HERITAGE AND DATA: CHALLENGES AND OPPORTUNITIES FOR THE HERITAGE SECTOR

Report of the Heritage Data Research Workshop held Friday 23 June 2017 at the British Library, London

Heritage Research
Heritage Futures
Alan Turing Institute
British Library

Report compiled by Rodney Harrison, Hana Morel, Maja Maricevic & Sefryn Penrose

Content

The event held at the British Library on the 23rd June 2017 was envisaged as an initial scoping and investigative research workshop, bringing together key representatives from the UK heritage industry and academic community from humanities and social and computing science to discuss challenges and opportunities that data presents to the Heritage Sector.
The workshop was organised as a collaborative event between the AHRC Heritage Priority Area, the AHRC-funded Heritage Futures research programme, the Alan Turing Institute and the British Library with an intention to create an interdisciplinary space for discussion of the role of data in heritage research, bringing together practitioners with members of the academic community to discuss these issues.
Key objectives were to:

- Identify key research question that are arising as heritage industry embraces data;
- Capture research interests and capability, including similarities and differences, across the sector that would have significant impact on the sector development;
- Develop a broader understanding of key issues across the sector;
- Establish next steps to address the issues identified at the workshop.

The cultural and natural heritage sector holds rapidly increasing volumes of data largely on human society and culture, past and present, which has become a new frontier of digital operations for many institutions.

Many heritage organisations have invested substantially in digitising and cataloguing analogue sources and are now gathering born-digital content at scale including:

- Electronic personal archives and digital information in a variety of formats;
- Data related to historic buildings and environment;
- Data related to entire collections or certain parts of collections, geographic and provenance data, archaeological data, ecological and biodiversity data;
- Data related to specific communities, or audience related data, etc.

This data provides:

- Foundation for new research into both historical (natural and cultural) phenomena and contemporary life;
- A basis for a growing range of new services to different audiences;
- New information source for industry, government and general public;
- New ways to facilitate organisational and broader stakeholders' planning processes for variety of infrastructure projects;
- Transformation of business processes in the heritage institutions; and
- Growing understanding of audiences' interests, behaviours and characteristics.

As with other sectors, data is presenting both new opportunities and new challenges. In many instances, organisations are now at the stage of working out how to deal with often heterogeneous and multi-format data that might be fragmented even at the institutional level and certainly across the UK digital cultural space. There is very little collaborative work that has been done to date to systemically move forward the sector's knowledge of how to deal with cultural and natural heritage data in such a way as it can be linked, analysed, processed and understood. There is also a need to consider underlying standards, ethical issues and sustainability in the way that is appropriate for heritage.

Ahead of the workshop, we developed an initial list of challenges to stimulate discussion. These are presented below:

- Digital heritage: it is unclear what the current status in relation to digital heritage is, what is being collected, why and in which formats; or how organisations and their users intend to make use of this material.
- Digitisation and data development processes that lead to data creation in heritage are not completely understood. However, decisions made at this stage could influence outcomes related to this data at a later stage.
- Governance issues relating to privacy, ethics, provenance and other key considerations.
- Metadata and standards are underdeveloped and fragmented and there is no sufficient understanding about sustainable ways to approach this across the sector.
- Preservation of this data is happening in continuously changing environment. As this is one of the key remits for many of the organisations involved, the effective and robust methods that can future-proof preservation policies, practice, and technologies is an important aspect to consider.
- Opening heritage data democratises access, but what technical and ethical issues are implied in this process, and how might these be managed?
- Multimodal nature of data challenges in heritage and culture.
- Automation and machine learning opportunities and challenges arising in the sector.
- Potential of data research to improve understanding of audiences, targeting of programmes and activities, and improving social inclusion. There is a vast difference in levels of investment and engagement with these issues amongst different domains or fields of practice within sector. This variability also arises from the limits on resources which are felt particularly acutely by small to medium and independent heritage organisations. The project aims to identify ways of supporting not only the bigger national organisations, but also the needs of small to medium and independent organisations across the sector.

The list was intended to encourage discussion, with an expectation that these themes will be changed in the course of the workshop. The aim was to stimulate a sector-wide discussion, working with participants to validate and define key issues arising in their professional practice and research.

This report captures key points from the presentations given during the workshop, as well as bringing together the common themes identified by participants and key points from the extensive range of discussions. The report also highlights areas which were seen as potential future research by participants during the Collaborative Dialogue sessions. The workshop participants were also keen to agree next steps for potential future activity, which is also captured in this document.

References are used to attribute points raised to particular guest speakers, although in some cases this is not possible.

10.00-10.05	Welcome. *Maja Maricevic*, Head of Higher Education, British Library
10.05-10:30	Keynote: Data challenges and opportunities, wider perspective. *Professor Patrick Wolfe*, Exec Director, UCL Big Data Institute & non-Executive Director and Trustee, Alan Turing Institute
10.30-10:50	Introduction and setting the Workshop goals. *Rodney Harrison*, UCL
11.00-11.20	Data and Heritage - Case Study: National Archives *Sonia Ranade*, Head of Digital Archiving
11.20-11.40	Data and Heritage - Case Study: British Library *Adam Farquhar,* Head of Digital Scholarship
11.50-12.10	Data and Heritage – Case Study: Heritage Lottery Fund *Gareth Maeer,* Head of Research
12.10-12.30	Data and Heritage – Case Study: Historic England *Jen Heathcote*, Head of Strategic Research and Partnerships *Keith May*, Heritage Information Strategy Adviser
12.30-12.50	Data and Heritage – Case Study: British Museum *Dominic Oldman*, Senior Curator Ancient Egypt & Sudan and Head of ResearchSpace
13.40-14.10	Session 1: Collaborative Dialogue, Facilitator: *Maja Maricevic*, British Library Potential research questions and their significance for heritage organisations and the sector
14.10-15.00	Session 2: Collaborative Dialogue, Facilitator: *Sefryn Penrose*, UCL Barriers and opportunities including policy and rights, technology, collaboration, access to data and existing research strengths and weaknesses
15.15-16.30	Facilitators' Feedback from Sessions 1 & 2 Emerging concepts and next steps synthesis. *Rodney Harrison*, UCL

8 Attendees

The Heritage Data Workshop held on the 23rd June 2017 brought together a group of 38 practitioners from various cultural institutions based in the UK listed below.

