Cultural Heritage Information: Artefacts and Digitization Technologies


Since the 1970s, the gallery, library, archive, and museum sector has promoted and encouraged digitization - to increase access to cultural heritage material through various incarnations of digital media. Indeed, it is now expected by both users and professionals that institutions should be undertaking digitization programs, and best practices in this area are now well documented and understood. This chapter scopes out the background to the current digitization environment, giving an overview of the methods and approaches involved. It points to current developments, highlighting the use of both two and three dimensional capture methods for the creation of digital surrogates of objects and artefacts, indicating the potential for further development in the sector, whilst drawing attention to current issues faced when digitizing objects and artefacts including cost, sustainability, impact evaluation, and expectation management in the changing information environment. The affordances of previously prohibitively expensive techniques – such as multi-spectral imaging and 3D scanning – are now available at relatively inexpensive rates, which also raises questions about digital literacy and our understanding of what it means, for both the end user and information professional, to create digital versions of our cultural inheritance.

1. Digitization of cultural and heritage content

“the conversion of an analog signal or code into a digital signal or code” (Lee 2002, 3),
information, culture, and heritage sectors were quick to embrace digitization technologies as they became available, primarily to facilitate access to items in collections by providing them in electronic format, firstly with the creation of electronic catalogues in the 1970s, then conversion of printed source material into digital files in the 1980s (van Horick 2005). One of the first digitization projects was the “Optical Digital Image Storage System (ODISS)” launched in 1984 by The National Archives and Records Administration (NARA) in Washington which aimed to test the utility of digital image and optical disk technologies for the reproduction, storage and retrieval of archival documents (see González 1992, 1998, and NARA 1991 for an overview).

As network infrastructure grew, and the cost of computational devices fell, users increasingly consumed digitized content (Naughton 2000), and libraries, archives, and museums continued to experiment with the appropriate application of digital technologies within their institutional remit, being

In the 1980s, digitization generally focused on specific
18th and 19th century journals (ILEJ 1999). At the start of the 1990s, then, “In the early days of the Web, museums provided some of the best content and some of the most compelling reasons to go on-line” (Peacock et al 2004). Towards the close of the millennium, “A decade of digitization and documentation for the Web … created a rich array of cultural and historical information across the museum, library and archive sectors” (ibid).
to uncover detail in cultural and heritage objects. An example of this is a project which has collected multi-spectral data on the degradation of an 18th century parchment by a series of physical and chemical treatments, in order to inform best practice on how to image similarly damaged texts (MacDonald et al. 2013).

The affordances of 3D capture and 3D printing are also being explored in the GLAM sector, such as the E-Curator project which established the value of 3D models for sharing detailed information about museum objects among curators and conservators (Robson et al. 2012).

Moving beyond capture practices, the use of social media, the relationship of social media to more staid digitization practices, and the relationship of institutions to individuals, are being investigated as the sector explores these relatively low cost mechanisms for dissemination of digitized material (Terras 2011). An example of this is the British Library's "Mechanical Curator" which randomly selects artwork from their database and posts it to a Tumblr blog every hour. The logical conclusion of this is the use of social media platforms in conjunction with digitized content to ask online users to help with tasks in the cultural and heritage sectors, through a process known as "crowdsourcing".

For example, the digitized manuscripts of the philosopher and social reformer Jeremy Bentham (1748-1842) have been put online via a wikimedia platform and are being transcribed by volunteer labour in the Transcribe Bentham project (see Causer and Terras Forthcoming): more than 3 million words of Bentham's writing have now been transcribed, speeding up the costly process of creating machine readable text from digitized images of manuscript material.

There is also huge interest in the cultural sector for the possibilities that are inherent in "Linked Data": the ability to link or merge data with sources that follow other standards, to allow different collections of digitized content to be searched, analyzed and visualized (de Boer et al. 2012). An example of this is the British Museum's Research Space which is developing a range of flexible tools to search across the two.

