1

Introduction

1.1 Introduction

Picture a scene: in a county record office somewhere in England, a young archivist is looking through the morning post. Among the usual enquiry letters and payments for copies of documents is a mysterious padded envelope. Opening it reveals five floppy disks of various sizes, accompanied by a brief covering letter from the office manager of a long-established local business, explaining that the contents had been discovered during a recent office refurbishment; since the record office has previously acquired the historic paper records of the company, perhaps these would also be of interest? The disks themselves bear only terse labels, such as 'Minutes, 1988-90' or 'customers.dbf'. Some, the archivist recognizes as being 3.5″ disks, while the larger ones seem vaguely familiar from a digital preservation seminar she attended during her training. On one point she is certain: the office PCs are not capable of reading any of them. How can she discover what is actually on the disks, and whether they contain important business records or junk? And even if they do prove of archival interest, what should the record office actually do with them?

Meanwhile, a university librarian in the mid-west USA attends a faculty meeting to discuss the burgeoning institutional repository. Introduced a few years ago to store PDF copies of academic preprints and postprints, there is increasing demand from staff to store other kinds of content in a much wider range of formats, from original research data, to student dissertations and theses, teaching materials and course notes, and to make that content available for reuse by others in novel ways. How, the librarian ponders, does the repository need to be adapted to meet these new requirements, and what must the library do to ensure the long-term preservation of such a diverse digital collection?

Finally, in East Africa, a national archivist has just finished reading a report from a consultant commissioned to advise on requirements for preserving electronic records. The latest in a series of projects to develop records management within government, he knows that this work is crucial to promoting transparency, empowering citizens by providing them with access to reliable information, reducing corruption and improving governance through the use of new technologies. The national archives has achieved much in recent years, putting in place strong records management processes and guidance. But how to develop the digital preservation systems necessary to achieve the report's ambitious recommendations, with limited budgets and staff skills, and an unreliable IT infrastructure?

This book is intended to help these people, and the countless other information managers and curators around the world who are wrestling with the challenges of preserving digital data, to answer these questions. If I had been writing it only a few years ago, my first task would have been to explain the need for digital preservation at length, illustrated no doubt with celebrated examples of data loss such as the BBC *Domesday* disks, or NASA's Viking probe.[1] Today, most information management professionals are all too aware of the fact that, without active intervention, digital information is subject to rapid and catastrophic loss – the warnings of an impending 'Digital Dark Ages' have served their purpose. Hopefully, they are equally alive to the enormous benefits of digital preservation, in unlocking the current and long-term value of that information. Instead, their principal concern now is how to respond in a practical way to these challenges. There is a sense that awareness of the solutions has not kept pace with appreciation of the potential and the problems.

Such solutions as are widely known are generally seen as being the preserve of major institutions – the national libraries and archives – with multi-million pound budgets and large numbers of staff at their disposal. Even if reality often doesn't match this perception – many national memory institutions are tackling digital preservation on a comparative shoestring – there is no doubt that such organizations have been at the vanguard of developments in the field.

The challenges can sometimes appear overpowering. The extraordinary growth in the creation of digital information is often described using rather frightening or negative analogies, such as the 'digital deluge' or 'data tsunami'. These certainly reflect the common anxieties that information curators and consumers have about their abilities to manage these gargantuan volumes of data, and to find and understand the information they need within. These concerns are compounded by a similarly overwhelming wave of information generated by the digital preservation community: no one with any exposure to the field can have escaped a certain sense of despair at ever keeping up to date with the constant stream of reports, conferences, blogs, wikis, projects and tweets.

In writing this book, my goal has been to demonstrate that, in reality, it is not only possible but eminently realistic for organizations of all sizes to put digital preservation into practice, even with very limited resources and existing knowledge. I have sought to do so through a combination of practical guidance, and case studies which reinforce that guidance, illustrating how it has already been successfully applied in the real world.

