

1

Digitization in the context of collection management

Introduction

Digitization projects take place within the context of a global market and institutional goals. The dramatic growth in digitization since the late 1990s is a direct response to the rapid development of the internet and corresponding changes in user expectations. This is part of an overall shift within the heritage sector as a whole and has had an inevitable impact on the way that collections are managed, and the role of the conservator.

The impact of new technologies

The internet has become the principle source of information on many aspects of life for millions of people of all generations. This digital revolution has fundamentally changed the way that we live and work. While there are great benefits to the availability of information via the internet, it has put new pressures on libraries, archives and museums. Increasingly, people expect to find and retrieve information online, especially the generation who have grown up with the internet and have often not learned to use and interrogate catalogues. Research using catalogue information typically leads to focused exploration of a subject in a particular context. In contrast, the internet works on a keyword search basis. Search results via keywords on the internet are vast, and the enormity of information thrown up often leads to an assumption that information not available on the internet is nonexistent. The growth of the internet has therefore fostered a new generation of users who not only expect all information to be available online but also expect it to be fully searchable.

At the same time, new technologies have brought new opportunities for cultural institutions – many libraries and archives have been involved with digitizing their collections for some years now. Digitizing collections has a great number of benefits, many of which go beyond the most obvious of increasing the general accessibility of a collection (Hughes, 2004, 9–16). Making a collection available online will facilitate access to it for a much broader audience. This can result in a shift from traditional users, such as academic researchers coming to use original material within an institution's reading room, to a much broader and more diverse user-base who will discover and subsequently use information that did not previously seem available to them.

Furthermore, digital technology can enhance access, allowing the user to browse and compare content and enabling an institution to link the content of its own collection with other relevant online material, thus building up a large pool of interlinked information. Digitization therefore offers the possibility to develop a collection through bridging the gaps within one's own collection by adding information from elsewhere. Clearly, networked information between different cultural institutions enhances the scope for research on item level. Examples include the work on *Codex Sinaiticus* (2010) and the work of the European Commission to combine Europe's multicultural and multilingual heritage in *Europeana* (2010) on a collection level. In addition, digitization coupled with other optical technology enables information to be read and interpreted in new ways, further advancing research, for example the Walters Art Museum's work on the *Archimedes Palimpsest* (2008). Lastly, new ways of using and looking at the collection have opened up through combining it with other available information, for example the use of georeferencing in Tate's *Art Map* project (2010).

As online presentation offers the possibility to put together content in any conceivable way, there is great scope to develop specific education packages, directly supporting national or even international school curricula, for example, in the UK, The National Archives Education Service (www.nationalarchives.gov.uk/education) and the Science Museum's classroom resources (www.sciencemuseum.org.uk/educators.aspx). At the same time, online presentation allows formerly inaccessible material, such as negatives and fragile documents, to be made accessible. This is also where digital access aids preservation, as digitization enables research without

touching the original. On a higher level, digitization will help to raise the profile of an institution, as an online presence will raise awareness of the existence of a collection and bring great opportunities to market it. In addition, it is possible to develop strategic links with related institutions, and thus re-position the collection within a broader or brand new context.

There is ample evidence that the possibilities offered by digitization together with new technology has expanded the vision of institutions. There are several prominent examples of institutions that have redefined their goals with a view to digital technology.

- The Smithsonian Institution in the USA makes ‘the increase and diffusion of knowledge’ its mission within the strategic plan 2010–15, imagining ‘access to all known information . . . with one touch of the screen’ (Smithsonian Institution, 2010).
- The British Library’s strategy for 2008–11 aims to ‘advance the world’s knowledge’ through seven strategic priorities, five of which aim to either advance the British Library’s digital infrastructure or use digitally available content of the Library (British Library, 2008).
- The Koninklijke Bibliotheek (National Library) of the Netherlands states as its mission to ‘bring people and information together’. Three of its strategic priorities in 2010-13 relate to digital content, with the top priority being to offer ‘everyone access to everything published in and about the Netherlands’ (Koninklijke Bibliotheek, 2010).
- The National Archives and Records Administration (NARA) of the USA revised their strategy in 2009 and aim to make their ‘holdings and diverse programs available to more people than ever before through modern technology and dynamic partnerships’ (National Archives and Records Administration, 2009).
- The National Archives of Australia (NAA) states in its corporate plan 2009–12 that it will ‘continue to explore ways in which technological developments can enhance access to our collections’ (National Archives of Australia, 2009).