---

2 https://www.ucl.ac.uk/museums/petrie/research/research-projects/3dpetrie/3d-projects
3 http://mechanicalcurator.tumblr.com/
4 www.ucl.ac.uk/transcribe-bentham
5 http://www.researchspace.org/
2. The Digitization process

Digitization is dependent on capturing a representation of existing, analogue material (as opposed to “born digital” media which was created on a computational system). It is important to establish that the act of digitization is one of translation: the resulting digitized representation of an original analogue object is not a replacement for the object. Computational systems depend on exact numerical strings. The ordinary, or “real” world, of our senses exists in a continuous flowing stream of signals across time and often space. A document – or even a traditional photograph of a document - exists in analogue, where a varying signal represents a continuous range of values. In order to record, copy, transmit, or analyse such a complex signal using computational methods, it is necessary to translate this into a form which is more simple, predictable, and processable. All telecommunication systems work with one underlying principle: the information to be sent is converted into signals which can be transmitted, and reassembled on reception, to be converted into something we can perceive as a fair copy of the original. Digital systems are those which rely on a sequence of discreet numeric values, rather than the unconstrained and continually varying qualities of analogue signals. Numeric values are used in digital systems for processing, display, transmission, and input: often sampling values from analogue sources in the process we call “digitization”. Digital systems rely on the binary numeric system, where all numbers are represented using only two symbols, such as 0 and 1, known as binary digits or “bits”. Strings of bits can build up a representation of text, image, sound, or 3D object, but the more complex the representation, the more bits are required to describe it, and the more complicated the mechanisms are that are required to capture, store, display, process, analyse, and convert the information held in the binary data stream.
Providing a digital representation of an analogue object has various advantages: the digital representation can usually be more easily copied, shared, accessed, analysed and processed than the original (providing users have access to computational infrastructure – such as machines, software, networks, and subscriptions, which should not always be taken for granted, see Gooding 2013). The creation of digitized versions of primary historical sources also opens up new potential for research using advanced computational methods, to show different facets, relationships, views, or details of the original content. However it should always be remembered, that although digital representations of historical artefacts can be seductive, they are not the historical artefact itself, as they are only a digital representation limited to what has been captured during the sampling process:

A digital representation of an artefact is a representation of certain relevant characteristics of the artefact. It is not the original and complete artefact, nor even a metonymy or simulacrum of the complete artefact. It is only a representation of some “relevant characteristics” (Arnold 2008, p. 127).

2.1 Digitization Technologies
The introduction of digital camera backs as a means of digitization has superseded the previous methods which were flatbed and drum scanners. Both of these scanning methods are now obsolete and the standard for high-quality digitization nowadays is the use of professional digital camera backs in medium and large format made by companies such as Mamiya Leaf and Phase One. The accuracy and quality of the optical lens system used by these cameras is essential in minimising distortion. This method of capture … is also far quicker than the traditional scanning method as well as producing a higher quality image. This is due to the increase in size of the image sensor, commonly a CCD (charge-coupled device) chip in digital camera backs as well as the difference of the image sensor being static rather than having to pane across underneath the slide in a flatbed scanner… One of the advantages when using digital camera back duplication setups is the greater control over the light source, which is a lightbox consisting of a photographic flash with a uniform colour temperature… Since the launch of professional digital back cameras, there has been little development in scanning technology (Weidner 2013).
2.2 Digitization Guidelines

Producing digital versions of holdings through digitization has become an industry in itself, and there has been much effort devoted to producing guides to best practice for undertaking the digitization of library and archive material to provide a framework in which those who wish to undertake digitization of primary source materials can operate. The issue is that there are now over forty different guidelines for undertaking digitization in existence, and each of those
is a complex, multifaceted production. Each represents a synthesis of experience, drawn in part from earlier versions of a given guideline and in part from secondary literature, workshops and conference presentations (and other forms of hearsay), and the specific experience of consultants and other experts who develop a specific guideline (Conway 2009, p. 10).