Arends, Bergit	Science Museum
Bell, Nancy	The National Archives/National Heritage Science Forum
Bonacchi, Chiara	UCL, Institute of Archaeology
Connolly, Edmund	British Library, BSO Higher Education
Dappert, Angela	British Library, Thor Project Manager
Denard, Hugh	King's College London, Assistant Professor Digital Arts and Humanities
Dommett, Tom	National Trust
Farquhar, Adam	British Library, Head of Digital Scholarship
Fitzgerald, Neil	British Library, Head of Digital Research
Goudarouli, Eirini	National Archives, Digital and Technology Research Lead
Green, Laura	Kew Science
Harrison, Rodney	UCL, AHRC Heritage Priority Area Leadership Fellow
Hauswedell, Terras	UCL, Centre for Digital Humanities
Heathcote, Jen	Historic England, Head of Strategic Research & Partnership Team
Jeffrey, Stuart	School of Simulation and Visualisation, Glasgow School of Art, Fellow
Lane, Alison	National Trust
Leeson, Adala	Historic England, Head of Social and Economic Research and Insight
Madsen, Christine	Oxford e-Research Centre
Maeer, Gareth	Heritage Lottery Fund, Head of Research
Maricevic, Maja	British Library, Head of Higher Education
May, Keith	Historic England, Heritage Information Strategy Advisor
Mcapra, Alastair	Heritage Science Forum, Chairman
McConnachie, Stephen	British Film Institute
Mia, Ridge	Museums Computer Group, Chair
Morel, Hana	UCL, AHRC Heritage Priority Area Research Associate
O'Donnell, Joe	The Heritage Alliance, Policy and Communications Officer
Oldman, Dominic	British Museum, Head of ResearchSpace
Padfield, Joe	National Gallery
Prescott, Andrew	University of Glasgow, AHRC Digital Transformations Leadership Fellow
Ranade, Sonia	National Archives, Head of Digital Archiving

McGregor, Sam	Alan Turing Institute, Senior Research Facilitator
Penrose, Sefryn	UCL, Institute of Archaeology/Heritage Futures research programme
Sexton, Anna	National Archives, Head of Research
Smith, Robin	National Library of Scotland, Head of Collections and Research
Ward, Marcus	Historic England
Weech, Marie-Helene	Kew Gardens
Wolfe, Patrick	UCL, Professor of Statistics, Exec Director, UCL Big Data Institute & Alan Turing Institute, non-Executive Director and Trustee
Worthington, Richard	Historic England, Head of Digital Marketing and Communications

Key Points Raised:

• The future use of data and the way in which it develops into heritage practice has a potential to have transformative effects on how we collect, curate and care for both natural and cultural heritage.

• There is an inherent trade-off between technology and data potentially improving core infrastructure in society as well as standards of living, versus its provocation of deep social concerns. There is a constant need to recognise the public's position and stance in the debate around data creation, usage and management, particularly as data is now collected in huge volumes without explicit knowledge. Without the necessary safeguards and regulations in place to ease the mind of the public, many opportunities as a result of data use are most likely lost.

• Shaping our own agenda through experience and evidence is a key element to developing decisions on data governance for the heritage sector. Organisations are themselves beginning to steer towards more effective management, but would benefit from greater recognition of this new area of responsibility and further sharing of best practice and collaborative working.

• Data introduces new levels of uncertainty and bias and creates unprecedented levels of profusion that requires expert judgement deployed alongside algorithms, statistical methods and machine learning.

• Although many heritage institutions have collections and data that have been actively and consciously 'given' to them, today much of the data generated and created is 'captured', as a by-product of some form of digital interaction. This requires new types of expertise and development of new processes outside of traditional views of collecting, but this also offers new opportunities in engaging audiences and creating new knowledge.

• Case studies reveal that the public have so far not been able to engage with data collections, and so organisations have taken a more human-centred understanding of how the public use the collections and what they want from it. More work is needed in involving public with these new aspects of heritage, including linking these new activities to educational opportunities and raising public awareness of digital changes in society.

• We have entered an era in which traditional notions of accountability, agency, and permission have been overturned, but while the excitement of data's potential and how we make it publicly accessible may take precedence in some activities, these core public concerns will not disappear.

• Working with other institutions within the heritage sector, and engaging with the public and their relationship with the data, will ensure that data governance addresses societal needs and reflects the public interest.

• Data is subject to error and variation, and can be compromised in various ways. Complete transparency offers a way to curb risks and concerns associated with data e.g. data quality, data accuracy, and data fusion.

• Data can help in offering new ways to engage with questions of what heritage is, what it consists of, what can we tell about it in statistical terms, as well as about its diversity and dynamics.

• A networked heritage, as in linking organisations to each other and sharing information across a system at local and national levels, may act as a catalyst which empowers local change, leadership and action, as well as enriches local debate and dialogue.

• With the rise of algometric decision-making and policy development, data science outputs themselves have become a matter of public record which needs to be preserved if we are to ensure future accountability. The question of how we preserve such outputs and hold code and algorithms to account is becoming increasingly important.

• There are significant and well-established critiques of the technocracy and datafication (e.g. Rico 2017) of heritage practices, especially in relation to their apparent democratizing effects on heritage and collections, which suggest that any developments in this area need to be assessed critically.

Data Landscape

Society and Technology

Science, technology and data are not only increasingly important to how we function as a society, they are also critical for the global economy (WEF, 2012; 2016). It is now almost impossible for companies and industries to remain globally competitive without the use of technology or automation, and without making huge investments in understanding behaviours and targeting audiences/customers through generated data. Not only do things get done quicker, but in many cases more efficiently. The McKinsey Global Institute (2011) noted that 'in a big data world, a competitor that fails to sufficiently develop its capabilities will be left behind'.

We know, however, that there is an inherent trade off. While technology and data may improve core infrastructure in society as well as standards of living, they equally provoke deep social concerns. These concerns are not new. Sharp rises in public alarm related to privacy, or the role of technology in destabilising employment opportunities are recurrent historically, as technology and data continues to impact upon society and the lives of individuals in novel ways.

There is a need to develop new approaches and frameworks for the governance and management of data, and new forms of technology and data can help achieve this. We must, however, always be mindful to new and unexpected uses, users and interests as we develop new approaches and frameworks for the management and usage of technology and data.

Data Science Landscape

The Workshop opened with Patrick Wolfe's[1] talk *The Data Science Landscape,* which set out key challenges in the wider data landscape. This enabled the rest of the workshop to proceed by placing heritage discussions into the broader context of contemporary developments in data science, to identify key areas where heritage can benefit from advances in data science and to inform our horizon-scanning for future planning.

Wolfe defined data science as an *'interdisciplinary field about scientific methods, processes and systems to extract knowledge or insights from data in various forms'.* He saw it as a field that often treats text, images and shapes, as well as data related to human subjects, thus making a strong

1 Professor of Statistics, UCL; Exec Director of UCL Big Data Institute and Alan Turing Institute

link with data formats most relevant in heritage. Making sense of large heterogeneous data is dependent on the advances of statistics, machine learning, systems, databases and applied maths, which creates potential for new research and strong alignment between heritage and data science.