Commercial partnerships have since developed in this sector, especially where the scale of the task makes internal resourcing impossible and the content is attractive to publishers and other online providers. The most

noteworthy examples within the library sector are Google Book Search and Microsoft Live Books, whose aim it was to create searchable databases of full-text books (Rieger, 2008, 4–9). Archives, on the other hand, have taken advantage of the increasing interest in genealogy, and embarked on commercial partnerships focusing on family history content (e.g. partnerships with companies such as Ancestry, Brightsolid or FamilySearch). The advantages of teaming up with commercial partners in both cases are that digitization, including transcription, online presentation and marketing are resourced through the commercial partner, while the institution will get a copy of the scanned images. It can also provide a steady revenue stream as long as the demand for the digitized content remains strong. At the same time, digitizing cultural heritage collections benefits the private sector by stimulating new parts of the economy and encouraging creative thinking to develop new kinds of services using these images.

Commercial partnerships are not the only way that digitization projects are funded. Many grant-giving bodies will fund digitization and there are now specific funding streams dedicated to supporting institutions in their use of digital technologies; for example, the Joint Information Systems Committee (JISC) was set up in the UK to support education and research institutions by supplying funding and expertise. However, even grant-funded projects may have a commercial aspect if an institution uses a contractor to carry out the imaging and online delivery. Either way, the introduction of a commercial element adds new dimensions to collection management: an institution's staff may now need to work closely with a commercial company and the user may have to balance the cost of travelling to see the original against paying to see content online. Working with commercial partners is not inevitable but it is part of an overall shift in collection management and the way in which collections are used and accessed.

Collection management

The desire of libraries and archives to make books and documents more widely accessible fits into the need for accountability, especially where the spending of public money is involved. This has required collection managers to take a fresh look at all activities carried out within an institution. Waller (2003, 29–31) distinguishes three main processes in collection management:

- 1 The development of the collection through adding or removing items. This also includes activities such as processing, cataloguing, restoring and researching.
- 2 The use of the collection, which includes any activity benefiting from utilization of, reference to, information about, or the existence of the items contained within it.
- 3 The preservation of the collection with the aim of maintaining its value to an optimum level.

In all three processes, it is the value of the collection that is the focus of activity. Accessibility then forms the link between the collection, the values it represents and the specific process (see Figure 1.1). Collection management involves making well informed decisions in order to prioritize actions and optimize the allocation of resources to maintain as much accessible value as possible.

The development of digital technology and the opportunities that this has brought have created new challenges for the management of collections. For the conservator, who has traditionally focused their attention on preserving and developing the collection, emphasis is gradually shifting to *using* and

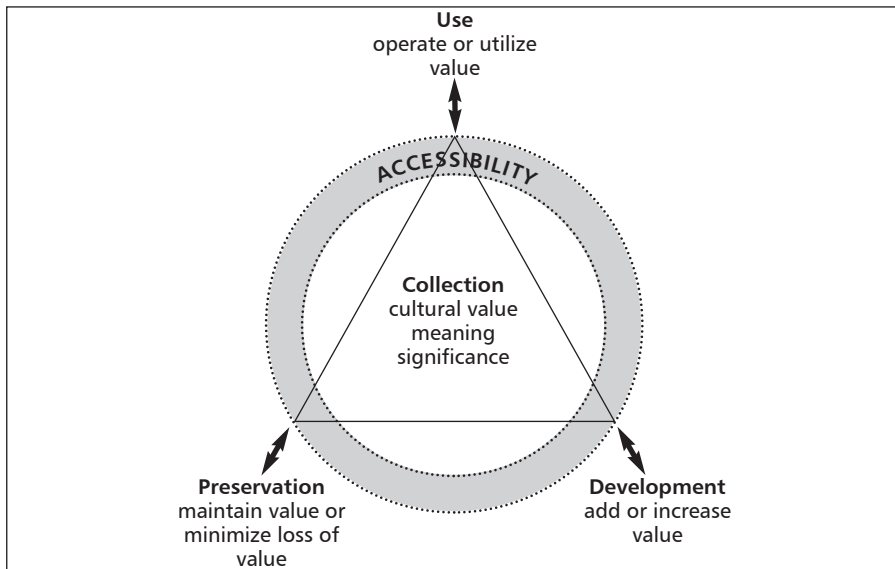


Figure 1.1 Collection management triangle linking collection value with accessibility (Brokerhof et al., 2009)

