<<https://www.youtube.com/watch?v=IL1GavATsXM>> [accessed 21 March 2022]

Computer Vision

Also see OCR & HTR

A type of machine learning tool that finds patterns in visual sources, such as photographs, videos, documents, maps, newspapers and books. Currently computer vision focuses on two types of task: labelling of content within visual sources and pretrained models such as [Amazon Rekognition](#), [Microsoft Azure Cognitive Services](#), [TinEye Reverse Image Search](#) and [Google Vision](#) do this; and computer vision to search by image, or compare images, for example Google Reverse Image (accessible through browser interface – doesn't have to be scalable alternative).

GLAM and cultural heritage potentials

Potential applications include searches that augment from museum-controlled vocabularies and categorisations, through visual search and categorisation, and content tagging. There are also potential applications for keyword tagging of collections data collection data when cataloguing.²³

GLAM and cultural heritage examples

- [Computer vision for digital heritage](#) - Research group at The Alan Turing Institute.
- [Deep Discoveries final report](#) - computer vision (CV) and explainable artificial intelligence
- [What the Machine Saw](#) - SMG applying computer vision tags generated with Amazon Rekognition to collection.

Further resources

- Angelova, Lora, Bernard Ogden, Jack Craig, Hari Chandrapal, and Dipu Manandhar, Deep Discoveries: A Towards a National Collection Foundation Project Final Report, 18 November 2021 <<https://doi.org/10.5281/zenodo.5710412>>
- Bergel, Giles, Alexandra Franklin, Michael Heaney, Relja Arandjelovic, Andrew Zisserman, and Donata Funke, 'Content-Based Image-Recognition on Printed Broadside Ballads: The Bodleian Libraries' ImageMatch Tool', 17 <http://ifla-test.eprints-hosting.org/id/eprint/209/1/202-bergel-en.pdf>
- Rees, A.J. (2020) 'Image Recognition as a Tool in Cataloguing Born-Digital Photography'. In Uimonen, P. and Hartig, K. eds. Connect to Collect: Approaches to collection social digital

²³ Rees, A.J. (2020) 'Image Recognition as a Tool in Cataloguing Born-Digital Photography'. In Uimonen, P. and Hartig, K. eds. Connect to Collect: Approaches to collection social digital photography in museums and archives. Pg 221- 231 <http://www.diva-portal.org/smash/get/diva2:1429411/FULLTEXT01.pdf>. 230

photography in museums and archives. Pg 221–231 <http://www.diva-portal.org/smash/get/diva2:1429411/FULLTEXT01.pdf>

- Stack, John, 'Computer Vision and the Science Museum Group Collection', *Medium*, 2020
<https://lab.sciencemuseum.org.uk/computer-vision-and-the-science-museum-group-collection-a6c20efb0ac9> [accessed 26 February 2022]

Crowdsourcing

See also LOD

A sourcing model where collective labour of participants is drawn upon for a specific outcome. The 'crowd' can be made up of a collection of people from local community, experts / professionals, and those with commercial interests.¹ It can be hard to get people to participate, and difficult to keep them involved in projects. One application of crowdsourcing is for tasks are difficult to automate therefore need human input, another is in the training of automated artificial intelligence for example CAPTCHA. [Wikipedia](#) and [Wikimedia](#), example of large scale successful ongoing crowdsourcing project.

GLAM and cultural heritage potentials

Crowdsourcing as a method is not an opportunity for free labour. It can be used to address inequalities, or gaps in data.

GLAM and cultural heritage examples

- [Galaxy zoo](#) - citizen science project classifying galaxies
- [Old Weather](#) - 'Help scientists transcribe Arctic and worldwide weather observations recorded in ship's logs since the mid-19th century.'
- [Communities and crowds](#) – with Daily Herald Archive at the National Science and Media Museum

Further resources

- BrodeFrank, Jessica, Blickhan, Samantha and Rother, Becky. "Crowdsourcing Knowledge: Interactive Learning with Mapping Historic Skies." MW19: MW 2019. Published January 15, 2019. Consulted March 1, 2022. <https://mw19.mwconf.org/paper/crowdsourcing-knowledge-interactive-learning-with-mapping-historic-skies/>
- Ridge, Mia, Samantha Blickhan, Meghan Ferriter, Austin Mast, Ben Brumfield, Brendon Wilkins, and others, '10. Working with Crowdsourced Data', in *The Collective Wisdom Handbook: Perspectives on Crowdsourcing in Cultural Heritage - Community Review Version*, 2021

<<https://britishlibrary.pubpub.org/pub/working-with-crowdsourced-data/release/2>> [accessed 1 March 2022]

- Sunoikisis DC, *SunoikisisDC 2022: Crowdsourcing and Cultural Heritage*, 2022
<<https://www.youtube.com/watch?v=taedqUaJlqU>> [accessed 22 March 2022]

Data Visualization

See also knowledge graph, API and GIS

Data Visualizations are graphic displays that communicate data. They can be digital – but have a longer history – however, an affordance digital live sources in data visualization is that they can be interactive and update in real time. Data Visualizations are the product of a series of choices. They can be made to answer to a specific question, but also could be exploration of certain type of data. Data Visualizations come from a tradition of rendering quantitative information and aim to make coherent large data sets.²⁴

GLAM and cultural heritage potentials

Representation of uncertainty and ambiguity is a major topic to be addressed, especially in humanities and museum collections data visualization. Data infrastructures of digitised collections content means construction of data visualisations can be live and respond to the collections as they evolve.