Care should therefore be taken as to which guidelines are followed and used for advice, and part of project planning is to decide the protocols, benchmarks and procedures that a digitization project will use within a specific institutional context. Jisc Digital Media’ (is a good first port of call for up to date digitization advice, but other commonly consulted and respected guidelines include

3. Digitization in Libraries, Archives, and Museums

---

7 http://www.jiscdigitalmedia.ac.uk/
There are many reasons why a program of digitization of cultural and heritage material may take place, and there are a variety of potential advantages, which are summarized by Deegan and Tanner (2002):
immediate access to high-demand and frequently used items; easier access to individual components within items (e.g. articles within journals); rapid access to materials held remotely; the ability to reinstate out of print materials; the potential to display materials that are in inaccessible formats, for instance, large volumes, or maps; 'virtual reunification' – allowing dispersed collections to be brought together; the ability to enhance digital images in terms of size, sharpness, colour contrast, noise reduction, etc.; the potential to conserve fragile/precious objects while presenting surrogates in more accessible forms; the potential for integration into teaching materials; enhanced searchability, including full text; integration of digital media (images, sounds, video, etc.); the ability to satisfy requests for surrogates (photocopies, photographic prints, slides, etc.); reducing the burden of cost of delivery; the potential for presenting a critical mass of materials (p. 32-33).

In addition to this, there is further potential for raising revenue from the licensing of digitized content, raising the profile of institutions, and increased public engagement and knowledge transfer opportunities (Terras 2012).

However, there are also good reasons not to digitize content: Hughes (2004, p. 50-52), summarizes issues which cause problems in digitization projects, including unresolved copyright issues, lack of adequate funding, lack of institutional support, technical drawbacks, the potential for digitization to damage or compromise fragile or rare original materials, and issues which arise when digitization is used instead of adequate conservation techniques or instead of robust cataloguing systems. It is worth stressing that digitization is a costly and time-consuming endeavor, and projects are expensive to deliver, manage, and sustain (Denbo et al 2008): not everything is worth digitizing, once things are digitized the resulting representations have to be maintained and looked after, and therefore strategic choices need to be made as to where resources are best employed.

3.1 Digitization in Institutional Contexts
The potential inherent in digitization is dependent on a variety of factors, including, but not limited to; the nature of the primary source content; the approach of the project; the technology available, appropriated, and used; and the aims of the institution. The potential and nature of digitization therefore varies greatly from program to program and project to project, even within institutions. Looking at recent digitization initiatives in the author's home institution – University College London – alone, can give some indication as to the different scope and remit of digitization initiatives.

Digitization projects can revolve around one object and what it means for a particular community: for example, a project based at UCL and the British Museum digitized a 12 metre long plank-built war canoe from the Melanesian Southwest Pacific, dating from 1910. 3D laser scanning, paired with anthropological research, delivered a holistic virtual reconstruction, multimedia interactive delivery, and a 3D printed colour replica of a detail of the boat for digital repatriation to the source community (Hess et al 2009). This project was therefore a research project as well as one to create content: from the ethnographic point of view, the team were exploring how digitization technologies could work with museums and communities to build links around a particular object which now physically resides in a museum on the other side of the world to the community who revere it. From a technical point of view, the team were researching how best to utilize 3D scanning within this context. There is therefore still opportunity for digitization to be a fruitful research topic: as new technologies develop their potential and affordances for the cultural sector should be investigated.

In a larger project at UCL, digitization can be seen to be used to build up an illustrated catalogue of a particular collection. UCL Art Museum holds over 10,000 paintings, drawings, prints and sculptures dating from the 1490s to the present day, including an important collection of 16th-century German drawings and Old Master prints by Dürer, Rembrandt and Van Dyck, later watercolour material by JMW Turner, Thomas Rowlandson and Joseph Wright of Derby and a unique archive of