We know that how data is now generated, collected and processed are leading to huge quantities of complex data, with some data collection intentionally generated while in other cases it is a by-product of pervasive and widespread use of digital technologies. Added to this is the growing difficulty – yet critical need – of ensuring quality of data and the uncertainty embedded in inconsistency of details generated.

What Wolfe raised as the first problem is that, effectively, the 'traditional data lifecycle' is no longer relevant. Traditionally - within the heritage sector, but equally relevant to all other sectors - we have collected data, processed data, and applied it to validate the hypothesis/problem which is directly linked to the methods consciously developed and used. We have then used those lessons and insights to inform further research. This gather-process-apply lifecycle has become increasingly complex with so much of today's common activities generating huge amounts of data, reaching beyond any linear structure or even awareness of what is being collected.

Wolfe highlighted that key to the vast amount of data collected is a myth that the more collected, the better one can understand the data. He asserted that making sense of large heterogeneous data is dependent on the knowledge of domain experts as well as advances in areas such as statistics, applied maths, databases, systems and machine learning.

Today's popular approach, the 'black box' approach, results in specific black boxes – or 'data lakes' – being fed with specific data-feeds using pre-set algorithms. These types of prediction approaches are adopted primarily by businesses but have little meaning in terms of research and understanding. However, having a mixed combination of literate data experts and research specialists, and contextualising the data with a clarity of vision, provides a different level of understanding data.

This level of greater understanding will be essential when we look at changes in policy and other legislative concerns, and it is this understanding rather than usage of the prediction paradigm that will enable us to see a shift in the data landscape. At this point in time, however, traditional governance, policy, and other interventions or societal steers for data collection, management and usage are simply no longer fit for purpose. Wolfe pointed out that there are many current unresolved regulatory issues related to data: in particularly the issues of privacy, ethics and

transparency are lagging behind practice. Public dialogue and policy are yet to fully address the issues of the ubiquitous data collection by private sector and governments.

Wolfe pointed to a range of emerging UK and International initiatives focusing on Data Governance including:

- Royal Society and British Academy project on data governance;
- Royal Society project on Machine Learning;
- IEEE project on Ethical Considerations in Artificial Intelligence and Autonomous Systems;
- Work done by the Information Commissioner's Office, Defra, GO Science, as well as European and US sources.

The UK is also investing significantly into relevant research via the Alan Turing Institute, the UK institute dedicated to data science, which is engaged across many different sectors and is looking to further understand key issues affecting heritage. But also, key capacity and investment is being made in key university groups and specific expertise of other national facilities such as the Francis Crick Institute, Farr Institute and the Leverhulme Centre for the Future of Intelligence. There are parallel developments in the US, Canada, Europe, China, India and many other countries.

Heritage Futures

In the next introductory session Rodney Harrison[2] introduced the AHRC Heritage Priority Area and the AHRC-funded Heritage Futures project.

The AHRC Heritage Priority Area is:

- Developing and leading on the intellectual agenda via its Future Heritage Research Strategy;
- Engaging with communities across disciplinary boundaries;
- Promoting collaboration within academia and beyond;
- Advising AHRC on needs and trends.

It was therefore important to link the dialogue about heritage data with the future of heritage research in order to inform this agenda, especially in terms of new needs and trends arising as data becomes a more prominent part of the heritage landscape.

Harrison also introduced the Heritage Futures[3] project. Heritage Futures is a 4-year research

2 Professor of Heritage Studies at UCL and the AHRC Heritage Priority Area Leadership Fellow
3 www.heritage-futures.org

programme (2015-2019) funded by the AHRC, and supported additionally by its host universities and partner organisations, such as the Heritage Alliance, IUCN, York Museum Trust, Kew Gardens, National Trust, the Frozen Ark and NordGen amongst many others. The project is carrying out ambitious interdisciplinary research to explore the potential for innovation and creative exchange across a broad range of heritage and related fields, in partnership with a number of academic and non-academic institutions and interest groups.

In his introduction Harrison emphasised the four key themes of the project:

Uncertainty: How is the uncertainty of the deep future conceived of and managed in different fields of conservation practice?

Transformation: What values are associated with heritage structures and landscapes that are allowed to undergo transformation and change?

Profusion: How do museums and people in their homes decide what to keep in the face of mass production and consumption?

Diversity: How are biological, cultural, genetic, and linguistic diversity categorised and conserved, and what can one field learn from another?

The above themes very closely match the key themes arising as a part of data discussions. We have already seen from the wider landscape themes introduced by Wolfe, that data introduces new levels of uncertainty and bias and that it creates unprecedented levels of profusion that requires expert judgement deployed alongside algorithms, statistical methods and machine learning. The future use of data and the way in which it develops into heritage practice has a potential to have transformative effects on how we collect, curate and care for both natural and cultural heritage. At the same time, we have already seen a variety of issues arising in practice, not least a danger of structuring emerging systems and practice in the ways that reinforces or even worsens the existing biases of biological, cultural, linguistic and genetic homogeneity.

Harrison in particular challenged the workshop participants to consider already apparent vast differences in levels of investment and engagement with data amongst different parts of the sector. This variability arises from the limits on research funding which are felt particularly acutely by small to medium and independent heritage organisations. He suggested that we need to find ways of supporting not only the bigger national organisations, but also the needs of small to medium and independent organisations across the sector.

Data In The UK Heritage Sector

While data collection, manipulation and prediction have been an established part of the heritage sector for some time, in the light of the big data revolution now taking place, we currently do not have an overview of existing practice within the sector. Equally, while some policy is derived from key national and international initiatives, we currently do not have heritage specific data policy frameworks. Just as observed by Wolfe, the practice is moving ahead of policy.

The sector does have a lively exchange and collaboration between digital practitioners working with data. There are also multiple collaborative initiatives and an active exchange with the academic community. The AHRC Digital Transformations theme[4] has, in particular, catalysed a lot of important research in this area. There is notable research and infrastructure presence of UK heritage organisations and the academic community within the relevant EU programmes as well. Many projects and institutions are enthusiastically embracing Open GLAM initiatives, using data initiatives to create new ways to engage the public with heritage data.

However, what has not been done to date is an attempt to understand and define if there is a common theme of Heritage Data. We still do not have a sector wide understanding of any common challenges and any systemic interventions that might be needed. Maja Maricevic[5] highlighted that the heritage sector, like many other sectors, faces challenges in relation to infrastructure, investment, agreed data standards and a dialogue with public and key stakeholders about data implications for heritage. Maricevic also pointed out that there is a 'shared sense of heritage data community', despite parts of the sector sometimes seeing data in different ways. Her key question to the participants was whether there is a way in which further dialogue could lead to a more formal understanding of common aims and challenges - in the way that, for example, a very diverse Healthcare sector discusses and understands a common challenge of utilising data opportunities in improving health outcomes. Can our work with data transform the way the public engages with and perceives heritage?