GLAM and cultural heritage examples

- [Faces of Sweden](#) - Oliva Vane and The Nordic Museum
- [Heritage Connector Knowledge Graph Visualization](#) - just Science Museum Group collection represented in visualization

Further resources

- Boyd Davis, Stephen, Olivia Vane, and Florian Kräutli, 'Can I Believe What I See? Data Visualization and Trust in the Humanities', *Interdisciplinary Science Reviews*, 46.4 (2021), 522–46
<<https://doi.org/10.1080/03080188.2021.1872874>>
- Drucker, Joanna. 2011. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly* 5(1) <<http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>> [Accessed 27 July 2020]

²⁴ Tufte, Edward R., *The Visual Display of Quantitative Information*, 2nd edition (Cheshire, Conn: Graphics Press USA, 2001)

- Tufte, Edward R., *The Visual Display of Quantitative Information*, 2nd edition (Cheshire, Conn: Graphics Press USA, 2001)

Emulation

Emulation is the process of imitating, or *emulating*, software or hardware to ensure continued use and access to digital objects. It can be in the form of hardware or software, creating an environment where one system can behave like another software system. It preserves the context of the original digital environment, the 'look and feel' and 'original behaviors' as well as 'informational context'.²⁵ Emulation is a digital preservation process.

GLAM and cultural heritage potentials

As museums, including the SMG, have collected and continue to collect born digital and hybrid objects emulation for research and preservation purposes becomes a pressing consideration. Emulation is one part of how to present and access born digital and hybrid objects in the future.

GLAM and cultural heritage examples

- Antonelli, Paola, 'MoMA | Video Games: 14 in the Collection, for Starters', 2012 <https://www.moma.org/explore/inside_out/2012/11/29/video-games-14-in-the-collection-for-starters/> [accessed 3 March 2022]
- Fino-Radin, Ben, 'A Predecessor of Today's Smartphones | Cooper Hewitt, Smithsonian Design Museum', 2019 <<https://www.cooperhewitt.org/2019/05/13/a-predecessor-of-todays-smartphones/>> [accessed 3 March 2022]

Further resources

- Arrigoni, Gabi, Albert Museum, Natalie Kane, Albert Museum, Stephen McConnachie, and Joel McKim, 'Preserving and Sharing Born-Digital and Hybrid Objects from and across the National Collection', 2022
- 'Preservation Action - Digital Preservation Handbook' <<https://www.dpconline.org/handbook/organisational-activities/preservation-action>> [accessed 3 March 2022]

²⁵ 'Preservation Action - Digital Preservation Handbook' <<https://www.dpconline.org/handbook/organisational-activities/preservation-action>> [accessed 3 March 2022]

Geographical Information System (GIS)

Also see crowdsourcing and data visualisation

GIS is a geospatial system to view and manipulate data. It is used to connect, analyze and visualize information within the context of a digital map. It uses discrete layers of information that are built up in the GIS. This information or data comes in two main types: Raster data and Vector data. Raster data is good for representing continuous spatial information, like pictures it can come in low and high resolution. Vector data is mathematical geometric data. The different formatting of these separate types of data is one of the reasons why GIS employs layers.²⁶ The layers of information build up to represent context and research findings. Examples of GIS software include ArcGIS (widely used but paid for platform) and QGIS (which is free).

GLAM and cultural heritage potentials

GIS can be employed as a geographical way of representing historical data. GIS can bring together different sources and show what are the silences in the dataset. Good way of understanding “silences” in data sets by visualizing them on maps.²⁷ It allows for the exploration and interrogation of sites that cannot be explored physically.

GLAM and cultural heritage examples

- GlobalXplorer – platform to facilitate crowdsourced analysis of satellite images of earth for archeologists.
- The Battle of Hong Kong 1941 - A Spatial History Project
- The Moseley Homestead - Digital Heritage Preservation Project

Further resources

- ‘A Gentle Introduction to GIS — QGIS Documentation Documentation’
<https://docs.qgis.org/3.16/en/docs/gentle_gis_introduction/> [accessed 22 March 2022]
- Sunoikisis DC, Sunoikisis DC Fall 2018 Session 04. Querying Geo-Data, 2018
<<https://www.youtube.com/watch?v=JcD7RoyRkOU>> [accessed 22 March 2022]

²⁶ Sunoikisis DC, *SunoikisisDC 2022: Geographic Information Systems, 2022*
<<https://www.youtube.com/watch?v=FjonOEnrLWQ>> [accessed 21 March 2022]

²⁷ Murrieta-Flores, Patricia, Christopher Donaldson, and Ian Gregory, ‘GIS and Literary History: Advancing Digital Humanities Research through the Spatial Analysis of Historical Travel Writing and Topographical Literature’, *Digital Humanities Quarterly*, 011.1 (2016)

- Sunoikisis DC, *SunoikisisDC 2022: Geographic Information Systems, 2022*
<<https://www.youtube.com/watch?v=FjonOEnrLWQ>> [accessed 21 March 2022]
- Weiss, Claire, and San Francisco, 'Determining Function of Pompeian Sidewalk Features through GIS Analysis', 10 https://proceedings.caaconference.org/files/2009/43_Weiss_CAA2009.pdf

Handwritten text recognition (HTR)

See also Computer Vision and OCR

HTR employs image analysis to transform images of handwritten text to machine-readable text. It can interpret words, letters and numbers that are handwritten in historical documents, in a similar process to OCR turning images of typeset text into machine-readable text. HTR is a difficult process because handwriting can be specific to an individual, in addition to historical periods having their own specific handwriting styles. As a result, there can be errors both at the word and character level and outputs are at a lower quality than OCR.

