

1

Why digitize? The costs and benefits of digitization

Introduction

The conversion of all sorts of cultural contents into bits and bytes opens up a completely new dimension of reaching traditional and new audiences by providing access to cultural heritage resources in ways unimaginable a decade ago.

(Mulrenin and Geser, 2001)

Over the last three decades, cultural heritage institutions (libraries, archives and museums) have integrated technology into all aspects of their mission and services. The first part of this chapter looks at these developments, and introduces case studies illustrating the wide range of reasons that an institution might consider digitization of its collections. The second part of the chapter will examine some of the new economic challenges and service paradigms associated with digital collections.

The potential of digitization

The libraries, museums and archives of the world are filled with materials recorded in many 'analogue' formats. These include paper and all its variants, for example vellum, papyrus, birch bark, wood and other substrates. Images can be represented on paper or canvas, as well as many surrogate forms including negatives, glass plates, and

microfilm and microfiche. Sound and moving image have been stored on film, videotape, audiocassette and LP records. Despite this variety of formats and playback devices with which it is associated, analogue information has three consistent qualities. Firstly, it is tied to a physical medium, meaning that analogue content is linear, bounded and fixed (Delany and Landow, 1994). Secondly, it is temporal, or bound to a sequential representation that is pre-determined by the author. Finally, it degrades when copied.

Digitization is the process by which analogue content is converted into a sequence of 1s and 0s and put into a binary code to be readable by a computer. Digital information also has common characteristics and qualities, regardless of whether the content is stored on DVD, CD-ROM or other digital storage media: it can be linked to other materials to create multimedia; it is not dependent upon spatial or temporal barriers, or hierarchies; it can be stored and delivered in a variety of ways; and can be copied limitless times without degradation of the original. Digital data can be compressed for storage, meaning that enormous amounts of analogue content can be stored on a computer drive, or on a CD-ROM. Digital content can be browsed easily, and can be searched, indexed or collated instantly. Most importantly, it can be linked to a whole 'web' of other content, either locally or globally via the internet.

The expansion of global computer networks and high-speed access to the internet has led to a proliferation of digital content, delivered to increasing numbers of computer users worldwide. There is a growing demand for immediate access to rich content and easily accessed, up-to-date information from news and media organizations. The development of 'digital libraries', a concept also known as the electronic library, the virtual library and the library without walls (Raitt, 2000), has preceded and anticipated much of this demand. Much of this development was anticipated by the work of visionary thinkers such as Vannevar Bush, articulated in his 1945 essay, 'As We May Think', where he famously posited the 'Memex' machine:

Consider a future device for individual use, which is a sort of mechanized private file and library. It needs a name, and, to coin one at random, 'memex' will do. A memex is a device in which an individual stores all his

books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.

(Bush, 1945)

The history of computing since Bush anticipated the notion of the scholar having access to infinite quantities of information at the desktop is one of rapid technological advances. These have led to a sea change in the accessibility, affordability and ease of use of computing and networked digital information. From mainframe computers of the 1940s, which were costly, labour intensive and maintained centrally by large organizations, via the introduction of micro and mini computers in the 1970s and 1980s, to the development of improved, inexpensive processors and memory which influenced the personal computing revolution of the 1990s, these changes have dramatically affected the way we live and work. In addition, access to networked computers, the internet, interactive materials and multimedia have created a technological infrastructure which has caught the popular imagination. These technological developments, and their rapid uptake by a large community of technology users, have underpinned the development of ‘Digital Collections’ and what we have come to call ‘the digital library’. This is defined by the Digital Library Federation as follows:

Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities.

(Greenstein, 2000)

Digitization in libraries, archives and museums

The use of technology has become a core part of the institutional mission of museums, archives and libraries around the world. Computer-based systems are now considered essential for many operational aspects of such Memory Institutions. These include collections management, as in the use of administrative databases and online cata-

logues; exhibit planning, including the management of loaned objects such as administering paperwork for insurance and transit; and user services and outreach, including the provision of online catalogues and reference materials, as well as public service websites with general information about mission, collections and services.

In addition to the use of technology for administrative purposes, more institutions are unleashing the 'added value' of their collections by developing digitization initiatives. Collections can be made accessible, via digital surrogates, in an enhanced format that allows searching and browsing, to both traditional and new audiences via the internet. Institutions of all sizes have seen such services multiply since the development of the world wide web in 1989. Consequently, many have become 'hybrid institutions', with a mission to manage both analogue and digital cultural resources, and to support and anticipate the demands of their patrons for both traditional and new resources. However, the dichotomy of preserving access to the resources such as the traditional card catalogue for some users, while also providing access to high-resolution images of key collection items and managing digital assets, is straining resources at some institutions (W. Arms, 2000).

There has also been significant growth of various national and international digitization projects in the last ten years, as libraries and universities all around the world have funded major initiatives to showcase their rich cultural and scientific heritage. Early pioneers included the Library of Congress in the USA (<http://lcweb2.loc.gov/>), the Bibliothèque Nationale de France (www.bnf.fr/), and the British Library (www.bl.uk/). The critical role that digitization plays in cultural heritage initiatives was recognized in the European Union's eEurope 2002 Action Plan (European Commission, 2000), aimed at stimulating European initiatives to realize opportunities created by the advent of digital technologies, and summarized by DigiCULT (Digital Heritage and Cultural Content) as endorsing the view that:

Digitisation contributes to the conservation and preservation of heritage and scientific resources; it creates new educational opportunities; it can be used to encourage tourism; and it provides ways of improving access by the citizen to their patrimony.