developing the collection. Preservation and use have thus far been seen as competing objectives; however, no library or archive keeps its collection just for the sake of keeping it. It is being kept with the expectation of being of value to current as well as future generations. Digitization has the potential to benefit both the development of a collection as well as its use. A well executed digitization project will result in detailed cataloguing information and metadata about a collection, while enhancing possibilities for its use in ways previously not conceived.

Doing this in a way that is sustainable, and that therefore benefits the preservation of the collection, as described in this publication, requires a pragmatic, risk-based approach from the collection manager. In addition, flexibility and the ability to understand the competing requirements of all interested parties, including commercial partners where those are involved, are pivotal in successfully balancing the preservation needs of the collection with the drive to make as much information as accessible as possible.

Access to content and context

There has been a long tradition amongst librarians and archivists to select and keep books and documents for the information they contain – information that was considered to add value to the collection as a whole within its specific context. Some items may have been selected for aesthetic reasons, such as the artistic value in their illuminations (even if the text could not be read or understood), but often little if any value has been attributed to aspects of their physical form, e.g. paper, parchment or media; or format, e.g. folding techniques, binding techniques or binding styles. At the same time, conservators have primarily focused on material attributes only, assuring readability and the ability to handle the object, while often staying oblivious to the content and context of an object.

Within the context of ever increasing digital possibilities, there are two seemingly opposing trends in the management of collections: on the one hand the desire to increase access to books and documents through online presentation, which conveys only the information and none of the material aspects, and, on the other to conserve material integrity as far as possible by carrying out less interventive conservation treatments, with the aim to preserve material as well as conceptual and contextual value for users now

and in the future. However, although these trends may seem to conflict, they actually facilitate and complement each other. If a collection is available online, then the original documents are handled less and there is consequently less need for interventive treatments to keep them robust. In this way online access is in line with an overall trend in conservation towards minimal intervention. For the digital user there is a trade-off: they have the convenience of accessing the digital collection online from their own home, but they see a digital image and so do not experience the physical document. In the future the growing emphasis on the virtual rather than the material may result in a loss of appreciation of the material itself, or conversely the opposite may happen – as experiences become predominantly virtual there may be a growing reverence for the material world. The long-term effects remain to be seen (Nichols and Smith, 2001).

There are now two different ways to access collections: i) a single user can gain direct access to the physical object within the reading rooms of an institution, or, but possibly at the same time; ii) multiple users may have indirect access to the virtual object through online platforms. In order to make an item physically accessible to a reader, an institution not only has to have it listed and described within its catalogue, it also needs to have appropriate retrieval systems in place to allow it to be found within the repositories and delivered to the reading room. Without an entry in a catalogue, a potential user has no way of knowing of an item's existence within the collection; without internal retrieval systems, staff are unable to locate and deliver it. Both of these systems are therefore absolutely pivotal to making a collection accessible. In contrast, indirect access through the virtual, digitized item does not require any physical provisions through the originating institution. However, it must not be forgotten that appropriate systems need to be in place for online presentation in order to satisfy the needs of a user. The information needs to be indexed, tagged with keywords or similar, and the hosting institution will have to ensure that digital systems are operating at appropriate capacity to allow virtual presentation. In return, an institution may exploit the possibility of increasing the knowledge about their collection through using the collective knowledge of the research community within their virtual presentation platforms, for example The National Archives UK's wiki site (<http://yourarchives.nationalarchives.gov.uk>).