GLAM and cultural heritage potentials

HTR systems (such as Transkribus) learn as they go, and therefore improve when the software has more information to be trained on. Although they may not be totally accurate on historical documents, the more documents and different handwriting styles that HTR software like Transkribus is trained on the better it will be and the less errors will occur when it digitises and transcribes texts.

GLAM and cultural heritage Examples

- [Transkribus](#)

Further resources

- Humbel, M., and J. Nyhan, 'The Application of HTR to Early-Modern Museum Collections: A Case Study of Sir Hans Sloane's Miscellanies Catalogue', In: *Proceedings of the Digital Humanities Conference 2019. Alliance of Digital Humanities Organizations: Utrecht, Netherlands. (2019)* (presented at the Digital Humanities conference 2019, Utrecht, Netherlands: Alliance of Digital Humanities Organizations, 2019) <<https://dev.clariah.nl/files/dh2019/boa/0802.html>> [accessed 9 March 2022]
- Müller, Markus, 'Uncovering censorship in the 16th century with Transkribus and Python. Episode I: OCR with Latin prints', *Digital Humanities Lab* <<https://dhlhlab.hypotheses.org/2022>> [accessed 9 March 2022]

Knowledge Graph

See also LOD

Knowledge Graphs are a way of organizing information through linked data. They combine the data management paradigms of database, graph and knowledge base.²⁸ Knowledge Graphs store and make searchable things and concepts based on how entities (that represent these things and concepts) relate to each other and are contextualized within the graph.²⁹ Google Search is a well-known Knowledge Graph.

GLAM and cultural heritage potentials

Knowledge Graphs can improve search and visualization of information. They are a possible alternative for organizing collections data: a step on from the relational databases that museum collection management systems are currently built on.³⁰ They have the potential to connect museum databases with other sources of information, to give further context and surface new connections to the collection, this was the case with SMG's Heritage Connector.

GLAM & Cultural Heritage examples

- Heritage Connector - suite of Demonstrators
- New York Public Library Archives and Manuscripts knowledge graph interface

Further resources

- Dutia, Kalyan, and John Stack, 'Heritage Connector: A Machine Learning Framework for Building Linked Open Data from Museum Collections', *Applied AI Letters*, 2.2 (2021), e23
<<https://doi.org/10.1002/ail2.23>>
- Singhal, Amit, 'Introducing the Knowledge Graph: Things, Not Strings', *Google*, 2012
<<https://blog.google/products/search/introducing-knowledge-graph-things-not/>> [accessed 17 February 2022]

Linked Open Data (LOD)

See also knowledge graph

²⁸ 'What Is a Knowledge Graph?', Ontotext <<https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/>> [accessed 23 March 2022]

²⁹ Singhal, Amit, 'Introducing the Knowledge Graph: Things, Not Strings', *Google*, 2012
<<https://blog.google/products/search/introducing-knowledge-graph-things-not/>> [accessed 17 February 2022]

³⁰ Such as Mimsy and Museum Index +

LOD is data that is published on the internet in a linked and open way. It aims to make data structured and machine readable so that it is easily navigated when shared online. It should use a standard format and refer to entities the same way that other LOD datasets do allowing interlinking. This is achieved by using uniform resource identifiers (URI) that represents entities. Linked open data should be made up of URIs within the Resource Description Framework (RDF), which is an abstract model to describe content semantically and link information. LOD should also use ontologies that use existing non-hierarchical vocabularies. LOD describes entities and relationships as triples: subject, predicate, object.³¹ LOD is not supported by most collection management systems (e.g. Mimsy XG and Adlib), so requires setting up and maintaining bespoke data infrastructure, frequently a knowledge graph.

GLAM and cultural heritage potentials

LOD has the potential to link multiple collections, objects, concepts, people and places that could be linked and then searchable.

GLAM and cultural heritage examples

- [A street near you](#)
- [Recogito](#)

Further resources

- Berners-Lee, Tim, 'Linked Data - Design Issues', 2009
<<https://www.w3.org/DesignIssues/LinkedData.html>> [accessed 26 February 2022]
- Blaney, Jonathan, 'Introduction to the Principles of Linked Open Data', Programming Historian, 2017 <<https://programminghistorian.org/en/lessons/intro-to-linked-data>> [accessed 26 February 2022]

Natural Language Processing (NLP)

Also see OCR, HTR, LOD

Natural language processing encompasses a collection of methods for computational processing of human language. It is an umbrella term inclusive of several machine learning methods that can be applied to human language so that they can be understood by computers. It is deployed as part of a pipeline of data processing rather than as a finished product. NLP can automate at scale, therefore can

³¹ Dutia, Kalyan, 'Knowledge Graphs, Machine Learning and Heritage Collections', Heritage Connector Blog, 2020 <<https://thesciencemuseum.github.io/heritageconnector/post/2020/11/06/knowledge-graphs-machine-learning-and-heritage-collections/>> [accessed 18 March 2022]

be useful on large data sets where finding meaning would be an overwhelming task for a single person. NLP is inclusive of processes such as sentiment analysis, speech recognition, translation, topic modelling, and named entity recognition (NER). NLP are the methods that can be employed once digitised and in plain text.

GLAM and cultural heritage potentials

Can be used to work with big text data sets quantitatively. There is potential for documents or text-based files that have already been digitised, to be analyzed at scale, this might include e identifying things (people, companies, places, events, etc.) to allow subsequent analysis or visualisation.