(DigiCULT, 2003)

It is easy to find similar testimonials to the potential of digitization from other sources around the world, from which it is clear that there are enormous benefits to be reaped by both the custodians and users of cultural heritage materials by the free delivery of cultural heritage collections at the click of a mouse. However, such statements are not the hollow pronouncements and promises of ten or 15 years ago, when early experimentation with desktop technologies and remotely accessible materials for instruction and research gave senior administrators in libraries and universities, as well as funding agencies and government departments, ideas that new technology would save millions of hours of teaching time and increase academic productivity, based on the assumption that a CD-ROM of a term's coursework could replace instructors and face-to-face classes. Such claims raised expectations unreasonably, and many enthusiastic 'early adopters' of digital technologies discovered at great expense that there are hidden costs and pitfalls to developing and using digital content. However, thanks to a period of extensive trial and error, experimentation and testing, a critical mass of digital content has been developed over the last two decades. This content, and the extensive experience of the practitioners and experts responsible for its creation, provides us with a valuable understanding of the digitization process, and its costs and benefits. This wealth of experience will realistically inform future project development, and provide information managers with the ability to assess accurately the potential of digitization for their collection, institution and patrons.

The most important lesson learned is probably that there are no short-term cost savings to be realized by digitizing collections. Such initiatives may save money in the long term, but start-up costs are not to be underestimated. Furthermore, technology has a short life cycle, which means expenditure in replacing systems after (an average of) three years, as well as significant investment required for staff to learn the latest systems and applications, which usually have a steep learning curve. Dealing with technologies that have such a short life cycle also means that the 'long term' – and the demand to see savings and returns on an initial investment – may come around sooner than anticipated. There can also be a problem with the available technologies. Systems developments are generally market led, not led by the needs of scholarship and research. Generic applications developed for business are

often all that is available (unless an expensive custom system is commissioned), and this can create frustrations with limitations of the technology. More significantly, the proliferation of digital data, coupled with the short life cycle of technology, has created a preservation problem for the future (discussed in the section on preservation in Chapter 7). There is also a concern about presenting access to a surrogate copy of the original, which can never truly be a satisfactory substitute for the artifact itself. The concerns that critics like Robert Hughes have expressed about slides and reproductions, which ‘destroy the sense of uniqueness and scale of the originals, and their physical presence’, are equally applicable to digital images, which are that they are simply: ‘an image of an image, not the thing itself but a bright phantasm, a visual parody whose relation to the original and actual work of art is the same as that of a shrunken head to the human being’ (Hughes, 1992). The question is further complicated by the question of the authenticity of digital data – we know that digital data can be manipulated, copied and altered with ease. How can such content ever be an acceptable substitute for the ‘real’ materials? How can institutions ensure that patrons understand the electronic materials they see have not been in any way manipulated – that they are seeing what the custodian of the originals deems to be a true representation of the original (A. Smith, 1999)? Most importantly, the lesson learned from earlier projects is that institutions must not neglect other activities when allocating resources for the establishment and maintenance of digitization services; the impact of a digitization programme on the institution’s other public service activities must be considered as a factor in informed decision making, and in keeping in perspective the investment made.

Advantages of digitization

In recent years, a growing understanding of the costs of digitization, in terms of both time and financial resources, has placed a greater focus on developing digitization initiatives and programmes that will realize tangible and strategic benefits for the institution and its users, rather than opportunistic or short term projects that are limited in their scope or focus. Consequently, it has been necessary to articulate clearly the concrete benefits of running digitization projects at the

outset. The best way to do this is to focus on developing resources that push the boundaries of what is possible in research or access by placing a focus on not merely transforming ‘pen to pixels’, but on developing projects that support the type of work that cannot be done in an analogue format. Digitization is a complex process, and there are concrete benefits to be realized from many types of digitization projects. These can be summarized as access, support of preservation activities, collections development, institutional and strategic benefits, and research and education. These themes are outlined in more detail below.

Access: broader and enhanced, to a wider community

The primary, and usually the most obvious, advantage of digitization is that it enables greater access to collections of all types. All manner of material can be digitized and delivered in electronic form, and the focus of the content that is selected for digitization varies across institutions. Some institutions have followed a policy of creating an electronic image of every item in their collection and placing it on their website. The National Gallery in London is one organization that has done so (www.nationalgallery.org.uk/). Other institutions, such as the British Library (www.bl.uk/), have chosen to put only the ‘greatest hits’ of their collections online. Another approach is to collect electronic images based around exhibition themes, or educational modules, and the Metropolitan Museum of Art in New York (www.metmuseum.org/) is among the organizations that have chosen this option.

Digital materials can be made available to a broader audience than those who have the resources or ability to travel to see the analogue collections, and access can be expanded to non-traditional audiences such as lifelong learners. Audiences can access the collections for often unanticipated and broad-ranging research interests – for example, historical materials may be used for local history or genealogical research, which has been one of the main attractions of the digitized records of the National Archives and Records Administration (www.archives.gov/). Activists and advocacy groups may access audio recordings of US Supreme Court proceedings, which are available via the Oyez project, developed by a professor of political science at

Northwestern University (www.oyez.org). The Gertrude Bell Archive at the University of Newcastle (www.gerty.ncl.ac.uk/) found that its collection of maps and photographs of areas around the borders of Iraq, as mapped out by Miss Bell in the 1920s, may have been of tremendous interest to a whole new audience in the spring of 2003 (Buchan, 2003).

Whatever the audience, their access to the materials is enhanced by the advantages of the digital format. With the application of the right technological tools, and careful attention to the design of the user interface, it is possible to search, browse and compare materials in useful and creative ways. Patrons may scroll or browse through thumbnails of the materials in image catalogues, including images of materials that were previously inaccessible, such as glass plate negatives, or oversized or fragile materials. Digital images or texts can be integrated with, and linked to, other materials, to provide an 'enriched' archive of materials. Examples of this approach include the Blake and Rossetti Archives at the University of Virginia's Institute for Advanced Technology in the Humanities (IATH; <http://jefferson.village.virginia.edu>). Both integrate searchable collections of images, texts, commentaries and glossary materials, as well as advanced imaging applications to 'zoom in' on manuscript images.