Examples of Institutional Practice

The workshop format did not allow us to present the whole breadth of institutional practices related to data, nor to ensure an even representation of different types of organisations working

4 http://www.ahrc.ac.uk/research/fundedthemesandprogrammes/themes/digitaltransformations/
5 Head of Higher Education, British Library

with data in heritage. However, it was deemed important to look at some current examples of work undertaken in the institutions that are active in this area before attempting to discuss key challenges and opportunities.

The challenging timelines in organising the workshop inevitably meant that we captured examples and organisational practice in the institutions where there is an already developed data framework in the organisation. Even then, this was the first time that data was discussed in its own right by the institutions dealing with different aspects of heritage data - such as archives, images, text, buildings, land, objects and the natural environment. The organisers deemed that it was valuable to bring together views and experiences from these different heritage domains in order to start exploring key common issues and opportunities.

This practical solution reinforces the challenge voiced by Harrison about differing levels of investment and engagement in relation to data in different parts of the sector. This should be considered as a significant limitation of this discussion which should be addressed in any future discussions.

Presentations were invited from the National Archives, the British Library, Heritage Lottery Fund, Historic England and the British Museum. The presenters were asked to present on either any significant data projects or their organisational approach to data. It was hoped that this approach would enable us to engage with both existing data projects and institutional ways of addressing this emerging field through organisational policies. We also wanted to gauge the level of technical accomplishment and challenges that already exist in the sector.

The five presentations were inevitably very different in their approach, but they all showed that data developments were modelled to closely follow existing institutional missions and priorities. It was clear that data practice in these institutions has already matured beyond experimentation towards attempts to articulate and deploy data in advancing key strategic outcomes.

Sonia Ranade[6] in her presentation *From paper to code: Data Science at The National Archives* framed her discussion within the National Archives' ambition 'to become a digital archive by instinct and design'. This articulated the key challenges across all key areas of the institutional activities in collecting, preservation and access.

6 Head of Digital Archiving, The National Archives

The National Archives has set an ambitious digital strategy to preserve a wider range of records and to enable better, user-friendly access to these records for readers and data-users.

In relation to collecting, this means grappling with 'unstructured heaps of digital information' as large as 1PB. The traditional institutional functions of identifying what is of value and what is sensitive is therefore experiencing a profound challenge. Although many heritage institutions have collections and data that were actively and consciously 'given' to them, today much of data generated and created is 'captured', as in it is a by-product of some form of interaction. Ranade explained that while before it was skilled people who read data and knew what to keep or dispose of, now it is simply impossible for people to make these decisions and carry out these processes individually. This echoed Wolfe's point that more data always requires more skilled people to understand it.

Another important theme for the National Archives was preservation and accountability, including new issues for public records. With the rise of algometric decision-making and policy development, data science outputs themselves have become a matter of public record which needs to be preserved if we are to ensure future accountability. The question of how we preserve such outputs and hold code and algorithms to account is becoming increasingly important.

This adds to the overall complexity and 'long-tail' of file formats that are emerging and that require new preservation solutions. This data is distinctly different to the current archival catalogues. Ranade illustrated this point using the project *Traces Through Time*[7], which uses the latest technology to link and suggest records that may relate to the person being researched. The new techniques enable reconnecting personal data using probabilistic matching of records. The project is showing future promise and will benefit from more research that will enable use of algorithms and machine-learning to reveal new connections and untold stories.

What we see emerging is the move to act immediately to make short term advances, Ranade said, but that there is now a need to invest and partner for more robust mechanisms.

7 http://www.nationalarchives.gov.uk/about/our-role/plans-policies-performance-and-projects/our-projects/traces-through-time/

In his presentation *Data and the British Library*, Adam Farquhar[8] echoed many themes introduced by Ranade. Just as the National Archives case, at the British Library, data is being embedded in the organisational strategy with the vision 'that data are as integrated into our collections, research and services as text is today'. The goal is to enable the British Library's users to consume data online though tools that enable it to be analysed, visualised and understood by non-specialists. This has led to identification of four core themes for the strategy: data management, data creation, data archiving and preservation and data discovery, access and re-use.

Farquhar explained the genesis of the strategy linked to the British Library's legislative remit to implement the UK 'non-print' legal deposit, which effectively means capturing and preserving the UK digital domain including e-books and e-journals, newspapers, sheet music, maps, web and other digital objects 'published' in the UK, generating many petabytes of new data in every format and every subject. Looking just at the UK Web Archive, which is downloading and preserving the UK web domain, this means a growth of around 70TB per year, generated via automated crawls. However, this new data frequently comes with numerous legal restrictions in relation to its use.

Another important source of data is the British Library's extensive digitisation of historic collections, such as digitisation of historic newspapers that currently contains just under 20 million newspaper pages, or the *Two Centuries of Indian Print*[9], a project currently digitising 1,000 rare Bengali printed books and 3,000 early printed books.

Farquhar explained how a dedicated team of Digital Curators is supporting computationally driven research using the British Library's data, most notably within the BL Labs project[10]. This project, supported by the Mellon Foundation, is just embarking on its third phase. BL Labs encourages researchers, developers, educators and artists to develop research and other projects using the British Library's digital content. To date some of the initiatives include a release of 1 million images on Flickr, which led to a range of projects from artistic reuse to experimentation with machine learning using neural networks for automated tagging of images. Neural networks are a set of algorithms, modelled loosely after the human brain, that are designed to recognise patterns. Many of these projects have been exploring difficult and underdeveloped areas, including constant work on optical character recognition (OCR) and algorithms that need much development to develop their use for difficult digital content such as images, sound or historic

8 Head of Digital Scholarship, British Library
9 https://www.bl.uk/projects/two-centuries-of-indian-print
10 http://labs.bl.uk/

scripts such as in Bengali books. In its third phase BL Labs are aiming to integrate the processes and tools that have developed to date into the British Library's regular service offer to users.

Farquhar also pointed out some key challenges from the British Library's perspective, including the continuing effort needed to transform physical into digital, transforming digital collections into usable information and data, engaging researchers and enabling greater discovery and use of data.

Our third presentation from Gareth Maeer[11] introduced us to the *Heritage Index*[12], a joint project from the Heritage Lottery Fund and RSA. The Heritage Index uses data to enable exploration of local heritage. It uses 120 different indicators, which are combined and mapped to provide data for all 390 local authority areas in the UK. It combines data for heritage assets such as buildings and nature reserves, and data for heritage activities such as volunteering, investment and community initiatives.

Maeer moved our data discussion into an entirely new domain, helping us to see data as a new way that can begin to help us engage with the questions of what heritage is, what it consists of, what can we tell about it in statistical terms, as well as about its diversity and dynamics.