---

8 http://www.mhm.ucl.ac.uk/mhm-research/western-solomon-islands-war-canoe.php
9 http://www.ucl.ac.uk/museums/uclart/
works by staff and students from the Slade School of Fine Art, one of the leading art school's in the world, based at UCL. A digitization project based in the UCL Art Museum is gradually cataloguing and digitizing prints to create a definitive online catalogue of the art museum's holdings over a number of years, dependent on a variety of relatively small internal and external sources of funding. The online database is a work in progress as records are checked and updated, and images of digitized content are added, as they are created, and at time of writing includes around eighty percent of the collection. This online resource allows those interested in the contents of the UCL Art Museum to understand the scope of their holdings, and to find and refer to works in the collection more easily. The eventual addition of digital images of the art to its associated record will mean that users can access this catalogue online prior to visiting the Museum and requesting the artwork in question. Digitization, in this project, is then primarily used to assist access and record keeping, as well as to promote the content of the UCL Art Museum, and to assist both casual browsers and researchers in understanding the scale of its holdings. The digitization is being undertaken in-house as funding allows, with the resulting catalogue being hosted on UCL's own library servers, as part of a institution-wide activity to catalogue and make accessible the collections of various museums within UCL. The museum also links closely into UCL's teaching programme: with work placement students from the Master's degree in Digital Humanities aiding in the project, and suggesting and implementing new ways in which the image based content can be used. It has been noted, using UCL museums as a case study, that university libraries and museums can provide an ideal platform to experiment with different modes of digital delivery of cultural and heritage content (Nelson and MacDonald 2012). At the other end of the scale of digitization, the mass creation of digital records of a range of collections can be achieved to foster access to anyone in the internet-wide community who may wish to use them, collaborating in larger, federated digitization efforts. Another project at UCL, Europeana Travel11, digitized "over a million resources including maps, manuscripts, photos, films, books and postcards on the themes of travel, tourism, trade routes and exploration" to "enable public access to..."10

10 http://artcat.museums.ucl.ac.uk/
11 http://www.ucl.ac.uk/ls/europeanatravel/
ICT Policy Support Programme. The Enumerate project runs for three years from February 2011 and aims to “create a reliable baseline of statistical data about digitization, digital preservation and online access to cultural heritage in Europe” (ibid).

13 http://www.europeana.eu/portal/
14 http://www.enumerate.eu
The survey ascertained the state of digitization within each organisation, whilst also asking about expenditure, access to digital collections, and the institutional digital preservation strategy. 83% of the institutions which responded were partaking in digitization activities in 2011, and 100% of National Libraries have a digital collection. Only about 20% of the materials which are deemed to be important enough to be digitized currently are: Art Museums have the highest percentage of digitized content, averaging at 42% of their holdings, whilst National Libraries only have 4% of their content digitized, although they have a target to get 62% of their collections digitized within time. Photographs are the most digitized object type – 66% of respondents have digitized them, with 46% of institutions digitising archival records, 35% digitising drawings, 32% digitising posters, 31% digitising postcards, 29% digitising rare books, 29% digitising paintings, and less than a quarter of institutions digitising non-rare books, newspapers, serials, sheet music, microfilms, and video and audio recordings. The least digitized formats were 3D objects (7%) and whole monuments and sites (8%), showing that there are still issues in dealing with the organizational complexities and temporal and financial investment needed, to capture these. 34% of institutions have their own written digitization strategy, and almost one third (half in the case of national libraries) are included in a national digitization strategy. 31% of institutions have some kind of policy for the use of their digital collections, and 23% of institutions have a written digital preservation strategy. 85% of institutions say that they use web statistics to measure the use of their digital collections, yet only 42% of them say that they actively monitor use. These statistics indicate that there is much more to be done to ensure that digitization is a planned, targeted activity where the ramifications of the investment of limited resources are understood.
cultural heritage institutions are working on digitization). The costs of digitization are in the range of EUR20,000 to EUR40,000 per full time member of staff, unless the project involves audio-visual digitization, when the costs rise to EUR103,000 per full time member of staff.

The use of volunteers is common, being mostly used at archives and records offices. Funding for digitization is sourced from internal budgets in 87% of the institutions, although 40% mention some form of public grant or subsidy, 5% say they raised finances via private investment, and 4% indicate that they had some commercial sponsorship.

It can be seen, then, that digitization is now a core activity across the heritage sector, although there is a significant way to go until all the content that institutions deem worthy and necessary of digitization is captured, and that this will require much further financial investment and support from the sector, over a long period of time, to reach the goals they have set themselves in delivering digitized cultural and heritage material.

It should also be considered that the more accessible digitized content that there is, the more useful existing content will be, as rich connections can be made across collections, encouraging excellence in research, learning, and teaching.