Access can be provided to materials in all formats. The National Gallery of the Spoken Word (NGSW; www.ngsw.org/), a collaborative project based at Michigan State University, is creating a significant, fully searchable online database of spoken word collections spanning the 20th century, and will be the first large-scale repository of its kind. NGSW provides storage for these digital holdings and public exhibit 'space' for the collections. These include the Vincent Voice Archive, recordings of the spoken word and sounds, originally collected by G. Robert Vincent, who began recording voices in 1912 at the home of US President Theodore Roosevelt. He went on to amass the largest private collection of recordings of voices, believing that there was no substitute for hearing the actual voice – which can transmit meaning and inflections that cannot be conveyed by the written word. When he retired in 1962, he donated the recordings to MSU. He also donated his time and assisted in cataloguing the entire collection, meaning that the recordings have accurate and detailed catalogue entries. The col-

lection houses taped speeches, performances, lectures, interviews, broadcasts, etc. by over 50,000 people from all walks of life, from Abbott and Costello to Graf Ferdinand von Zeppelin.

Another example of online access to multimedia resources for the remote user is The Experience Music Project in Seattle (www.emplive.com/), a collection of materials promoting and illustrating the history of popular music. Materials from the museum's collection are presented alongside interactive, audiovisual tools, interviews with contemporary musicians, a sound-lab, and an ever-changing selection of content from the permanent collection.

Supporting preservation

Developing a digital surrogate of a rare or fragile original object can provide access to users while preventing the original from damage by handling or display. This was the motivation behind the digitization of many priceless artifacts, most famously the Beowulf Manuscript at the British Library which is too fragile for use or consultation by scholars without special permission. The Library carried out high-resolution imaging of the original, which created digital images that can be subject to advanced imaging analysis including ultra-violet and x-ray photography. This has had the dual benefit of increasing scholarly understanding of the original while protecting the original. The multi-site Making of America project was similarly inspired, in their case by making digital copies of brittle copies of 19th-century journals accessible online. This is a common motivation for digitization. Often, the fragile condition of collections prevents their use. Digitization is not a substitute for traditional preservation microfilming, however. The digital format is too unstable, and issues related to the long-term preservation of digital media have not yet been resolved (see Chapter 8 on rare and fragile materials and the section in Chapter 7 on preservation for more on these topics).

Collections development

The provision of digital materials can overcome gaps in existing collections. Primarily, there is an opportunity for collaborative digitiza-

tion initiatives to allow the re-unification of disparate collections. It is often the case that materials that were originally part of a complete collection are now held in far-flung locations, and there is a growing desire to present at least a 'virtual' sense of what the entire collection would look like. Many projects have been motivated by the goal of virtually 're-unifying' such materials.

One example is the Arnamagnaean Institute (AMI; www.hum.ku.dk/ami/aminst.html) at the University of Copenhagen. This project is making a web-accessible catalogue of medieval Icelandic manuscripts, and proposes to use this catalogue to achieve a 'virtual reunification' of the two halves of the Arnamagnaean collection, which is now divided between Reykjavik and Copenhagen. The AMI is also planning a full digitization of all manuscripts in its possession, and the catalogue records will link to these images as they become available (Driscoll, 1998).

Similarly, the Canterbury Tales project (www.cta.dmu.ac.uk/projects/ctp/) plans to develop CD-ROMs containing digital images, and transcriptions, of all extant manuscripts of the books of the *Canterbury Tales*, regardless of where the original resides. This will facilitate a unique comparative analysis and collation of the *Tales*. In addition to the advantages of seeing the folio pages in comparison with each other, the texts can be searched, browsed and collated to examine different usages of words in the different manuscripts.

Digitization is also a means of creating resources that can be re-purposed for unforeseen uses in the future. Changing research trends may alter the demand for items in a collection: the development of new fields of study (such as the study of popular culture) means that collections once perceived as ephemeral, or of low research value are now heavily researched. Similarly, collections of items that were once in high demand are now banished to offsite storage for lack of use (Price and Smith, 2000). Ephemeral materials – including magazines, pamphlets, badges and the like – may also be fragile, so digitization is especially advantageous for maintaining access to such materials. For example, at the University of Bournemouth the library and the University's media studies department are starting a project to digitize and make accessible their copies of *TV Times*, a guide to independent television programmes in the UK from 1956 to 1985. A magazine that was once perceived as a disposable weekly purchase to

help households plan their television viewing selections is now a valuable record – in some cases the only remaining record – of programmes, cast lists and production information. It is sometimes the only record of particular programmes made by independent television in the UK (see www.bournemouth.ac.uk/library).

Furthermore, libraries are increasingly under pressure to provide access to materials in response to user requests, and are transitioning policies from collecting material ‘just in case’ someone will need it, to one of developing relationships which allow the library to deliver material from elsewhere ‘just in time’ to answer a user’s needs. Providing access to digital material from many sources and places can facilitate this shift to on-demand delivery (Deegan and Tanner, 2002).

Institutional and strategic benefits

There is no doubt that digitization programmes can raise the profile of an institution. Projects to digitize priceless national treasures or valuable scholarly materials, if done well, can bring prestige to the whole institution. Raising the profile of an organization by showcasing digital collections can be a useful public relations exercise. Digital collections can also be used as leverage with benefactors and funders by demonstrating an institutional commitment to education, access and scholarship. Certain funding opportunities exist for digitization, and it may be expedient for an institution to use them as an opportunity to accelerate a digitization programme (this is discussed in more detail in Chapter 6). Internally, there can be benefits in several areas. Access to digital catalogues improves collections management in general, by creating detailed records about the collections. Online catalogues also provide detailed information about collections to users, or even by including browsable digital images in alignment with the catalogue entries. By thus enhancing services there may even be a reduction in costs of certain types, for example, delivering heavily used materials such as short loan collections online.