The Heritage Index can be accessed through the RSA website which offers an exploratory map-based search as well as the full data download. It is organised under six domains: historic built environment; museum, archives and artefacts; industrial heritage; parks and open spaces; landscape and natural heritage, and; culture and memories. The Index was developed by using and combining many existing data sets rather than collecting new data. It is designed to be open and inclusive in its definition of heritage.

Maeer emphasised the dynamic nature of this data resource, its main use and value being in opening questions rather than providing answers. The Index has a variety of functions – to promote tourism, to help local authorities understand local heritage and coordinate activity, and for use by local activists and the public. Maeer illustrated some of the uses through the example of Warrington, where a low Index score energised local stakeholders into action and enabled them to plan the way forward.

11 Head of Research, Heritage Lottery Fund
12 https://www.thersa.org/action-and-research/rsa-projects/public-services-and-communities-folder/heritage-and-place

The project presented us with another case where value judgements are playing a significant part in developing and using data resources, not least in ensuring fairness and flexibility needed to use data to catalyse change by enriching local debate, helping communities learn, and making better-informed decisions on the local and national level. Maeer also introduced the idea of networked heritage, which is a concept expressing potential to link organisations to each other and to share information across the system at local and national levels in the way that empowers local change, leadership and action.

On the national level DCMS have used Heritage Index as an indicator in a White Paper, so it will be interesting whether this continues with a third version of the Index expected in 2018.

The presentation for Historic England covered an extensive heritage domain spanning objects and artefacts, archaeological data, monuments and buildings. Jen Heathcote[13] introduced us to a set of organisational priorities related to data. As the Government's advisory body looking after the historic environment, Historic England has to fulfil statutory functions and provides constructive advice to owners, guardians and the public on managing change to it. As well as protecting heritage, Historic England carries out research to help people understand and care for heritage and understand its value to society. Data plays a significant role in enabling Historic England to fulfil these roles. Heathcote introduced it using the following broad categories:

Research

- Examining how we can improve use of spatial analysis to identify risk and opportunity;
- Improving understanding of digital archiving and dissemination.

Engagement

- Examining how we can show the value of heritage & counter assumptions that it is a barrier to growth;
- Exploring transfer and sharing of sensitive heritage data (e.g. Listed Buildings Owners Survey);
- Improving social inclusion by better understanding how heritage impacts society.

Comms and Marketing

- Improving Historic England's ability to share data;
- Identifying what metadata – context, taxonomies, structure – makes it more accessible to others.

13 https://www.thersa.org/action-and-research/rsa-projects/public-services-and-communities-folder/heritage-and-place

Further detail regarding data work at Historic England was provided by Keith May[14], who reflected on two key areas of activity – digital archiving and digital data structures. Digital archiving strategies described by May moved discussions forward by introducing other essential parts of heritage data infrastructure, such as the Archaeological Data Service (ADS) which is strongly embedded within the sector. Since 1999, Historic England requires data from the projects they fund be deposited with the ADS. Also, Historic England undertakes other projects related to digital archiving, such as Big Data project in 2004 which focused on GIS/Lidar/Laser Scan/Sonar data.

Historic England's current focus was more on archiving methods and metadata rather than analytics, and it follows the typical lifecycle for archaeological data. Similarly, their work on digital data structures was focused on establishing ontologies and vocabularies that allow for more appropriate descriptions of the heritage domain. This work has a strong archaeology focus and includes events such as the creation of objects, their loss, deposition, discovery, interpretation and conservation. The ontology model used is the CIDOC Conceptual Reference Model (CRM), which provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation.

Historic England puts a great deal of attention in contributing and developing data infrastructure that is interoperable across the sector. May pointed out extensive collaborative work related to the Linked Open Data (LOD) vocabularies[15] which enables interoperability and data exchange underpinning thesauruses supporting activities such as the classification of monument type records, archaeological objects, construction materials, sampling and processing methods and materials and periods, and time-based entities.

May emphasised the importance of persistent URIs enabling exciting new developments such as inclusion of multilingual labels.

Project such as STAR, STELLAR and SENESCHAL have been essential in moving forward the development of semantic technologies for archaeological resources. However, May also referred to continuing challenges such as tensions between being descriptive vs. controlled at the point of data entry.

14 Heritage Information Strategy Adviser, Historic England
15 http://heritagedata.org

Heathcote and May asked some key research questions informed by the Historic England's research agenda:

- What sort of information will be most useful to those who plan and take decisions affecting the historic environment?
- How can we encourage the sharing, linking and interoperability of historic environment data and information?
- How do we ensure the consistent development, application and enforcement of existing technical information and data standards and promote this to others?
- What is the best approach to providing services for the public and research professionals to access and use information?
- How can we harness the enthusiasm of the general public and other groups to help improve the heritage record?

The discussion about semantics and categorisation continued in the presentation given by Dominic Oldman[16] from the British Museum - *Issues of Cultural Heritage Data Quality*. Oldman started with the challenge that our current catalogues and systems should not be replicated for data. Current heritage information systems such as SPECTRUM and MARC are designed for a closed environment which depends on classical categorisation models demanding that things that share common properties belong in the same category. Similarly, the British Museum's Collection Online search could be understood only by people working in museums, and even more specifically archaeology museums, due to specialist classifications and terminology.

He highlighted the dangers of publishing raw data that is designed to work in conjunction with curator's advice, this bringing us back to Wolfe's introductory theme of requirement for expert interpretation of data. The British Museum's *ResearchSpace*[17] project is attempting to address this problem, concentrating on content and data, utilising Linked Data and Semantic Web techniques as well as a range of other digital research methods like natural language processing. Changing emphasis leads to a new paradigm which reduces focus on programming code, meaning that software becomes a revealer of original semantics, and not a system providing semantics. The ontology used is CIDOC CRM, previously also introduced by Historic England.

ResearchSpace enables the British Museum to return its experts to the centre of the system, thus returning their knowledge back into the narratives previously dominated and distorted by

16 Head of ResearchSpace, British Museum
17 http://www.researchspace.org

technology requirements. The new search allows searching against objects, people, places, events, periods and concepts - also providing context for each other. This allows the user to start with a particular entity and then derive other entities from the results - people can lead on to places, and places to events and then to objects and so on. Oldman illustrated this through a demonstration of Hokusai woodblock digital mapping, embedding expert knowledge relevant to this subject in a flexible way. *ResearchSpace* also provides a user-friendly environment that enables more intuitive discovery, visualisation and new and more meaningful interaction between users and content, including enabling users to create their own datasets. It provides an integrated environment for contextual data, as well as a variety of research tools such as semantic search, semantic annotation, image manipulation and annotation with IIF etc.

Collaborative Dialogue

Some key challenges were raised here, outlined below:

- Archives are also data which opens up new types of research and questions about access;
- There is a need to link people and places by enabling navigation within and between collections;
- We need to create a user-friendly service for data;
- There is a growing difficulty to keep up with different filing systems and file formats.