It has been suggested that digitization will make the need to facilitate direct access to documents through reading rooms redundant in future. Ramsholt (2009) suggests that digitization will result in a decline in reading room visitor numbers. However, the picture is not consistent. The National Archives UK, for example, had put over 80 million documents online by 2009 and while within a single month up to 12 million images might be downloaded, the number of document deliveries to the reading rooms – approximately 50,000 per month – has not declined but rather has remained constant. In fact, it has been shown that online presentation tends to generate interest and people who would previously not have thought of coming to the archive will now be drawn to visit. Strategic considerations to digitize the ten most popular record series would account for less than 20% of the total number of documents delivered into the reading rooms over the course of a year. At the same time, reducing document deliveries through this level of digitization has been estimated to require about six times the current annual operating budget of The National Archives UK, while the remaining 80% of physical document deliveries into the reading rooms would still need to be resourced. A further difficulty is that these 80% are often documents required by specialists such as journalists and academics – as opposed to the more predictable requests made by genealogists – and are therefore impossible to predict and pre-empt through digitization (Morley, 2009).

Planning and processing with minimal risk

Often, the value of cultural heritage is measured by the level of participation or use, resulting in an emphasis on numbers. This favours items for digitization which are of interest to many over those which might benefit for preservation reasons (de Lusenet, 2004). In addition, the idea that digital images constitute a preservation copy in a similar way that microfilm does has been disputed, mainly due to the uncertainties of digital preservation (Hughes, 2004, 8). As long as the challenges of digital preservation have not been fully met, digitization of any book or document cannot be seen as a preservation measure for the original itself. However, digitization projects will still help to improve the long-term prospects of a collection. The availability of digital copies reduces handling of the original, while an imaging project can be an opportunity to carry out

rehousing or improve storage conditions. So while digitization is unlikely to solve all the problems of preservation there can be tangible benefits.

Alongside the preservation benefits there are, however, risks associated with the imaging process. The process must be sympathetic to the document, causing as little damage as possible. Experience so far has shown that it is highly likely that access to the original will be required again, be it for access to information and material in the near future, or for re-imaging projects in the medium or long term. Image capture does not need to cause damage. There have been ample examples of extremely carefully considered projects over the years, for example, the Special Collections at the University Library Graz (2010), digitization of the *Lindisfarne Gospels* for the online gallery at the British Library (2010) and digitization of the *Book of Kells* at Trinity College Dublin (2010). In all cases, digitization involved either single manuscripts or small parts of a special collection. Yet the required throughput of projects involving thousands or even millions of pages does pose risks to the documents: equipment itself may cause damage, and high-volume handling and time pressures can necessitate working processes that may compromise best practice. The resulting relatively minor damage must be weighed against the many benefits to be gained from digitization and this balancing act is at the heart of the collection manager's approach.

A fact that is often overlooked is that the documents themselves can pose challenges to the image capture operation. There is a risk that the capture rate is tortuously slow because of the need for careful unpacking and unfolding, or that the resulting images may not capture all information because documents are damaged. In the worst case scenario image capture may have to stop altogether if documents are found to be contaminated with mould, or are too damaged to be handled. These scenarios can be anticipated and managed with careful, well informed planning that involves the input of a collection manager. The key role of the collection manager is to consider the physical attributes of the items, and how these attributes will influence the imaging operation. Experience has shown that it is essential that the collection manager is involved from the outset so that the planning of the imaging operation takes into account the nature of the physical items and their preservation requirements. The role of the collection manager is therefore very much that of a facilitator, using their

skills and knowledge to ensure the smooth running of the imaging operation while minimizing the risk to the documents.

The collection manager is not necessarily a conservator, but there are certainly parts of the project that require a conservator's skills and knowledge, in particular when assessing the condition of items and when preparing damaged items for imaging. Approaches to conservation differ according to the context, and a digitization project requires an approach that balances the need to address damage and the need for expediency, since projects often have tight budgets and timescales. In practice this means that interventive work is kept to a minimum and that treatments may be adapted and carried out in batches to maximize efficiency. Likewise, the documentation of treatments may be kept to a minimum. Since every digitization project is different it may be appropriate to adapt the treatment according to the type of equipment used. As previously mentioned, this overall approach is in keeping with current trends towards minimal intervention in conservation, except that in the context of the digitization project, the conservator must consider the way in which the items will be handled during imaging and the need for all information to be captured. These themes, concerning practical aspects of a digitization project, run through the following chapters of the book.