Developing digital projects can have long-term benefits for the institution, although it may take many years to realize these benefits fully. Such initiatives may create an opportunity for investment in the technological infrastructure, and can create an opportunity to develop the overall technological skills base among staff. Staff themselves will bene-

fit from access to digitization programs that give them an opportunity to learn about new technologies. If managed correctly, internal digitization units can provide a tremendous opportunity for staff development. One institution that is now realizing such benefits is the New York Public Library (NYPL; <http://digital.nypl.org/>), where an initiative established to support digital projects is now providing programmatic support for the whole organization. NYPL's Digital Libraries Program was developed to support the NYPL Visual Archive (which was formerly known as ImageGate). The project dealt with over 600,000 images from all four research collections of the NYPL, including many different types of visual materials, such as printed ephemera, maps, postcards and woodprints. In order to support this undertaking, major investments in staff, technology and infrastructure were made. In particular, a team of almost 30 staff was developed, covering a broad scope of expertise in all aspects of digitization and technology infrastructure, including databases, web publishing and high-resolution imaging, as well as metadata and library standards. Now that the team is in place and fully equipped, and has completed some of the earlier projects, they are able to support additional projects and initiatives for the whole institution (Bickner, 2003).

Many funding opportunities are contingent on collaborations and partnerships between several institutions, so this can be an excellent opportunity to develop strategic liaisons with other institutions. Such initiatives are often developed under the auspices of a national digital library programme. For example Denmark's Electronic Research Library (DEF; www.deff.dk) is creating a portal for Danish research libraries. This will provide access to all the information resources managed by the individual libraries via a national infrastructure, with a common user interface and access system, enabling cross searching of all collections. This is a major undertaking, but it has led to a great deal of investment in the infrastructure of Danish research libraries, and the technological upgrading of library systems. Added benefits will include the negotiation and acquisition of 'national licences' for electronic journals and information databases; the provision of funding for the digitization of selected collections; a retro-conversion of paper-based catalogues; and development of the Danish Research Database and initiatives for electronic publishing.

Research and education

Digitization of cultural heritage materials can have tremendous benefits for education. Many institutions present educational ‘modules’ on their websites, presenting ‘packages’ of educational material based around their collections. Museums have been particularly successful in this respect, as most organizations have in-house educational departments, which have been charged with developing materials that will exploit the potential of technology for delivering educational resources to all levels of learners. The Hunterian Museum at the University of Glasgow boasts that its digital collections are used by schoolchildren ‘from Barra to Brooklyn’ (www.hunterian.gla.ac.uk/). The New Museum of Contemporary Art’s Virtual Knowledge Project (www.newmuseum.org/) is an outreach programme that facilitates online discussions between museum staff, artists and schoolchildren around themes of contemporary art. Similarly, the Minneapolis Institute of Arts (www.artsmia.org) has put digital images of 5000 works from their collection online (out of 100,000 objects in the whole museum). These are organized thematically to allow in-depth study of key ideas and concepts, such as ‘modernism’ and ‘myths and legends in art’, using items from the museum’s collection to develop teaching packs. This sort of outreach has become an essential way for many museums to fulfil their obligation of ‘public education’ in many parts of the USA, where a combination of budget cuts in school districts and security concerns have all but ended school visits to museums in many urban areas.

The advantages to academic research and advanced scholarship are equally impressive, and the potential of networked technologies to create a dynamic reading and scholarly environment is driving digitization initiatives at many institutions. John Unsworth has posited that networked digital information can support the fundamental elements of scholarship, the ‘scholarly primitives’, which he suggests are the ability to do the following with research materials: discovering, annotating, comparing, referring, sampling, illustrating and representing. These activities are basic to scholarship and common to all eras, disciplines and media. All are activities that can be enhanced considerably in scholarship that is based on digital information, and in particular, networked digital information (Unsworth, 2000). While

the fundamental aspects of scholarly methodologies are still in place, there are assumptions that digital materials can be 'read' in new and creative ways, and that because of this, production and delivery paradigms for scholarly materials are shifting. No one model of electronic delivery is definitive; indeed, the nature of the format allows many representational models for different types of information, data and content. Both publishers and academics are starting to think about new ways to represent scholarly information. Digital library systems, which customize information upfront and create a dynamic reading/browsing/studying environment, can facilitate these goals, and also develop new and shifting paradigms in the relationship between scholars, users, publishers, cultural institutions and libraries. These changes in relationships work on many levels. The user is able to engage with the source materials in what has become known as an 'enriched' fashion: it is possible to not just read text or view an image on the screen, but to browse, search, annotate and compare materials. Digital collections offer flexible and interactive access to the materials, and enable new scholarly imperatives.

Another example of the potential to change the essentials of scholarship is the Chopin 1st Editions Project, based at Royal Holloway College, University of London. This project is developing an online variorum edition of Chopin's work, and is using this to analyse the creative history of Chopin's music. The variorum could also be used by performers to create their own editions by combining elements from a range of different sources.