5. Impact and Expectation

Although there has been significant investment in digitization in the cultural and heritage sectors for decades, it can be difficult to prove the value of this activity. This is increasingly necessary in an economic climate where “impact” of publicly funded activity is measured and evaluated: “The global economic decline that began in 2007 has led to serious cuts in funding for almost all humanities and cultural heritage initiatives, including the development of, and support for, digital collections” (Hughes 2012, p.2).

Yet it is difficult to establish value: “Digital resources are valuable to different audiences for different reasons, and some value may not be realized immediately… Value is subjective, changes over time and has different meanings that are contingent on external factors” (ibid, p. 5-7).

In response to this issue, Tanner (2011) suggests that there are five different modes of value in digitized resources, in order to provide an impact assessment of the often intangible benefits of digitization.

There is Option Value to digitized resources: when people enjoy the digitized content,
and produce some form of output, such as research output by academics. There is Prestige Value, where people derive utility from knowing that a digitized resource is cherished by individuals both inside and outside of their community. There is Education Value, where people are aware that their digitized resources contribute to their own or other’s education, training, and knowledge. There is Existence Value, where people benefit from knowing that a resource exists, even if they do not use it themselves. And finally, there is Bequest Value, where people derive satisfaction from the fact their communities will be able to access these resources in the future (p.34).

Tanner’s modes of value are an extremely useful taxonomy for those attempting to frame discussion of the potential—or known—impact of their digitization project, to show the benefits that it could possibly have to a range of users, and perhaps persuade funding sources of the necessity for digitization.

An alternative means of assessing impact is provided by TIDSR (Toolkit for the Impact of Digitized Scholarly Resources), funded by the JISC e-Content Impact and Embedding Programme, which provides a framework of both qualitative (interviews, focus groups) and quantitative (log analysis) methods for assessing the use of a digital resource.

At a time of rapid development and creation of digitized resources, we are only now being able to frame discussions about the use and usefulness of them: “The value, impact and use of collections take time to evolve and to be understood, and this needs to be reconciled in a world of responsive, short-term funding opportunities” (Hughes 2012, p.10).

The current lack of external funding for the creating of digitized material can be juxtaposed with issues of user expectation regarding digitized content. With ongoing changes in networked technologies, including the emergence of ever-connected smartphones in the digital environment, the number of potential users of digitized cultural and heritage content has increased enormously. Expectations have been raised that all content is available digitally (even though, as we have seen above, only a small fraction of institutional holdings are currently available in digital format). Expectations and needs of potential user groups differ, and it can be difficult for heritage organizations to be agile enough to respond to a rapidly changing.
Digital tools have worked their way into every aspect of our lives. Even those people with high levels of digital fluency take for granted the increased levels of speed, access and mobility that as little as three years ago were hard to imagine. For many, the possibilities still feel a little bit sci-fi, but as William Gibson observed so insightfully in 2003, “the future is already here – it’s just not evenly distributed”. This is as true for the cultural sector as anyone, with many organisations struggling to embrace the new reality of audience behaviour, let alone go boldly into a future of big data, the semantic web and seamless participation (Finnis 2013, quoted in Malde et al 2013, p.2).

A report recently published by Culture24, a non-profit cultural organization which supports arts and heritage venues to reach audiences, suggests that institutions should look at the range of digital material they are creating, and the way it is delivered, and to encourage: the measurement of value (both what the institution and their users value, before exploring how these can be enhanced through digital channels); regular reporting and analysis of what works for the institution and what doesn’t; a focus on user behavior and demand, moving away from the institutional supply-driven model of digital content; understanding the changing behaviours of audiences; being honest about the effort and time it takes for digital activities to be carried out; understanding web, mobile, and social behaviours; and to conduct experiments in the way content is created, communicated and delivered (Malde et al 2013, p.5-6). This will, then, affect the digitization policy and strategy of an organization: “The starting point should… be the mission of the organisation and the needs of the target audience. You need to know what you want to achieve and who it is for” (ibid, p.4.). Perhaps the biggest change in digitization in recent years is this necessary move towards understanding user needs and communities, rather than creating “scan and dump” or “build it and they will come” digitization projects (Warwick et al, 2008). Although, as we have seen in the Enumerate survey above, only one third of major cultural and heritage institutions across Europe have considered their digitization activities in enough detail to have such a policy.