Further discussion during the collaborative dialogue at the meeting pointed to complications of data within the heritage sector in terms of the contextualisation of data: non-sensitive data can actually hold sensitive insights through linkages with other datasets or analysis with new techniques. As transparency and open access movements take prevalence to counter any concerns of usage and encourage transparency and access, anonymisation is becoming an increasing challenge due to the landscape becoming more interlinked and working as part of an open network.

Some themes that emerged from the session are highlighted below.

Building Trust

The Workshop highlighted the constant need to recognise the public's position and stance in the debate around data creation, usage and management. Nowadays, data is sometimes created in huge volumes without explicit intent and design as a part of digital operational processes and everyday interactions. In some cases, this leads to new opportunities to link previously limited data sources with other datasets revealing more than perhaps agreed or consented to.

It could be argued that actual genuine consent is simply unfeasible if not impossible. However, there remain some critical questions as to the inherent trade-offs (or debates) between building trust and providing safeguards for the public resulting in slow incremental improvements due to society anxiety, or prioritising the pursuit of knowledge and process at all costs.

Indeed, without the necessary safeguards and regulations in place to ease the mind of the public, many opportunities as a result of data use are most likely lost (Session 1).

Concerns and Anxiety

In the Collaborative Session, perceptions of misuse and anxiety about misuse were discussed at length. It was also highlighted that history has demonstrated many examples by which the implementation of new technologies has led to public concerns or even controversies, overshadowing the benefits certain technologies offer.

Today, for example, there are huge barriers to data sharing despite the potential to do so, due to concerns relating to data protection and usage. Public confidence in data governance is crucial, but tensions between recent proposals for using data (such as headlines suggesting a breach of data, *'Hospital data sold without patients' consent to boost profits of private drugs companies'* (Daily Mail, 2014) or *'Patients records should not have been sold, NHS admits'* (Telegraph 2014)) have led to new questions about data governance and controversial issues about data application. There have also of course been on-going, and as yet inconclusive, global debates related to the appropriateness use of data by governments, covering issues from mass surveillance to the emerging use of private data in political processes, which have also had an impact on the public perception of Big Data.

Privacy

Privacy issues are incredibly complex not least because there is no definite concept of what it is, and any attempt to define it reveals competing and contradictory dimensions. For example, just because something is in the public sphere, it does not necessarily mean it is not simultaneously private. While privacy is central in its association with core concepts such as liberty, democracy and freedom, it also involves issues of power, control and covert or overt surveillance. These complexities are exacerbated by the ways in which policy-makers' approaches to privacy generally tend to address and solve concerns for more specific (and less holistic) issues (Solove, 2008).

Additionally, 'at a foundational level, different cultures and groups share different notions of privacy, setting boundaries about what is considered private or not. Compounding the differences between groups, attitudes to privacy are highly context-specific and tied closely

to the purpose for which data is used' (British Academy & Royal Society, 2017: 30; Wolfe). For example, workshop participants touched on the idea of commercial opportunities and whether turning data into an income generator was a viable option. The reality, however, is that for now, the public may well have more support for public data being available and free for public use and within the public sphere, but have huge concerns bordering privacy breaching when their data is used for commercial and profitable applications (Collaborative Session).

Ownership and Responsibility

This leads on to questions of ownership. During the workshop there was a clear recognition that data ownership and protection are linked to the recognition of its value i.e. practitioners recognise the value of the information in their possession and protected by their organisation, and thus recognise their responsibility towards it. Another point raised was that communities which provide data feel organisations need to have more responsibility in terms of making this data meaningful, and ensuring it has some sort of longevity or is properly archived.

The notion of ownership, particularly in the heritage sector, clearly constitutes a key challenge for future research and policy making.

Fairness and Transparency

Although many heritage institutions have collections and data that have been actively and consciously 'given' to them, today much of the data generated and created is 'captured', as a by-product of some form of digital interaction (Ranade). Wolfe introduced fairness and transparency, saying that 'we do not get a great feeling of transparency right now which has stopped the UK and US government from using strategies' for data collection. He added that capturing data 'needs to be justified to society'- as in, why are we as government, institutions and organisations automatically collecting data from the public's daily lives in a sort of surveillance society? He also mentioned that consent becomes 'hazy with ubiquitous data collection' and that there is a level of discomfort as big technology companies or governments acquire personal data. He asked the question, 'Do we wish to give our data away for perpetuity?'.

The workshop also focused on transparency and fairness in terms of access and ability to use data now being made increasingly available to the public. In this sense, it was seen in terms of the uncertainty of data. Case studies, as in the case of the British Museum, revealed that the public

have so far not been able to engage with data collections, and thus the BM have taken a more human-centred understanding of how the public use the collections and what they want from it (Oldman).
There was also a clear recognition that work and analysis of data are linked to a wider culture of delay in publication, and that there should be a change in striving for perfection by instead revealing uncertainty through being more transparent about analysis (Collaborative Dialogue).

Fairness, equally, was seen from a different approach, perhaps more akin to fair representation and exclusion within datasets. Ranade raised the issue that, for example, records of women were harder to validate and triangulate through data fusion due to the fact that women were more likely to change their names, suggesting that using maiden or marital names or next of kin would be something that could help produce a fairer result.

Ethical Considerations & Engagement with the Public

While the Heritage Data Workshop's preliminary targets were to address some of the arising issues the sector faces by the voluminous flux of available data, the public remain central to any further progress in big data, analysis and application. We have entered an era in which traditional notions of accountability, agency, and permission appear to be changing, but while the excitement of data's potential and how we make it publicly accessible may take precedence in some activities, these core public concerns will not disappear. Particularly within the heritage sector, uncertainties that arise from all areas of practice have great potential to require social and ethical consideration. Working with other institutions within the heritage sector, and engaging with the public and their relationship with the data, means that it will be important that data governance addresses societal needs and reflects the public interest.

Key questions raised at the workshop include:

- Should we push the public's comfort zone just because technology has progressed farther? (Wolfe)
- How do we approach diversity (of data) if it is only as diverse as the categories we use to understand diversity? (Harrison)
- How can we use data to build up knowledge, and how can government and the heritage sector look into requirements needed for open access and linked data? (Maricevic)
- How do you identify what is of value and what is sensitive? (Ranade)

- How do you balance the release of imperfect research with the potential to be used by government? (Maeer)
- How can the sector encourage access, use and sharing of data and information within the sector and for the wider public? (Heathcoate & May)
- In light of data's positioning with discourse on civil rights and counter-cultural views of technology development, we should keep in mind that we are being questioned quite a lot in today's climate (Oldman).
- How do you enable the vision of data to become meaningful to everyone?
- Where is the line between making something available for the public, freely available, or freely available with the need of expensive software?
- Are we tracking our data and finding out how it is used for the right reasons?
- How do we best express the limitations of data served to different audiences?
- How do we justify the workload of dealing with big data, or the need for expertise?