Digitization can also be the first step in conducting advanced research on historical materials. Ancient documents present a prime candidate for digitization because of their historical import, combined with centuries of exposure and degradation. At the Rochester Institute of Technology, an important site for research into the digitization of ancient documents has emerged in a collaborative project between the Xerox Digital Imaging Technology Center and the Chester F. Carlson Center for Imaging Science. Their primary mission has been an effort to enhance and clarify ancient writings, with a particular emphasis on the Dead Sea Scrolls (www.cis.rit.edu/research/dir.shtml). This project has developed a purpose-built imaging software and digital camera station. Electronic sensors and digital image processing are combined

to permit multispectral analysis. Multiple digital images of a single scroll are recorded at different wavelengths of light. The images are recorded by an electronic camera, which converts the light intensity in each section of the image into an electrical signal to be read by a computer. To aid in capturing the different wavelength ranges, coloured glass filters are placed over the camera before making the exposure. After the images are gathered, they are processed with software developed by the Xerox Corporation. The software permits the images to be analysed and combined in different ways. In many instances, this two-part technique of imaging and processing has revealed characters no longer recognizable to the human eye, granting translation scholars access to material not seen in thousands of years. The project has also conducted research with other fragile ancient documents, written on clay, papyrus or vellum. Some of the material consists of long scrolls, while other material consists of small pieces of documents, often numbering in the thousands. High-resolution scans are made and then manipulated by a variety of applications, including histogram and threshold adjustments, combined with hue and saturation manipulations following the initial scan. Experimentation of this nature is revealing ways in which advanced digital imaging, and digital cameras capable of reading a spectrum from the ultra-violet to the infra-red, can reveal characters in the otherwise unreadable manuscripts, increasing the overall accuracy of translation and interpretation.

Integration of technology: a case study of an incremental approach

Although there are many reasons to adopt computer technology in cultural heritage institutions, no one reason will predominate, and it is important to emphasize that most institutions will integrate many different technology-based projects over a long period of time. Some of these projects will overlap, some may ultimately contribute to an institutional ‘digital library’, while others may become known as ‘legacy projects’, leaving preservation concerns and headaches for future caretakers. Certain priorities will take precedence at different stages in an institution’s history, and these initiatives may or may not be consistent with what technology is available at the time. Conse-

quently, it is instructive to look at the history of digitization at one organization to see that reasons for digitization can be pragmatic and can change over time to adapt to funding and other considerations.

Established in the early 19th century, the UK's National Gallery of Art (www.nationalgallery.org.uk/) in London contains over 2000 works, including some of the most important European paintings in the world. Artists such as Botticelli, da Vinci, Titian, Rembrandt, Monet, Renoir and Van Gogh are well represented within the collection. To assist with its various conservation efforts, the National Gallery established a Scientific Department in the mid-20th century. The Department has since become an important site for conservation research, and more recently, the home of the Gallery's digitization efforts. Over the last decade, projects have included the development of scanning and photographic equipment capable of highly accurate colour images, as well as a colour separation system, which can print the images on a conventional four-colour press.

Initially, digitization efforts at the National Gallery were implemented to create archival colour records of paintings within the collection. These records could then be used for regularized comparison, often five or ten times a year, to monitor deleterious change within the works, particularly light-induced changes in pigment. With this goal in mind, the Scientific Department implemented the VASARI project in the late 1980s, a system for acquiring high-resolution digital images to facilitate a surface analysis of paintings. The system included a high-resolution monochrome camera, an accurate positioning system, a light projector containing a set of filters, image-processing software and a workstation. By the early 1990s, the Scientific Department had become interested in moving from mere digital acquisition to publication, resulting in a project known as MARC (at the National Gallery, MARC is an acronym for Methodology for Art Reproduction in Colour, and has nothing to do with the MARC standard for MACHine-Readable Cataloguing!). The primary results of the MARC project were the creation of a digital camera more portable than the system used for the earlier VASARI project, and the development of a colour separation system for four-colour printing. Yet by the mid-1990s, network access to the images generated by the VASARI and MARC processes, often over one gigabyte

each, remained unrealized. Since that time, the Department has been conducting research into various file formats and worked to develop a network image viewer and central indexing system. Ultimately, standard JPEG and TIFF formats were selected.

In addition to the aforementioned series of projects, the National Gallery has realized the potential for systematic digitization and is currently creating digital surrogates of its entire collection. The digital images will be incorporated into a larger database of the Gallery's entire holdings, which can be used to record and manage the collection by curators, conservators and scientists alike. However, the Gallery's digitization efforts are not solely aimed at an internal audience. In the early 1990s, the Gallery was one of the first galleries to have computers for public use, in the Micro Gallery of the Sainsbury Wing, as well as making their collections available on CD-ROM. In the summer of 2000, every painting within the National Gallery's permanent collection was made available on the web and in 2002, the Scientific Department collaborated with a private firm to develop an innovative image enhancement technique for visitors to the site. The technique, resulting in high-definition scans which may be zoomed in on minute details, allows viewers extremely close access to prominent paintings within the Gallery's collection (including, in the summer and autumn of 2003, a beautiful representation of Raphael's 'The Madonna of the Pinks', prominently displayed on the Museum's home page). To deter copyright infringement, a discreet logo is embedded within each of the images. Eventually, this technique will be available for paintings throughout the permanent collection (for more information, see www.nationalgallery.org.uk/about/press/2003/zoom.htm).

The impact of digital collections on institutions

The development of digital collections and the proliferation of such content through the global 'information explosion' (Gill and Miller, 2002) are changing the way that information is used and managed. The 'digital library', the 'online archive' and what Martha Wilson of Franklin Furnace has called the 'desktop museum' (see www.franklin-furnace.org/ and Wilson, 2001), are enabling new paradigms for schol-

arship and access. In order to capitalize on these developments, new strategic visions and economic models are emerging, as administrators start to examine the way that digital collections can be managed and funded for the long term. The challenges of the digital age are moving memory institutions into new business models, and developing institutional enterprises around digitization. However, this transition is not from one static and identifiable paradigm to another static paradigm (S. Smith, 1998). Instead, the rapidly changing technology is facilitating a period of experimentation and evaluation of new models for scholarship and access, and an examination of new funding models.