GLAM and cultural heritage examples

- American Museum of Natural History Case study in *The Museums and AI Network - AI: A Museum Planning Toolkit*
- NER in Heritage Connector

Further resources

- Block, Sharon, 'Doing More with Digitization', *Commonplace: The Journal of Early American Life* <<http://commonplace.online/article/doing-more-with-digitization/>> [accessed 17 March 2022]
- McGillivray, Barbara, Thierry Poibeau, and Pablo Ruiz Fabo, 'Digital Humanities and Natural Language Processing: Je t'aime... Moi Non Plus', *Digital Humanities Quarterly*, 014.2 (2020) <<http://www.digitalhumanities.org/dhq/vol/14/2/000454/000454.html>> [accessed 17 March 2022]
- Nguyen, Dong, Maria Liakata, Simon DeDeo, Jacob Eisenstein, David Mimno, Rebekah Tromble, and others, 'How We Do Things With Words: Analyzing Text as Social and Cultural Data', *Frontiers in Artificial Intelligence*, 3 (2020) <<https://www.frontiersin.org/article/10.3389/frai.2020.00062>> [accessed 17 March 2022]

Network analysis

Also see Knowledge Graph and DV

Network analysis is a useful methodological tool to understand relationships within data sets. It can visualise – and thus aid in the understanding of – different types of network actors e.g., people, things,

institutions and places, and the ways they are related to each other.³² These actors and their relations are represented with nodes and edges, which then make up the network that forms the analysis.³³ It is an effective way to map out complex connections and relationships,³⁴ and can be a useful research tool for humanists (for example historians) although not always appropriate.³⁵ Network visualisations can be generated with software including Onodo, Palladio, Gephi, Nodegoat and Node XL.³⁶

GLAM and cultural heritage potentials

Network analysis has the potential to communicate and interrogate complex networks to audiences. It is therefore a useful interpretive tool to reveal connections and relationships within a collection or exhibition. Rendering network analysis in digital form means that these can be interacted with and engaged with further by audiences.

GLAM and cultural heritage examples

- 'MoMA | Inventing Abstraction | Connections'
<<https://www.moma.org/interactives/exhibitions/2012/inventingabstraction/?page=connections>> [accessed 6 July 2022]
- Rutherford, Toni, 'Untangling the Web – Visualising the Networks Involved in the Construction of the Royal Albert Hall', *V&A Blog*, 2022 <<https://www.vam.ac.uk/blog/museum-life/untangling-the-web-visualising-the-networks-involved-in-the-construction-of-the-royal-albert-hall>> [accessed 6 July 2022]

Further resources

- Düring, Marten, 'From Hermeneutics to Data to Networks: Data Extraction and Network Visualization of Historical Sources', *Programming Historian*, 2015

³² Düring, Marten, 'From Hermeneutics to Data to Networks: Data Extraction and Network Visualization of Historical Sources', *Programming Historian*, 2015 <<https://programminghistorian.org/en/lessons/creating-network-diagrams-from-historical-sources>> [accessed 6 July 2022] & Rutherford, Toni, 'Untangling the Web – Visualising the Networks Involved in the Construction of the Royal Albert Hall', *V&A Blog*, 2022 <<https://www.vam.ac.uk/blog/museum-life/untangling-the-web-visualising-the-networks-involved-in-the-construction-of-the-royal-albert-hall>> [accessed 6 July 2022]

³³ Düring, Marten, 'Cheat Sheet: Social Network Analysis for Humanists', Digital Humanities LAB at CVCE Powered by Uni.Lu <<https://cvcedhlab.hypotheses.org/106>> [accessed 2 August 2022]

³⁴ Düring, Marten, 'From Hermeneutics to Data to Networks: Data Extraction and Network Visualization of Historical Sources', *Programming Historian*, 2015 <<https://programminghistorian.org/en/lessons/creating-network-diagrams-from-historical-sources>> [accessed 6 July 2022]

³⁵ Weingart, Scott, 'Demystifying Networks', *The Scottbot Irregular*, 2011 <<http://www.scottbot.net/HIAL/?p=6279>> [accessed 6 July 2022]

³⁶ Although Node XL requires some python knowledge.

<https://programminghistorian.org/en/lessons/creating-network-diagrams-from-historical-sources> [accessed 6 July 2022]

- Weingart, Scott, 'Demystifying Networks', *The Scottbot Irregular*, 2011
<http://www.scottbot.net/HIAL/?p=6279> [accessed 6 July 2022]

Optical character recognition (OCR)

Also see HTR, NLP and computer vision

OCR is a type of computer vision that transforms an image of print text into machine encoded text. Documents and other textual sources are scanned and using computer vision and image recognition technologies converted to text. These files need to be 'cleaned' as even best OCRs have errors.³⁷ OCR can be used to convert and interpret sources at large scale, e.g. the [Google Books initiative](#). Accuracy of results has significantly improved in recent years due to the use of machine learning to refine results.

GLAM and cultural heritage potentials

The method is established and once complete operations like keyword searches, categorization, NLP and then NER can be applied to text from the documents.³⁸

GLAM and cultural heritage examples

- cmaquarie, 'Experiments with Digital Tools in the Archives — OCR', *Brought to Light*, 2018
<https://blogs.library.ucsf.edu/broughttolight/2018/06/05/experiments-with-digital-tools-in-the-archives-ocr/> [accessed 9 March 2022]
- Bean, Iain, 'OCRguessr: Making Posters into a Guessing Game — Science Museum Hangman | by John Stack | Science Museum Group Digital Lab', *Medium*, 2021
<https://lab.sciencemuseum.org.uk/ocrguessr-making-posters-into-a-guessing-game-science-museum-hangman-8220e79adbed> [accessed 9 March 2022]