The four phases of digitization

In order to meet the objectives of both preserving a collection and increasing access to it, it is essential to acknowledge and plan for all phases of digitization projects from the very beginning through to the very end. A framework for such projects has been suggested by Rieger (2008, 16) and the Public Record Office Victoria (2010), although activities can be grouped into four distinct phases (see Figure 1.2). From this model it is apparent that image capture is only one part of a complex process.

Phase one

Phase one includes the selection of material (see Chapter 4 'The process of selection'), anticipating its use as well as assessing any copyright issues,

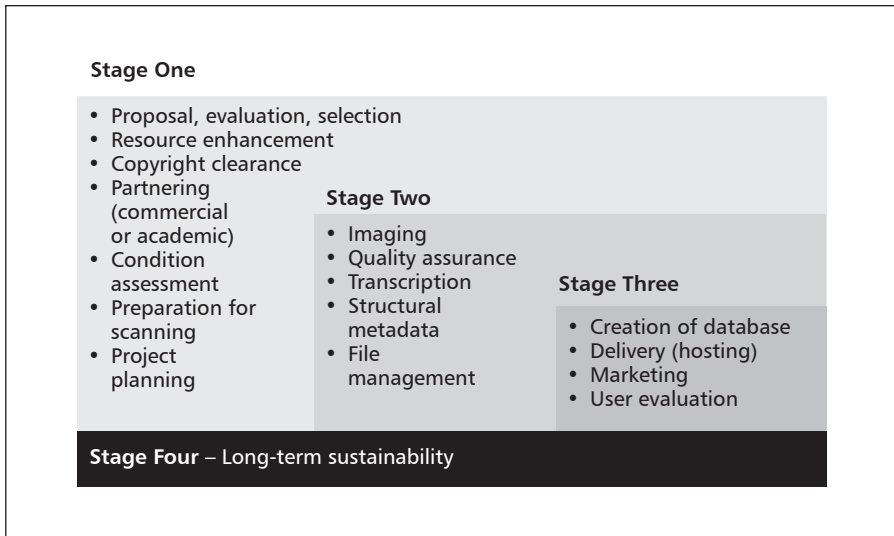


Figure 1.2 *The four phases of digitization*

which might be the deciding factor for some projects. At this point it is also necessary to consider the funding of the project, as digitization requires considerable financial resources. The content of the collection may point towards external funding sources. Any commercial company interested in selling digitized content will have a very specific remit, and therefore specific content will potentially attract specific commercial companies. For example, archival documents containing people's names and whereabouts might interest companies specializing in family history. Equally, there might be academic partners willing to fund projects if content furthers their learning and research aims, or grant sources where digitization fulfils the main remit of the funder. Equally, preservation may be the primary driver for a funder to grant money. Phase one will include a statistical survey of the collection by a collection manager or conservator to assess physical attributes so as to inform resource requirements and aid project planning. At this point, a decision regarding how to prepare documents for scanning will be made. This phase may therefore include the preparation of documents by a conservator, where it is possible to do so in advance.

Phase two

The second phase of a digitization project includes the actual image capture, the stage where the collection manager or conservator will be heavily involved. However, digitization does not only mean the actual taking of an image, it must also include processing the resulting images, the creation of metadata, the structuring of the image set and the creation of an archival master set of files (see Chapter 3 'The digital image'). Quality control must also be built in here. It should include, first, the control of the actual operation from a preservation point of view, ensuring as little damage as possible occurs, and, second, must include rigorous quality control of the resulting images, as it is more efficient to re-take unacceptable images while having the documents still at hand. Phase two may also include transcription or optical character recognition (OCR), where this is possible and desired.

Phase three

Phase three will consist of preparing the dataset for online presentation. There is also likely to be an element of website development for presenting the actual product as well as marketing and promoting it. Lastly, it should include incorporation of user feedback where possible, and the set up of ongoing user support. If digitization was carried out within the context of a commercial enterprise, then charging and related financial aspects will need to be set up and managed. This third phase already contains aspects that are ongoing, such as marketing and managing finances. These aspects must be recognized from the start; simply owning a set of digital images will not make information any more available than not having them digitized. Digitization will always result in long-term financial implications for the originating institution.