Those who are developing and managing the technology do so in the hope that new technologies will enable the extension of the reach of research and education, an improvement in the quality of learning, and new methods of scholarly communication (A. Smith, 1999). Digital collections have enormous potential for changing the way that information is used, and for developing new ways of preserving, collecting, organizing, propagating and accessing knowledge (Witten and Bainbridge, 2003). At many institutions, electronic 'spaces' are seen as a resource to augment learning. As more and more libraries devote space to computer terminals, and museums develop kiosk systems, the physical presence of technology in memory organizations cannot be ignored. At the University of Hertfordshire, a new Learning Resource Centre provides a large space for computers in the library, equating and integrating student and faculty computing needs with information needs. A similar space is being built at Glasgow Caledonian University, incorporating library collections, computers and teaching space. These developments suggest that administrators have seen that the creation of a digital library or online archive enables the creation of new space even if the institution cannot buy any more physical space.

At one extreme, this idea has led to the notion of the 'virtual campus', the idea that the physical campus is no longer required when 'learners' can have access to all the content they need via an electronic library. Institutions such as the University of Phoenix – an entirely virtual campus offering extended education modules – attracted an enormous amount of attention in the mid 1990s (not coincidentally, also the years of the dot.com boom and bust), and many administrators were beguiled by the prospect that universities

and libraries could package and sell academic content (their ‘product’) via online teaching materials. While there is a tremendous amount of potential for distance education, the reality is that the technology available at this time doesn’t fully support such initiatives, and that the business plans of many such initiatives overestimated the market for such resources. The failure of Fathom.com, a high-profile for-profit online education initiative based at Columbia University but incorporating a number of prestigious partners including the British Library, is one such example indicating that reports of the demise of the physical campus were much exaggerated.

Nonetheless, there have been significant changes in the delivery of scholarly content, shifts in the relationship between content creators and users, and shifting paradigms in the ‘delivery chain’ of published materials. Notably, we see a shift from the traditional model of a publisher creating material, which is bought by a library and then distributed to users. Now, there are many different delivery models, such as from publisher to service provider (such as JSTOR) who creates an aggregate resource to which libraries subscribe. Users of the library are then able to access this material. This model raises questions about the provision of long-term access to such resources, as we see a movement away from the system of libraries purchasing, storing and preserving books and journals on paper (Guthrie, 2001). There are also changes in the relationships between scholars, users, publishers and cultural institutions and libraries (Deegan and Tanner, 2002). For example, scholars are investigating whether self-archiving of their research (including making available pre- and post-print publications) might resolve some of the difficulties associated with academic publishing. These archives could be published on their own websites, maintained by their employing university (Oppenheim, 2002).

Observation of such developments indicates that there is a role for a carefully managed institutional repository of electronic information that allows active engagement with electronic resources. Though faculty, librarians, archivists and curators all create electronic content that can become part of a digital library, it will only be through developing an understanding of how to properly manage this content that the economic potential of electronic information will truly be realized and understood. Many institutions would

like to change the current paradigm in which they pay faculty to create scholarly content, which is then given to publishers and then sold back to the university through journal subscriptions. Experimentation with the concept of an 'institutional repository' attempts to address this issue.

It is also important to understand the difference between an 'institutional repository' and a digital library. An institutional repository seeks to exploit the intellectual capital produced by the institution and therefore 'owned' by it. A digital library, on the other hand, is a broader collection of not just these materials, but materials published elsewhere and licensed and distributed to users of the library. It is an aggregated and accumulated system, allowing access to interconnecting information created at many different locations and in many different media types. This information is subject to different interpretations, classifications and purposes (the core elements of the 'scholarly primitives' outlined above), which should be supported by the underlying infrastructure of the digital library. One approach to developing such an infrastructure is the Open Archival Information System (OAIS; <http://ssdoo.gsfc.nasa.gov/nost/isoas/>), a conceptual framework for an archival system that can be adapted and expanded to preserve and maintain access to digital information over the long term. Many library standards organizations (including RLG and OCLC) are looking at ways in which the OAIS model might be adapted for a digital library environment, and the relevance of the OAIS model is discussed extensively in Deegan and Tanner (2002).

There are a number of ongoing initiatives developing tools and architectures for institutional repositories and digital libraries, notably MIT's DSpace (www.dspace.org/) and the FEDORA (Flexible Extensible Digital Object Repository System) Project (www.fedora.org), a collaboration between the Universities of Virginia and Cornell. DSpace is a digital repository, created to capture, distribute and preserve the intellectual output of MIT (and organizations that are involved in the DSpace partnership) by providing stable long-term storage for digital content in a secure preservation environment and repository which is accessed via an easy-to-use interface for faculty depositing the materials. The FEDORA Project is creating a repository management system with an extensible architecture for manag-

ing the digital content so that it can be re-used and re-purposed for many interpretations.

Such initiatives raise a number of important questions. If libraries and other institutions are digitizing content and making it available to mass audiences, are they becoming more like publishers? What are the economic implications of this, and how does this affect research and culture? And who should pay for these initiatives? DSpace is presently supported by MIT's core library budget, as well as by charges to users of 'premium' services offered by the repository (such as metadata creation) besides external grant support and in-kind support from members of the DSpace federation who are participating in the development. This raises the question of what an institution can charge for this sort of repository service. There is little information available on what the market will actually bear in terms of paying for such services. These kinds of models require strong institutional support, leadership, and business and operational planning operated in parallel with the research and development process to build the system, not after it has been created, when it might be too late (Barton and Walker, 2003).

New economic models

New economic models are emerging as digitization initiatives develop at various organizations. What are the economics of having services on the desktop that, until very recently, could only be obtained by physically going into a library? What is the cost to the library of offering this sort of service online at no charge to the user? And is there a saving to the institution now that they no longer have to provide the traditional services (Lesk, 2003)? Such questions are beginning to affect some of the ways we think about digitization, as we try to resolve the question of how we can pay for digital collections. Presently there are several possible sources of funding and revenue for digital projects, including:

- institutional subscriptions
- individual sales
- outside grant support

- institutional support from the host institution
- revenue generation, for example by the provision of digitization services.