Further resources

- Akhlaghi, Andrew, 'OCR and Machine Translation', *Programming Historian*, 2021
<https://programminghistorian.org/en/lessons/OCR-and-Machine-Translation> [accessed 9 March 2022]

³⁷ Crump, Jon, 'Generating an Ordered Data Set from an OCR Text File', *Programming Historian*, 2014
<https://programminghistorian.org/en/lessons/generating-an-ordered-data-set-from-an-OCR-text-file> [accessed 9 March 2022]

³⁸ Gupta, Maya R., Nathaniel P. Jacobson, and Eric K. Garcia, 'OCR Binarization and Image Pre-Processing for Searching Historical Documents', *Pattern Recognition Society*, 2015
<https://doi.org/doi:10.1016/j.patcog.2006.04.043>

- Crump, Jon, 'Generating an Ordered Data Set from an OCR Text File', *Programming Historian*, 2014 <<https://programminghistorian.org/en/lessons/generating-an-ordered-data-set-from-an-ocr-text-file>> [accessed 9 March 2022]
- 'Optical Character Recognition (OCR) - How It Works' <<https://www.nicomsoft.com/optical-character-recognition-ocr-how-it-works/>> [accessed 9 March 2022]

Persistent identifiers (PIDs)

See also LOD

Persistent identifiers are a unique and universal identifiers that do not change over time or across systems. PIDs are part of data and research infrastructure. They are automatically generated, rather than manually created, through a process called "minting".³⁹ PIDs are IDs given to digital objects that are stable and long lived, and mean that digital object they refer to is citable and traceable. PIDs should do three key things: they need to be actionable, they need to be globally unique, and they need to be persistent for long term use.⁴⁰ Actionable means that by adding some additional information to the PID string can make it into a live URL (web address).

GLAM and cultural heritage potentials

PIDs enable better understanding of how collections are being used because they enable and support tracking of objects and subjects online. For example, DOIs (digital object identifiers) make scholarly research outputs interoperable and measurable.⁴¹

GLAM & Cultural Heritage examples

- [Towards a National Collection – Heritage PIDs](#)
- [ORCID ID – unique persistent identifiers for researchers](#)

Further resources

- Towards A National Collection, *TaNC: Persistent Identifiers (Rachael Kotarski) and Locating a National Collection (Gethin Rees)*, 2021 <https://www.youtube.com/watch?v=jLbK4E_ebMM> [accessed 2 March 2022]

³⁹ Koster, Lukas, 'Persistent Identifiers for Heritage Objects', *The Code4Lib Journal*, 47, 2020 <<https://journal.code4lib.org/articles/14978>> [accessed 2 March 2022]

⁴⁰ Towards A National Collection, *TaNC: Persistent Identifiers (Rachael Kotarski) & Locating a National Collection (Gethin Rees)*, 2021 <https://www.youtube.com/watch?v=jLbK4E_ebMM> [accessed 2 March 2022]

⁴¹ Koster, Lukas, 'Persistent Identifiers for Heritage Objects', *The Code4Lib Journal*, 47, 2020 <<https://journal.code4lib.org/articles/14978>> [accessed 2 March 2022]

- Koster, Lukas, 'Persistent Identifiers for Heritage Objects', *The Code4Lib Journal*, 47, 2020 <<https://journal.code4lib.org/articles/14978>> [accessed 2 March 2022]
- Tóth-Czifra, Erzsébet, 'Developing Identifiers for Heritage Collections: Q&A with Rebecca Grant (F1000) and Frances Madden (British Library), Authors of the Open Educational Resource', *DARIAH Open* <<https://dariahopen.hypotheses.org/1210>> [accessed 3 March 2022]

Virtual reality (VR)

Also see 3D scanning / 3D imaging and AR

VR is a visual, virtual and spatial technology that creates immersive fully generated 3D environments.⁴² It is a visual simulation that people understand as real. It is possible to build an artificial world within VR, to move around and interact with. VR is primarily visual but can also involve audio or even smell. VR technology needs specialist hardware with headsets such as [Google Cardboard](#), [HTC VIVE](#), and [Oculus Quest 2](#).

GLAM and cultural heritage potentials

Collaborations between 3D specialists and researchers means VR has the potential to be used as a scholarly digital artefact.⁴³ VR can be employed as a method to recreate in 3D space artefacts and places not accessible physically.

GLAM & Cultural Heritage examples

- [Dare to Discover Virtual Reality experience](#) - The National Maritime Museum, Amsterdam in the Netherlands
- [Google Arts and Culture VR tours](#)
- [Tim Peake Space Decent VR](#) – Science Museum

Further resources

- Kuchelmeister, Volker, 'The Virtual (Reality) Museum of Immersive Experiences', 2018 <<https://doi.org/10.14236/ewic/EVA2018.39>>
- Ma, Xiaocheng, 'VR and Cultural Heritage Recreation', *AMT Lab @ CMU*, 2021 <<https://amt-lab.org/blog/2021/10/vr-and-cultural-heritage-recreation>> [accessed 19 March 2022]

⁴² Sunoikisis DC, SunoikisisDC 2022: Geographic Information Systems, 2022 <<https://www.youtube.com/watch?v=FjonOEnrLWQ>> [accessed 21 March 2022]

⁴³ Ma, Xiaocheng, 'VR and Cultural Heritage Recreation', *AMT Lab @ CMU*, 2021 <<https://amt-lab.org/blog/2021/10/vr-and-cultural-heritage-recreation>> [accessed 19 March 2022]

- Sunoikisis DC, SunoikisisDC 2022: Geographic Information Systems, 2022
 <<https://www.youtube.com/watch?v=FjonOEnrLWQ>> [accessed 21 March 2022]

Input/output type table

This table sets out the various DH methods for digitising and working with digitised cultural heritage collections, and the type(s) of source they use or produce. If a DH method uses or produces types of sources (catalogue data and metadata, 2D images, text, 3D visuals) this is indicated below, if a requires an additional pre-processing step this is indicated and the DH method initials are included as a ‘via’ step.