The starting point should… be the mission of the organisation and the needs of the target audience. You need to know what you want to achieve and who it is for” (ibid, p.4.). Perhaps the biggest change in digitization in recent years is this necessary move towards understanding user needs and communities, rather than creating “scan and dump” or “build it and they will come” digitization projects (Warwick et al, 2008). Although, as we have seen in the Enumerate survey above, only one third of major cultural and heritage institutions across Europe have considered their digitization activities in enough detail to have such a policy.
It is also worth returning, here, to the issue of representation in digitization, in the light of increased user expectation and regular use of digital media. As people turn to the digital as a convenient means to view and access a wide variety of digitized cultural and heritage content, we have to ask: do they understand the processes which created it, and what they are looking at for evidence, or sources of information? The persuasive nature of digital visualisation and display can mean we do not stop to question the very nature of digitized content: its labor of production has been concealed and therefore bears less evidence of authorship, provenance, originality, and other commonly accepted characteristics attributed to physical objects. For these reasons the digital object's materiality is not well understood (Cameron 2007, p. 70).

Understanding digitisation is then a particular extension of digital literacy: the confident and critical use of Information Society Technology (IST) for work, leisure and communication. It is underpinned by basic skills in ICT: the use of computers to retrieve, assess, store, produce, present and exchange information, and to communicate and participate in collaborative networks via the Internet (Digital Literacy, 2010). This is combined with information literacy: knowing when and why you need information, where to find it, and how to evaluate, use and communicate it in an ethical manner (CILIP 2013).

Those creating digital material, and particularly digitized cultural and heritage content, have an obligation to their users to explain the nature of the resources they are publishing, and fully documenting and describing them (something which seldom happens). Indeed, there may be a further obligation to inform the user community about issues of representation so they fully understand the material that they are accessing: this particularly comes into play when dealing with advanced representations of material, such as 3D scans, or multi-spectral images, where it is difficult to understand the relationship of the primary historical source to its digital representation. If institutions, and by extension, their users, do not understand the digitization process, can our interpretation of digitized material ever be robust? This is an issue that the sector needs to explore with greater awareness.

Conclusion
This chapter has provided an overview of the current digitization environment in the cultural and heritage sectors. After a period of adoption and experimentation we are now at a juncture where institutions are expected to provide digital versions of their holdings, although there is much work to be done to create digital representations of all that is deemed to be digitization-worthy. Digitization has been shown to be a complex and costly process, in which the translation of analogue content into a digital form is only a small part: much is dependent on the institutional framework, resources, and aims in which the digitization project operates, rather than merely considering technical issues about capture and storage. Best practice in newer areas of digitization, such as 3D capture and printing, or the use of multi-spectral imaging, is still being investigated, and it will continue to be the case that as new technologies emerge their affordances should be explored for the particular use requirements of the cultural and heritage sectors. In addition to technical aspects, there are pressing issues regarding the use and usage of digitized resources, and how we can show that the digitization process adds value to the user experience, and to society at large. We are only just beginning to understand how institutions can best respond to a rapidly changing information environment, and evolving user expectations – which may not be combined with the same advances in digital literacy within our use communities.

Throughout this chapter, the importance of understanding institutional contexts and aims has been shown to be paramount when planning, or considering undertaking, a digitization project. Organizations must now take a more holistic view of the digital environment in which they operate, to understand their users, their resources, and the every growing potential inherent in the creation of digitized versions of historical and culturally important content. A wealth of digitized heritage material has now been created, with many more digital treasures to follow in the forthcoming decade. Attempting to understand the possibilities this delivers across the sector, to a wide range of users, is now part of the role of every individual digitization project, as we look to a future where further advancements in digital capture, access, search, analysis, and dissemination will affect how users interact with, perceive, and understand their cultural inheritance.
8. Bibliography


Causer, T. and Terras, M. (Forthcoming) “Crowdsourcing Bentham: beyond the traditional boundaries of academic history.” Accepted, International Journal of Humanities and Arts Computing.


Terras, M. (2012). Digitization and Digital Resources in the Humanities,