These models are based on the development of business practices from the print environment. In addition, most models are based on the considerations of particular collections, and the funding structures of individual institutions – there is no one size, or model, that will fit all conditions (Wittenberg, 2003).

Cost savings: indirect costs

However, the more significant question is how to actually realize the dividend from our investment in digitization. In examining this question, it is necessary to look at the indirect costs of digitization, and to examine ways in which cost savings might be turned into revenue. This involves examining some of the institutional practices and logistics associated with acquiring, storing and delivering electronic information, and looking at potential savings created by electronic storage, access and circulation (Lesk, 1996). In both the digital library and the institutional repository many cost models and potential sources of revenue, including advertising and direct taxes, have been investigated. Some of these are discussed by Michael Lesk in his essay ‘How to Pay for Digital Libraries’ (Lesk, 2002). Lesk concludes that no one model of funding dominates. We see a mix of models: free distribution; institutional funding; and some sales and subscriptions.

Electronic journals are an example of the shifting paradigms in delivery of resources. Libraries now ‘rent’, rather than purchase, serials. The costs of renting versus buying journals are very different. Costs related to buying serials include the cost of storing, shelving, retrieving and cataloguing the materials, as well as costs related to the physical storage of the content: the costs of building libraries, the cost of power for heat, light and air conditioning, which are a direct cost to the library. The shift to renting electronic content has reduced the costs of maintaining the physical materials, but has increased the cost of preserving the content. Who is paying or is willing to pay to preserve this digitized information (Guthrie, 2001)? It may take less

space to store collections electronically, but how can these kinds of savings be captured? For example, buying JSTOR and other electronic journals will save library shelf space, but will this saving on space be so large that it will only be necessary to build a new library in 12 years, not ten? Furthermore, many institutions continue to maintain the paper publications as well as subscribing to the electronic serials, and indeed publishers will often require that an institution purchases a paper version of the journal in order to qualify for a discounted rate on the electronic journal. Buying the electronic journal alone is often a more expensive option.

Costs such as storage come out of different parts of the overall budget, and are 'indirect'. They are monitored at the most senior administrative level (such as the vice-chancellor in the UK or the provost in the USA). As such, these costs are rarely seen, let alone able to be truly accessed by libraries. Universities, for instance, often fail to recognize or directly charge departments for all indirect costs (e.g. most university libraries don't pay rent for their building) and so a library may not realize, in deciding whether or not to buy an electronic publication in place of a paper one, the extent of shelving and cataloguing costs that are saved by going electronic. These issues are tied up with complex questions on the 'value' of information, making it almost impossible to put a numerical value on delivering information to the desktop instead of the library reading room (Lesk, 2003). There is a lack of real figures on which to base these assumptions, as we don't have enough experience with these resources and funding models to develop properly predictive figures. An additional complication is that technology and network costs have decreased dramatically in the last 20 years, making comparative calculations relating to the cost of digitization over a long period of time almost impossible. The savings that we see at present are also aggregate, that is, they are shared by a large number of institutions. Collectively, this could add up to a significant figure – but individually, the sums involved probably do not yet offset the cost of digitization. This is one reason why the Library of Congress digitization initiatives do not focus on the large-scale conversion of books. It is better to focus on the conversion of unique materials that would otherwise have limited use.

If these savings could be captured, they could provide a significant fund for further digitization. But in order to evaluate what these savings might be, it is necessary to take account of all elements of the financial equation, including the long-term implications for building plans, capital costs and maintenance. Few institutions think in such terms, preferring to see digitization as merely another competitor for inclusion in an already strained acquisitions budget. But this is not the way decisions of this kind should be made. In the digital world, a broader institutional perspective needs to be applied to resource allocation decisions, and to evaluate how this revenue can be quantified. The larger academic community needs to work together to realize the economies of scale that are possible. It is also necessary to look at added value benefits (such as user satisfaction or the advancement of scholarship) and work out a methodology for putting a value on them at some level, in order to gain an understanding of the true benefits, both financial and scholarly, of digitization (Waters, 2003).

Cost savings: widening the evaluation

We now suspect that digital resources *should* be creating cost savings for institutions (especially libraries) at some part of the digital life cycle. Proving this is another matter. In order to quantify this sort of revenue, more research is needed on the economics of hidden costs. For example, the trend towards using public domain materials for digitized courseware saves payment of copyright fees to authors and publishers for 'course packs'. Other costs that could be re-allocated to digitization might include resources such as travel grants awarded to scholars and PhD students to visit large research collections and archives. For example, if the series of medieval judicial materials at the National Archive in London (including Common Pleas, King's Bench, Ancient Indictments and Gaol Delivery Rolls) could be digitized, how many scholars would not have to seek research travel grants to work on these materials? (See Byrd et al., 2001, for a quantification of savings to the organization by the use of online patron access.) Similarly, there will be an overall saving to the institution if digitization eliminates or reduces curatorial and librarianship costs (W. Arms, 2000), and such cost savings could be explored to develop

digitization funding. Again, it is extremely difficult to quantify the sums involved in this type of saving, to unravel from whose budget it is coming, or to fully understand how such savings can be exploited.

Developing a critical mass of digital content may enable savings elsewhere in the institution by, for example, reducing the hours that a reserve or short-loan collection needs to be open, reducing the time spent re-shelving bound journals, but taken to its logical conclusion, this line of argument about savings of library staff time and reducing salary lines could be at the cost of redundancies for librarians (W. Arms, 2000). This isn't practical at any level, especially as we know from experience that even if librarians' time is saved, it is just moved to other tasks – such as developing training programmes on how to use electronic resources.