	Catalogue data & metadata	2D Images	Text	3D visuals
Application programming interface (API)	X	X	X	
Augmented reality (AR)	via DV			X
Computer Vision (CV)		X		
Crowdsourcing	X	X	X	
Data Visualization (DV)	X	via CV	via NLP	
Emulation*				
Geographical Information System (GIS)	X		via NLP	
Handwritten Text Recognition (HTR)		X	X	
Knowledge Graph (KG)	X	via CV, OCR, HTR and NLP	via NLP	
Linked Open Data (LOD)	X	via OCR, HTR and NLP	via NLP	
Natural Language Processing (NLP)	X	via OCR and HTR	X	
Network analysis	X and via KG			
Optical character recognition (OCR)		X		
Persistent identifiers (PIDs)	X	X	X	X
3D scanning / 3D imaging				X
Virtual reality (VR)	via DV			X

* Emulation primarily works with software as a content source.

Appendix II – Existing SMG DH projects

Listed here are DH projects, or projects that employ the same digitisation processes as DH methodologies, that have been produced or are in production at SMG.

3D Scanning / 3D imaging

- 3D Photogrammetry bust of James Watt – VR and 3D printed outcome - 2010
 - Hess, Dr Mona, 'James Watt - a VR Encounter with the Engineer', *Engineering-the-Future* <<https://lguariento.github.io/Engineering-the-Future/27.html>> [accessed 30 June 2022]
 - Hess, Mona, and Stuart Robson, 'Re-Engineering Watt: A Case Study and Best Practice Recommendations for 3D Colour Laser Scans and 3D Printing in Museum Artefact Documentation', *D Printing*, 9
 - 'Portrait Bust of James Watt | Science Museum Group Collection' <<https://collection.sciencemuseumgroup.org.uk/objects/co8231699/portrait-bust-of-james-watt-portrait-bust-cast>> [accessed 30 June 2022]
- SMG collection items on Sketchfab, ongoing
 - 'Science Museum Group (@sciencemuseum)', *Sketchfab* <<https://sketchfab.com/sciencemuseum>> [accessed 30 June 2022]
- 3D scan Shipping Gallery. Science Museum, London, 2012
 - Hindmarch, J., Melissa Terras, and S. Robinson, 'On Virtual Auras: The Cultural Heritage Object in the Age of 3D Digital Reproduction', in *The Routledge International Handbook of New Digital Practices in Galleries, Libraries, Archives, Museums and Heritage Sites*. (London: Routledge, 2019), pp. 243–56 <<https://melissaterras.org/2019/11/19/new-book-chapter-on-virtual-auras-the-cultural-heritage-object-in-the-age-of-3d-digital-reproduction/>> [accessed 21 January 2020]
 - 'Shipping Galleries | Science Museum' <<https://scanlabprojects.co.uk/work/shipping-galleries/>> [accessed 2 August 2022]
 - 'Shipping Galleries | Science Museum', *ScanLAB Projects* <<https://scanlabprojects.co.uk/work/shipping-galleries/>> [accessed 15 August 2022]

Application programming interface (API)

- Circuits of Practice, 2022

Natale, Simone, Ross Parry, and Petrina Foti, *Circuits of Practice Research Report: Narrating Histories of Computing and Digital Media in Museum Environments* (Loughborough University, 25 February 2022)

<https://repository.lboro.ac.uk/articles/report/Circuits_of_Practice_research_report_narrating_histories_of_computing_and_digital_media_in_museum_environments/19317764/1> [accessed 30 June 2022]

- Never Been Seen, 2021
 - ‘Never Been Seen | Science Museum Group Collection’
<<https://thesciencemuseum.github.io/never-been-seen/>> [accessed 30 June 2022]
 - Stanley, Will, ‘Never Been Seen’, Science Museum Group, 2021
<<https://www.sciencemuseumgroup.org.uk/blog/never-been-seen>> [accessed 4 July 2022]
- SMG collections API, 2017 - present
 - ‘Using Our Collection API’, *Science Museum Group*
<<https://www.sciencemuseumgroup.org.uk/about-us/collection/using-our-collection-api/>> [accessed 30 June 2022]
 - Stack, John, ‘Exploring Museum Collections Online: Some Background Reading’, Medium, 2018 <<https://lab.sciencemuseum.org.uk/exploring-museum-collections-online-some-background-reading-da5a332fa2f8>> [accessed 4 July 2022]

Augmented reality (AR)

- Archive to AR Immersive storytelling, National Railway Museum, 2019-2020
IT, Hive IT UX Design and Development Sheffield | Hive, ‘Immersive Rail’, Hive IT UX Design and Development Sheffield | Hive IT (Hive IT UX Design and Development Sheffield | Hive IT),
<https://www.hiveit.co.uk> <<https://hiveit.co.uk/our-work/immersive-rail/>> [accessed 1 August 2022]
- Journeys of Invention app, 2013
‘Journeys of Invention’ (London Science Museum: The Red Green & Blue Co Ltd, 2013)
<<https://apps.apple.com/gb/app/journeys-of-invention/id685965924>> [accessed 30 June 2022]
- Wonderlab AR, 2022
‘Wonderlab AR on the App Store’ (Science Museum) <<https://apps.apple.com/gb/app/wonderlab-ar/id1635122389>> [accessed 3 October 2022]