Phase four

Phase four has to be seen in this context, because sustaining the collection for most institutions will involve the management and financing of long-term preservation of the digital images as well as the originals. The preservation benefits of online access will only last as long as the digital images are available for online users. When collection managers consider the

preservation of their physical collections they typically think in terms of the next 100 years. As yet, there is still no certainty regarding the feasibility of the preservation of digital surrogates over such a time span, nor is there reliable data on the cost of digital preservation, even though current trends suggest that digital technology will become cheaper over the course of time (Chapman, 2003; Palm, 2005). Planning to keep digital content accessible for at least ten years should be feasible, but the resource requirements for maintaining a digital collection might even be higher than the continuing preservation needs of the original. At the same time, new technologies will continue to evolve, making it likely that collections will be redigitized, exploiting new optical possibilities in the same way that collections currently available as black and white microfilm surrogates are now getting digitized in full colour; and even items previously digitized are being redigitized in order to allow, for example, for 'turning the pages' technology. In the face of these developments, it is all the more important to balance the need for widened access to information with the preservation of the original, as the idea of a never-to-be-touched original has already been proven wrong.

Fitting digitization into collection management

There are some significant benefits for collection management brought on by digitization projects: the intensive work, especially when involving large parts of the collection, goes hand in hand with the acquisition of new knowledge about it. Materials and typical damage phenomena become apparent, which can feed into strategic aims for the collection. Equally, every digitization project results in improved cataloguing and metadata about this part of the collection, which in turn will enhance access and help to preserve the original by ensuring that staff find the desired item without much searching and the reader can see exactly what they were looking for.

Managing a collection in terms of its preservation, use and development, the contribution of digitization is two-fold (see Figure 1.3): digitization will help to preserve the collection by maintaining the value of the collection through a reduction in access to the original and allowing it to be stored more cost-effectively in terms of storage location and environment, where possible; at the same time, digitization allows the information contained within a book or document to be made accessible online on a much broader basis, therefore

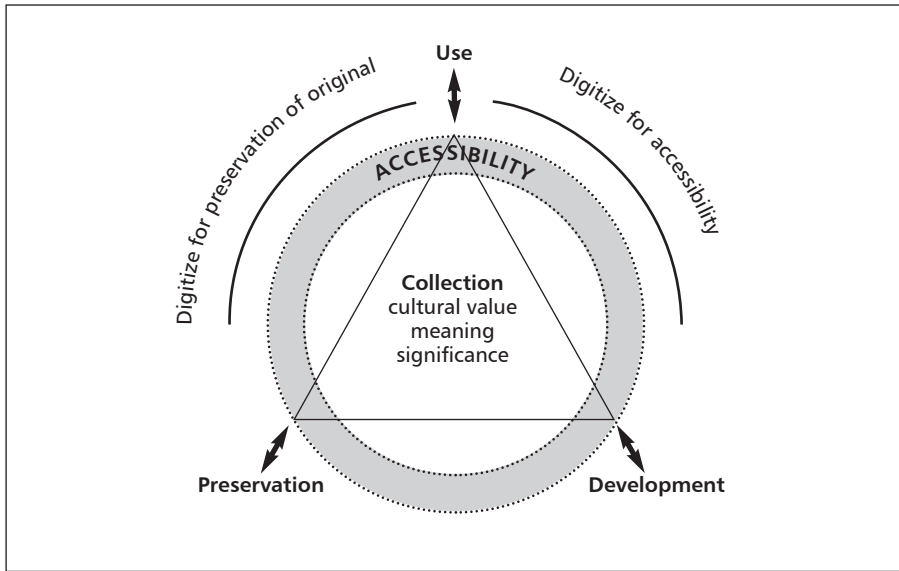


Figure 1.3 Collection management triangle linking preservation, use and development through digitization (Brokerhof et al., 2009)

supporting the use of the collection. It is for these reasons that digitization should be fully embedded within any collection management strategy, involving those charged with the preservation of the collection as well as those charged with making it accessible to the user.

Chapter summary

- Institutions can reach new audiences through online access, which also presents opportunities to build connections with other institutions. Online resources can utilize new technologies and provide a wealth of contextual information. Digitization aids the preservation of a collection by reducing the handling of the originals.
- Digitization might introduce a commercial element that is part of an overall shift in collection management that places more emphasis on the use and development of collections.
- Digitization is not in conflict with preservation but can complement and facilitate it. However, in the long term it is unlikely to result in an overall reduction in demand for access to original documents and is not a solution for

all preservation problems.