It is also important to realize that the costs of digitization are just beginning at the time of starting digitization projects:

The programmatic capacity to distribute and maintain electronic resources, and to migrate them to new forms as original digital platforms fail and formats and software are superseded, is fundamental to long-term efforts . . . rising user expectations may require that existing digital files be reprocessed in new ways. When OCR software is perfected, for example, unsearchable bitmap images of texts could be thought unsatisfactory. Projects that do not plan for change may become obsolete, and therefore irrelevant.

(Hazen, Horrell and Merrill-Oldham, 1998)

Nonetheless, some institutions are looking to cost recovery models based on potential savings as digitization replaces and improves some existing services. Especially when they are feeling the strain of having to support both analogue and digital resources with the same number of staff and with the same budgets as in previous years (literally in many cases – during the present US economic crisis, many institutions have had budgets frozen). Consequently, some digitization funds are being diverted away from other collections-based activities, and may even be taken from budgets dedicated to acquisitions. This is a strategy that will not pay off for the institution, unless funds are diverted from existing activities based upon a strategic approach and an assessment of where

technology is actually saving money. Using this approach, we can examine several activities and services, including the following:

- Transitioning from analogue to digital photography has seen real savings at many institutions as photographic order backlogs are cleared.
- As buildings and spaces within the institution are refurbished to include 'wired' classrooms and meeting spaces, it is no longer necessary to pay for 'Campus Media' service organizations – projector rentals, staff to set up equipment on an event-by-event basis, etc.

In addition, emerging initiatives and ways in which institutions are going digital may, in the long run, realize some savings in terms of library staff and space:

- As informational websites become the norm for most institutions, and online access becomes the preferred delivery method for certain collections, there may be some savings to staff time as they find themselves having to deal with fewer face-to-face queries. However, library staff invariably report that any anticipated time savings are instead spent addressing questions regarding electronic resources. Also, as noted elsewhere, digital initiatives are a wonderful advertisement for the institution, and may increase requests from users to see the analogue resources.
- The development of electronic reserve, or short-loan, collections will certainly be a great service to library users, as will electronic access to past examination papers. Such services may realize cost savings at some institutions as the library hours, or staff time, needed to administer such popular and labour-intensive collections are reduced.
- Some institutions may realize savings from other forms of publication or distribution. In some library contexts (as in many business contexts), some simple substitutions may provide new revenue sources. For example many non-profit organizations offer certification programmes as a major source of income (for example, many library schools in the United States offer certification for public librarians). Such operations can replace the post-

ing of print documentation and print test materials with web-based documents and tests, and the savings in postage and administration can pay for the whole online operation.

The models outlined above, and the qualifications associated with each, illustrate that it is still premature to anticipate digital ‘cost recovery’ in existing service areas; because many real costs are hidden, it can be difficult to see such opportunities for what they are – so it is not economically advisable to invest in such initiatives with the expectation of cost saving, although this may be an agreeable outcome. Above all, resources should not be diverted from some service or acquisitions area to start a digitization programme, unless the decision can be defended in the terms set out above.

Another approach is to consider if there are hitherto untapped funds that can be used for digitization, such as developing approaches to leverage college tuition fees for digitization. One area of exploration is providing continuing access to digitized content provided by a university as an ongoing benefit for its alumni as part of their tuition fees. This would scale especially well for professional education, such as medicine and the sciences, by giving alumni access to new research. Other ideas come from organizations like Digital Promise (www.digital-promise.org/), a US lobbying agency trying to have funds from the sale of unused, publicly owned telecommunications activities (mandated by Congress) allocated to a national Digital Opportunity Investment Trust. Whatever the future sources for digitization, whether tuition or government grants for technology development funds, digitization will be far easier to sustain if there is a guaranteed revenue for the long-term preservation of and continuing access to these programmes.

Conclusion

Digitization of cultural heritage materials is changing the ways in which collections are used and accessed. Many materials are amenable to digitization, including scarce, fragile and ephemeral materials, as well as the whole spectrum of moving image and audio materials. All can be safely used by a wider audience in digital form. Research and interrogative tools for digitized source materials can also make digital surrogates more

amenable to certain types of interpretation, such as full-text searching and indexing, as well as comparison of materials for multiple sources. Nonetheless, there will always be times in which no digital surrogate will be adequate for scholarship, and it will be important to be able to evaluate whether or not digitization is truly worthwhile before undertaking a digitization initiative (Nichols and Smith, 2001). Many factors will come into play when evaluating the 'value' of digital resources, but these factors may help in assessing when digitizing collections can be cost effective. Valuable digital resources, which will bring prestige to the institutions that create and maintain them, will be those that can support scholarship without any loss of the benefits of working with the originals.

With no definitive evidence base to give concrete numbers about the economic value of digitization to an institution, assessing the value of digital resources is a question of also assessing whether digitization is also causing information to 'lose' some of its value: for example, what is the loss to scholarship if electronic resources cannot be browsed in the same way as conventional library stacks? In a recent presentation, Michael Lesk gave a compelling example of the value of information, and of the 'serendipity of the stacks', which should be preserved in the digital library, in telling the story of Sir Alexander Fleming and the lucky discovery in the library (by a browsing scholar) that led to the discovery of penicillin:

Fleming (a doctor) first discovered that some substance from the mould *Penicillium* killed bacteria in 1928, and wrote a paper about the substance, hoping for help from a biochemist. But little happened for over a decade. Prompted by the Second World War to look for antibacterial agents, Sir Ernst Chain, a researcher at Oxford, found Fleming's 10-year-old paper in the British Journal of Experimental Pathology. This discovery in the stacks led Chain and Lord Howard Florey to test and then exploit the first modern antibiotic, to the great benefit of medicine and humanity; Chain, Florey, and Fleming shared the 1945 Nobel Prize.

(Lesk, 2003).

This describes the type of research and discovery that should be replicated in the digital collection, and which will ensure that digital collections have value to all users in a digital future.