Computer Vision (CV)

- Colour & Shape, 2020
Sleeman, Cath, 'Colour & Shape: Using Computer Vision to Explore the Science Museum Group Collection', *Medium*, 2020 <<https://lab.sciencemuseum.org.uk/colour-shape-using-computer-vision-to-explore-the-science-museum-c4b4f1cbd72c>> [accessed 30 June 2022]
- What the Machine Saw, 2019
 - 'What the Machine Saw' <<https://johnstack.github.io/what-the-machine-saw/>> [accessed 30 June 2022]
 - Stack, John, 'Computer Vision and the Science Museum Group Collection', *Medium*, 2020 <<https://lab.sciencemuseum.org.uk/computer-vision-and-the-science-museum-group-collection-a6c20efb0ac9>> [accessed 26 February 2022]

Crowdsourcing

- Communities and Crowds, ongoing
 - 'Communities and Crowds', *Science Museum Group* <<https://www.sciencemuseumgroup.org.uk/project/communities-and-crowds/>> [accessed 30 June 2022]
 - 'Opening up the Daily Herald Archive to Citizen-Led Research', *National Science and Media Museum Blog* <<https://blog.scienceandmediamuseum.org.uk/opening-up-the-daily-herald-archive-to-citizen-led-research/>> [accessed 30 June 2022]
- Railway Work, Life & Death, 2016 – current
'Railway Work, Life & Death', *Railway Work, Life & Death* <<https://www.railwayaccidents.port.ac.uk/>> [accessed 30 June 2022]

Data Visualization (DV)

- Colour & Shape, 2020
Sleeman, Cath, 'Colour & Shape: Using Computer Vision to Explore the Science Museum Group Collection', *Medium*, 2020 <<https://lab.sciencemuseum.org.uk/colour-shape-using-computer-vision-to-explore-the-science-museum-c4b4f1cbd72c>> [accessed 30 June 2022]

Emulation

- Manchester Baby, Science and Industry Museum, 2018
'Why Emulate the Manchester Baby? | Museum of Science and Industry', *Science and Industry Museum Blog* <<https://blog.scienceandindustrymuseum.org.uk/why-emulate-baby/>> [accessed 30 June 2022]

Geographical Information System (GIS)

- Locating Communications Heritage, 2009-2010
University College London, 'Locating Communications Heritage: Engaging the Mobile User | GTR' <<https://gtr.ukri.org/projects?ref=AH%2FH018247%2F1>> [accessed 30 June 2022]
- The East Coast Time Line, National Railway Museum, 2012
'The National Railway Museum's First Ever Smartphone App', *National Railway Museum Blog*, 2012 <<https://blog.railwaymuseum.org.uk/the-national-railway-museums-first-ever-smartphone-app/>> [accessed 30 June 2022]

Knowledge Graph (KG)

- Heritage Connector, 2020-22
 - Dutia, Kalyan, and John Stack, 'Heritage Connector: A Machine Learning Framework for Building Linked Open Data from Museum Collections', *Applied AI Letters*, 2.2 (2021), e23 <<https://doi.org/10.1002/ail2.23>>
 - 'Heritage Connector Blog', *Heritage Connector Blog* <<https://thesciencemuseum.github.io/heritageconnector/>> [accessed 30 June 2022]
 - 'Heritage Connector Demonstrators' <<http://heritageconnector.sciencemuseum.org.uk/>> [accessed 30 June 2022]
 - 'Heritage Connector Knowledge Graph Visualisation' <https://thesciencemuseum.github.io/heritage-connector-demos/3_visualisation/index.html> [accessed 30 June 2022]
 - Winters, Jane, John Stack, Kalyan Dutia, Jamie Unwin, Rhiannon Lewis, Richard Palmer, and others, *Heritage Connector: A Towards a National Collection Foundation Project Final Report* (Zenodo, 9 February 2022) <<https://doi.org/10.5281/zenodo.6022678>>

Linked Open Data (LOD)

- Tristan, Roddis, *Linked Data at Science Museum*, 2013
<https://www.w3.org/2013/04/odw/odw13_submission_62.pdf>

Natural Language Processing (NLP)

- Congruence Engine, 2022-ongoing
 - 'The Congruence Engine: Digital Tools for New Collections-Based Industrial Histories', *Science Museum Group* <<https://www.sciencemuseumgroup.org.uk/project/the-congruence-engine/>> [accessed 4 October 2022]
 - Boon, Tim, and Anna-Maria Sichani, *The Congruence Engine: Digital Tools for New Collection-Based Industrial Histories* (Zenodo, 4 October 2022)
<<https://doi.org/10.5281/zenodo.7144075>>
- Heritage Connector, 2020-22
(See above)

Network analysis

- Congruence Engine, 2022-ongoing
(See above)

Optical character recognition (OCR)

- OCRguessr, 2021
Bean, Iain, 'OCRguessr: Making Posters into a Guessing Game — Science Museum Hangman | by John Stack | Science Museum Group Digital Lab', *Medium*, 2021
<<https://lab.sciencemuseum.org.uk/ocrguessr-making-posters-into-a-guessing-game-science-museum-hangman-8220e79adbed>> [accessed 9 March 2022]
- OCR for Making Modern World online, 2004
OCR of documentation so that it could be edited and form part of Making the Modern World online, to accompany the Making the Modern World permanent galleries.