These were just some of the thoughts raised, which all point to the need for further facilitation and engagement of debate to help shape future practice and policy.

We need to move beyond traditional data models and protections. The workshop highlighted the nature of data as linked to technical, social, and legal developments but also demonstrated the importance of transparency as the speed of data-centrality in our society surges social anxiety. Access and reuse of data in ways which surpass the original purpose of collection provides space for imagination and innovation, but the notion of profiling and feeling of a loss of control causes unease.

Issues of Heritage Data Quality

One of the key points raised in the workshop was uncertainty in the context of diverse sources producing diverse, or even incomplete or unreliable, data leading to questions surrounding the integrity of the data or simply its compromised ability to reveal trends/patterns/results.

Data is subject to error and variation, and can be compromised in various ways. Data about data (as in, metadata) can be transferred and transformed over time with errors, mistakes, gaps and so on, compromising the information and increasing complications of verifiability (Ranade; Oldman).

The British Academy & Royal Society (2017: 22) points out:

> Knowing in advance which data sets are of poor quality or misrepresentative is far from simple. This was arguably a simpler calculation to make when the logic of data collection and its use were more tightly coupled. As data streams are purposed and repurposed, the reliable and useful signals of provenance become more important yet harder to achieve. Furthermore, as data is used to train algorithms and insights from data become embodied in algorithms that are traded, knowing where data comes from is likely to become significantly more difficult.

With this in mind, the need to address the lack of skill, expertise and resources was raised numerous times in the workshop. With much experience fragmented across institutions and the sector itself, the ability to rigorously justify data creation and the work that goes into it can be difficult.

Additionally is the need for experts across the heritage sector - and beyond – to feed back their own experiences with data so that unidentified questions and unresolved strategies can be explored through dialogue.

The data conversation is intricately linked with communities, but also tied to government, the public sector, the private sector and academia. Shaping our own agenda through experience and evidence is a key element to developing decisions on data governance for the heritage sector. Participants in the workshop pointed towards success stories promoting the sharing of scientific data within a 'trusted environment' (see E-RIHS "DIGILAB").

Changing Our Capacity to Deal with Data

Data Governance: Existing Data and Changing Management

During the Collaborative Dialogue, 'how much there is, what it means, and how terrifying it is for a small poorly resourced team to make a meaningful entry in that world of data' was pointed out by co-organiser Sefryn Penrose[18]. What is data? Big data has a 'big horizon'. There are inordinate definitions as to what data is, and great heterogeneity within each collection. Not only does this make it technically difficult to handle but it promotes heavy scrutiny. Penrose added that 'all eyes are on data' and that there is an issue of us 'not keeping up with the latest data technicalities nor

18 UCL, Institute of Archaeology/Heritage Futures research programme

building in long-term visions or longevity for this'. Interoperability is rare; renewal or migrations of data as systems develop are out of reach for most heritage organisations, though such issues do not tend to hamper collection efforts. The issue of a 'perfection problem' was raised in the Collaborative Dialogue, as an idea of 'completeness' has a tendency to lead data collection. Smart sampling strategies are not always possible to implement. Funding streams have not yet taken into account the growing importance of existing data, its uses and users, leaving support for projects that might improve it trailing. A parallel emphasis on innovation also creates a perennial backlog in system efficiency. As one delegate put it, 'we're always looking to the last big thing'.

As the British Academy and Royal Society (2017: 18) points out:

> The ease of collecting and managing large volumes of data in 'big data' platforms and the availability of new tools to analyse such data – such as machine learning – means that large volumes of data can be collected, integrated and analysed in ways that generate unexpected patterns or insights which go far beyond the original intended purpose of data collection.

This generation of unexpected patterns and insights makes Ranade's call to become a 'digital archive by instinct and design' ever more relevant. She raised the need to promote an understanding, or organisational 'vision', so that all staff can be involved in recognising the benefits of data, and the need to change our approach design.

This turns to the issue of data management: the current benefits and risks associated with data creation and usage are pushing the need for more regulation and clear guidelines related to data. Delay in policy means organisations are themselves beginning to steer effective management (Science and Technology Committee, 2015; Royal Statistical Society, 2015).

The impact of this data revolution on concepts (or assumed rights such as freedom, privacy and equality) integral to social democracies is at the heart of how data governance will develop. Meanwhile, vast developments in technology and data usage bring noticeable benefits to society. The heritage sector can and is playing a role in the development of governance, as government struggles to catch up in terms of policy and regulation.

Areas such as the potential commodification of personal data, the usage for private interests or group profiling, inaccurate analysis leading to policy or legislation, concerns over security, and the datafication of everyday life all cause concern to the wider public and need to be considered by those responsible of large datasets.

Data Discovery, Assess and Reuse

The heritage sector is exploring varied methods of analysing data, which is accumulated from multiple sources over time and has the potential to provide unexpected insights. However, data interoperability across organisations and sectors remains challenging. While data is becoming ubiquitous, the basic discovery of data remains difficult and its usage is still relatively low.

Several workshop participants emphasised the importance of data standards and referred to the work institutions are doing to support development of relevant standards in order to enable effective sharing of data and long-term preservation of data (May; Farquhar). While the work is advancing in the range of areas such as vocabularies, persistent identifiers and resolution of multilingual challenges, there are still many issues arising. Challenges that were mentioned include difficulties in dealing with legacy systems, issues in tackling complex and messy data and multiple formats, need to address software issues as much as data itself etc. It was also discussed that there is more connectivity needed between different domains – e.g. historic environment, museum objects, archives and libraries.

It was also pointed out that in some cases there were significant issues remaining around legal frameworks – e.g. copyright and IP – especially affecting our understanding what use and re-use is permitted (Farquhar; Ranade). This is especially urgent in relation to new research uses of text and data mining, which is becoming more prevalent but is still in many cases limited due to legal or technical restrictions. The workshop concluded that while there are many useful initiatives, there is no current roadmap of how we can move towards linked data in more strategic way.

The participants were clear that greater use and reuse of data would benefit both research and public audiences. However, there is much work to be done to mainstream relevant tools and services, and even to fully understand the needs of these audiences.

Digital Skills

The issue of skills for managing data and developing future data services was mentioned throughout workshop. While some relevant skills are in existence in the larger organisations, the sector does not have a full understanding of the emerging roles and competencies required to manage its data needs. Shortage of data skills is prevalent in all sectors of economy and relatively low salaries in heritage sectors make recruitment and retention of specialist staff difficult.

Many participants pointed out that we need better understanding of digital and data roles required by institutions, what are educational and career pathways for any such roles, and how these roles fit with organisational strategies.