- Image capture can be done on a small scale without causing damage but in a large project, when thousands of documents are processed, some handling damage is inevitable. The collection manager must be involved at an early stage not only to ensure document welfare but also to ensure the smooth running of the operation. Physical attributes of the documents will influence the image capture operation.
- Digitization should be embedded in an institution's collection management strategy because it brings benefits to the preservation, use and development of the collection.

Bibliography

- The Archimedes Palimpsest Project (2008) *The Imaging of the Archimedes Palimpsest*, www.archimedespalimpsest.org/imaging_initialtrials1.html.
- British Library (2008) *The British Library's Strategy 2008–2011*, www.bl.uk/aboutus/stratpolprog/strategy0811/strategysummary.pdf.
- British Library (2010) *Online Gallery: sacred texts: Lindisfarne Gospels*, www.bl.uk/onlinegallery/sacredtexts/lindisfarne.html.
- Brokerhof, A. W., Ankersmit, B., Chin-a-Fo, J., Luger, T. and Scholte, T. (2009) *Values and Criteria: towards a mother model for valuation of cultural heritage*, Instituut Collectie Nederland, internal discussion paper.
- Chapman, S. (2003) Counting the Costs of Digital Preservation: is repository storage affordable?, *Journal of Digital Information*, 4 (2), article 178, <http://journals.tdl.org/jodi/article/view/100/99>.
- Codex Sinaiticus (2010) *Codex Sinaiticus: participants*, www.codexsinaiticus.org/en/project/participants.aspx.
- de Lusenet, Y. (2004) *Keeping Things that Work: preservation aspects of digitization*, EURESCO Conference on 'Philological Disciplines and Digital Technologies: computational philology: tradition versus innovation' (Il Ciocco, 6–11 September 2003), www.knaw.nl/ecpa/publ/ciocco.pdf.
- Europeana (2010) *Think Culture*, www.europeana.eu/portal/index.html.
- Hughes, L. (2004) *Digitizing Collections: strategic issues for the information manager*, Facet Publishing.
- Koninklijke Bibliotheek (2010) *Strategic Plan 2010–2013*,

- www.kb.nl/bst/beleid/bp/2010/index-en.html.
- Morley, O. (2009) *Digitisation and Our Options for The National Archives Future Strategy*, The National Archives UK, internal presentation.
- National Archives and Records Administration (2009) *Preserving the Past to Protect the Future: the strategic plan of the National Archives and Records Administration, 2006–2016, revised 2009*,
www.archives.gov/about/plans-reports/strategic-plan/2009/nara-strategic-plan-2009-2016-update.pdf.
- National Archives of Australia (2009) *Corporate Plan 2009–2012*,
www.naa.gov.au/Images/CorporatePlan2009-10to2011-12_tcm2-28444.pdf.
- Nichols, S. G. and Smith, A. (2001) *The Evidence in Hand: report of the task force on the artifact in library collections*, Council on Library and Information Resources.
- Palm, J. (2005) *The Digital Black Hole*,
www.tape-online.net/docs/Palm_Black_Hole.pdf.
- Public Record Office Victoria (2010) *Digitization Requirements*, PROS 10/02, Specification 1, www.prov.vic.gov.au/publications/publins/1002s1.pdf.
- Ramsholt, K. J. (2009) Digitization: threat or benefit to conservation?, *Journal of Paper Conservation*, **10** (2), 21–4.
- Rieger, O. Y. (2008) *Preservation in the Age of Large-scale Digitization*, Council on Library and Information Resources.
- Smithsonian Institution (2010) *Inspiring Generations through Knowledge and Discovery*, www.si.edu/about/documents/SI_Strategic_Plan_2010-2015.pdf.
- Tate (2010) *Art Map*, www.tate.org.uk/artmap.
- Trinity College Dublin (2010) *The Old Library & the Book of Kells Exhibition*, www.bookofkells.ie/book-of-kells.
- University Library Graz (2010) *Special Collections*, www.uni-graz.at/ubwww/ub-sosa/ub-sosa-handschriften.htm.
- Waller, R. R. (2003) *Cultural Property Risk Analysis Model*, Acta Universitatis Gothoburgensis.