Persistent identifier (PIDs)

- Persistent Identifiers, 2020-2022

'TANC HeritagePIDs - Index' <<https://tanc-ahrc.github.io/HeritagePIDs/>> [accessed 30 June 2022]

Virtual reality (VR)

- ExploreStation VR experience with National Rail, National Railway Museum, 2022
'ExploreStation VR Experience', *National Railway Museum*
<<https://www.railwaymuseum.org.uk/whats-on/explorestation-vr/>> [accessed 4 July 2022]
- Handley Page VR, 2017
'Handley Page VR: Science Museum's VR Experience', *PRELOADED*
<<https://preloaded.com/work/science-museum-handley-page-vr/>> [accessed 15 August 2022]
- Space Descent
'Space Descent VR', *Science Museum Group* <<https://www.sciencemuseumgroup.org.uk/our-services/partner-with-us/touring-exhibitions/space-descent-vr/>> [accessed 4 July 2022]
- Thresholds, National Science and Media Museum, 2018
 - 'Thresholds', *National Science and Media Museum*
<<https://www.scienceandmediamuseum.org.uk/what-was-on/thresholds>> [accessed 30 June 2022]
 - Wentworth-Weedon, Jack, and Sharan Dhanda, 'How "Thresholds" Comes to Life', *National Science and Media Museum Blog*, 2018 <<https://blog.scienceandmediamuseum.org.uk/how-thresholds-comes-to-life/>> [accessed 30 June 2022]

Appendix III - Ideas for DH projects in SMG

These are a list of potential project ideas for DH methods at SMG: firstly, SMG collection objects that could be digitised, enabling further research, or already digitised objects/parts of the collection that could be researched further through application of DH methods; and secondly, ways of working that could enable DH methods to develop. These were suggested as part of the workshops and interviews undertaken for primary research for this landscape report.

- **Punched card digitisation**

There are numerous punch cards in the SMG collection,⁴⁴ that would have been used to program machines such as Jacquard looms or early computers.⁴⁵ Suggestion to digitise these punched cards in order to access and research the binary code they represent.

- **Acquisition register digitisation**

HTR could be used to transcribe the Science Museum acquisition registers, currently only in their original physical leather-bound ledger form, for SMG collections provenance research.⁴⁶

- **Shared technical standards for parts**

Using 3D scanning and imaging methods to test for standardization of mechanical parts, i.e. screw threads, in the Mechanical Engineering collection. Data from this would enable provenance research, and potentially network analysis, through understanding any standardization of mechanical parts across collection objects.⁴⁷

- **XRF analysis**

XRF, or X-ray Fluorescence,⁴⁸ analysis enables detailed understanding of the composition of materials. Suggested use on SMG collection objects, and collection of this material composition information as part of cataloguing processes.⁴⁹

- **Material network analysis**

As research and interpretation method for understanding relationships between collection objects, through the movement and usage of raw materials.⁵⁰

- **Planned emulation at point of collection**

⁴⁴ 'Science Museum Group Collection | "punched Cards" Search'

<<https://collection.sciencemuseumgroup.org.uk/search?q=punched%20cards>> [accessed 6 July 2022]

⁴⁵ Science Museum Group. Collection of early punched cards used by Hollerith and BTM. 1984-544Science Museum Group Collection Online. Accessed July 6, 2022.

<https://collection.sciencemuseumgroup.org.uk/objects/co62511/collection-of-early-punched-cards-used-by-hollerith-and-btm>. | Science Museum Group. Jacquard Punched Cards, 1890-1910.. 2002-432Science Museum Group Collection Online. Accessed July 6, 2022.

<https://collection.sciencemuseumgroup.org.uk/objects/co503962/jacquard-punched-cards-1890-1910-punched-cards>. | Suggested Dr Rachel Boon

⁴⁶ Suggested Ben Russell. Dr Rachel Boon, Ben Russell, Jessica Bradford, and John Stack, Workshop: Digital Humanities methods for SMG collection, 2022

⁴⁷ Suggested Ben Russell. Dr Rachel Boon, Ben Russell, Jessica Bradford, and John Stack, Workshop: Digital Humanities methods for SMG collection, 2022

⁴⁸ Beckhoff, Burkhard, Birgit Kanngießner, Norbert Langhoff, Reiner Wedell, and Helmut Wolff, Handbook of Practical X-Ray Fluorescence Analysis (Springer Science & Business Media, 2007)

⁴⁹ Suggestion Ben Russell. Dr Rachel Boon, Ben Russell, Jessica Bradford, and John Stack, Workshop: Digital Humanities methods for SMG collection, 2022

⁵⁰ Suggested by Dr Rachel Boon, from Ben Russell point. Dr Rachel Boon, Ben Russell, Jessica Bradford, and John Stack, Workshop: Digital Humanities methods for SMG collection, 2022

When collecting artifacts that are born digital or have a born digital element, SMG could also collect accompanying software for emulation purposes. In that way, artefact and digital context are collected for later research and understanding.⁵¹

- **Innovation time**

Many open-source digital preservation (DP) tools built in spare time of DP professionals working in cultural heritage. What could be achieved if dedicated time set aside to experiment?⁵² Proposed half a day each week for dedicated play and experimentation time, with different digital tools and processes.⁵³

⁵¹ Suggested by Ed Bartholemew. Jack Kirby, Geoffrey Belknap, Ed Bartholomew, and Dr Tilly Blyth, Workshop: Digital Humanities methods for SMG collection, 2022

⁵² 'Tools Grid - COPTR', Community Owned Digital Preservation Tool Registry (COPTR)
<https://coptr.digipres.org/index.php/Tools_Grid> [accessed 6 July 2022]

⁵³ Suggested by Somaya Langly. Somaya Langley, Digital Humanities research project - interview, 2022