The important role of visualisation in conveying data has been mentioned as of particular importance. It has been noted that, in the cases when data in heritage works well, it is a particularly efficient route to engage new audiences and present new insights, and that it is also beneficial as it could lead to better data literacy of young people and public in general.

Below are some of the research questions that came to light during the Collaborative Dialogue.

The need for an information exchange space: how could a common hub be created and sustained?

Would a 'heritage data institute' be sustainable? Such an entity could provide a central knowledge exchange that worked across silos: a repository for standards and guidance; a platform for sharing experience and methodologies; for working out best practice across institutions; a training and skills base for continued professional development.

Perceptions of data: how could an analysis of the use of big data usefully address anxieties around its misuse (or perception of misuse)?

How do we tackle the issue of privacy? How can we express limitations of data presented to different audiences? How do we deal with the levels of certainty and ensure balanced and transparent interpretation?

What is 'Heritage'? How could data work towards an understanding of the way 'heritage' as a concept is constructed by various groups?

Focus on audiences and public value: how can we better understand needs and opportunities in relation to different audiences' engagement with data?

How can we develop new types of engagement and services with heritage by using data? What is good practice in analysing user needs in relation to digital resources? What is the new value that data can bring to the ways that different communities interact with heritage? What difference does this make to society? How do we capture and measure this value?

Heritage data infrastructure: what constitutes the UK heritage data infrastructure and are its constituent parts fit for purpose, adequately resourced and linked?

How do we develop better systems for finding and using non-text data (e.g. audio, visual or object based)? What work is needed to develop useful APIs? Do we understand the cost of data? How is data related to analogue collections? How can we ensure that there is more strategic view and better investment in long-term digitisation? How do we evolve our legacy systems? How do we improve data skills across the sector?

Structures of data: could more informed knowledge of how data is structured and constructed contribute to better practice for heritage data?

Can we define metadata standards that will adequately enable sharing, archiving, reuse and discovery of heritage data? Are current vocabularies working between heritage and technology? How can we effectively track usage?

Linked data and data aggregation: how can data/datasets be usefully aggregated and/or linked, and what data could be extracted?

For example, by aggregating or linking data around particular heritage objects, could they usefully contribute to their own care and maintenance? Would aggregation provide better insight into siloed datasets?

Legal and policy obstacles: what are systemic obstacles in developing heritage data, and how we can move forward?

How is copyright law enabling new data services, or is it preventing sharing and reuse of data? Are there new issues arising in understanding provenance of digital heritage and how policies need to develop to address this? Are research and cultural policies aligned in the best way to ensure links between cultural digital developments and the UK research agenda?

The workshop was successful and yielded practical immediate recommendations as well as longer-term and more complex recommendations. These include:

It is a myth that the more data is collected, the better one can understand the data: there is a huge need for the knowledge of domain experts. Institutions should ensure that data initiatives bring together data and domain experts.

It is critical to understand the population under study when generalisations are used (e.g. as part of automated algorithm driven operations) and to recognise that conclusions of any data are only ever relevant for the population of that data.

It takes repeated interactions to build trust amongst partners and the public that avoids the overselling, and hype, of big data.

We need to change our practice and thinking to become digital archivists by instinct and design, and pull away from traditional lifecycles.

Tackling resistance and public anxiety are intimately dependent on transparency and issues of privacy: these notions must be addressed explicitly.

There needs to be more conversation about what heritage data exists, what is recorded, and how it is being used at a local, regional, national and international level.

There is a need to identify changes that bring about challenges and benefits, and to map where to invest in relevant infrastructure and resources to stay ahead.

Understanding of value of digital resources and their use by different audiences is important. Further work is needed to understand specific ways in which this could be done in order to gain meaningful insights.

There is a need for the sector to be able to collectively embrace data opportunities and define its key challenges. Collective voice would be more effective in policy and investment related advocacy, as well as raising the profile of data with audiences.

Greater collaboration on infrastructure issues that enable interoperability and more integrated

discovery of heritage data is key requirement if the sector is to enable effective use and reuse of digital collections.

To ensure that this initiative continues progressively, participants of the workshop also discussed next steps. Moving forward, the initiative intends to:

- Produce this report on the discussions and concerns raised in the workshop.
- Explore the development of a virtual information exchange space, or hub, to share experience and best practice (noting the need to be mindful of developments in relation to the European Research Infrastructure for Heritage Science http://www.iccrom.org/e-rihs-new-eu-alliance-for-cultural-heritage/ and other European Open Data for Heritage developments).
- There is a need to engage successfully with natural heritage in our understanding of data science in the heritage sector as a whole.
- There was consensus of the usefulness in organising a series of workshops on:
    - Research methods
    - Immersive Research Development and Partnerships
    - Metadata and other standards enabling greater collaborative work to enable discovery, data sharing and preservation
    - Technology and data science innovation
    - Ethics and privacy issues and solutions
- Discuss and further explore how to grow engagement with the wider research community and public.

ASA, 2011. Ethical Guidelines of the Association of Social Anthropologists of the UK and Commonwealth. Available at:
https://www.theasa.org/downloads/ASA%20ethics%20guidelines%202011.pdf

British Library, 2017. British Library Data Strategy 2017
http://blogs.bl.uk/files/britishlibrarydatastrategyoutline.pdf

Cabinet Office, 2008. Data Handling Procedures in Government: Final Report. Available at:
https://www.gov.uk/government/publications/data-handling-procedures-in-government

European Commission, 2015. Creating Value through Open Data. Available at:
https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf

EU Data Protection Reform and Big Data, 2016. Available at:
http://ec.europa.eu/justice/data-protection/files/data-protection-big-data_factsheet_web_en.pdf

EU General Data Protection Regulation (GDPR). More details at:
http://www.eugdpr.org/

E-RIHS European Research Infrastructure for Heritage Science. More details at https://e-rihs.ac.uk/

Golant Media Ventures and NESTA, 2017. Evidence: The adoption of digital technology in the arts
http://www.nesta.org.uk/sites/default/files/difaw_gmv_e.pdf

HM Government, 1998. Data Protection Act. Available at:
http://www.legislation.gov.uk/ukpga/1998/29/contents

HM Gov, 2012. Open Data White Paper: Unleashing the Potential. Available at:
https://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf

HM Government: Horizon Scanning Programme, 2014. Emerging Technologies: Big Data. https://

www.gov.uk/government/uploads/system/uploads/attachment_data/file/389095/Horizon_Scanning_-_Emerging_Technologies_Big_Data_report_1.pdf

National Archives, 2015. Guidance on the implementation of the Re-Use of Public Sector Information Regulations 2015. Available at: http://www.nationalarchives.gov.uk/documents/information-management/psi-implementation-guidance-public-sector-bodies.pdf

Open Data in Europe. More details at: https://www.europeandataportal.eu/en/dashboard