1.2 Who is this book for?

This book is intended to be of value to anyone with an interest in the practice of digital preservation, but is primarily aimed at existing and prospective practitioners in:

- **smaller memory institutions**, such as libraries, archives, museums and galleries, which have a core mission to collect, preserve and provide access to information or artefacts
- **institutional archives and libraries**, which collect, preserve and provide access to the information resources created or used by their organizations in support of their core mission; examples include business archives and institutional repositories.

In other words, it is written for the vast range of organizations outside the national cultural memory institutions that want and need to develop the ability to collect, preserve and provide access to digital information. Although it should be of interest to policy makers within these organizations, it is intended primarily for those who are, or are hoping to be, responsible for digital preservation at a practical level.

The underlying aim of digital preservation can be stated very simply:

> To maintain the object of preservation for as long as required, in a form which is authentic, and accessible to users.

This book shows how you can build practical solutions to achieve that aim. It begins by looking at how to approach developing a digital preservation capability, from raising initial awareness, and gaining the necessary mandate and resources, to beginning an organized programme of work to put in place the appropriate people, systems and processes. It then examines in detail what the practice of digital preservation actually involves, from initially acquiring content to making it available to users. It should not be assumed that this requires monolithic IT systems; one of the central arguments of this book is that digital preservation is an *outcome*, which can be achieved by many different means, and at varying levels of complexity, to suit the needs and resources of the organization in question.

1.3 Minimum requirements

The entry level for digital preservation is actually very low – indeed, the premise of this book is that it is entirely realistic for small organizations to implement credible services. However, it must be recognized that there are minimum requirements for an organization to build a digital preservation service. These are:

- **Motivation**: First and foremost, an organization must have the desire to address the digital preservation challenge. Doing so is likely to be a lengthy process, by turns as frustrating as it is rewarding, and a substantial level of motivation is essential to persevere through this.
- **Means**: Second, an organization must have the wherewithal to turn that desire into reality. This may take the form of:
    - □ **expertise**: to establish the detailed case for digital preservation, define the organization's requirements, and oversee their implementation and future operation
    - □ **financial resources**: to fund staff, services and infrastructure
    - □ **infrastructure:** to underpin an operational digital preservation capability.

Of the three, either expertise or financial resources are the most critical: expertise can make best use of limited resources and help to secure more resources in future, while money can be used to buy in expertise. The minimum infrastructure required is very variable but, as will be demonstrated later in this book, should be within the reach of most organizations.

1.4 Some digital preservation myths

There are a number of widespread myths and misconceptions about digital preservation, which together serve to foster the image that it is too scary, complex and difficult to be contemplated as a practical proposition by smaller organizations. In particular, it is often perceived that digital preservation:

- can only be tackled by national bodies
- requires huge budgets
- requires deep technical knowledge
- can be left until next year to tackle.

This book serves to counter those myths with some digital preservation realities:

Digital preservation can only be tackled by national bodies

While such institutions have undoubtedly taken the lead in developing digital preservation as a discipline, the existence of mature, affordable, practical tools and services means that it is now not only realistic, but also imperative, for organizations of every size and type to address the issue.

Digital preservation requires huge budgets

You can spend as much or as little on digital preservation as resources allow. While the US National Archives and Records Administration has spent an estimated $500 million on building its Electronic Records Archive,[2] a working digital repository was developed at the English Heritage Centre for Archaeology at the cost of a few hundred pounds and the author's time (see Chapter 4, 'Models for implementing a digital preservation service'). This book demonstrates how much can be achieved using readily available tools and resources, as well as with more complex systems.

Digital preservation requires deep technical knowledge

While it can undoubtedly lead into very technical territory, especially at the cutting edge of research, digital preservation practice does not require deep technical knowledge. Practitioners today come from hugely varied backgrounds, ranging from traditional library and archives roles and IT, to astronomy and archaeology. Adaptability and enthusiasm are the most important characteristics for any would-be digital archivist. While most have developed their skills on the job, there are now an excellent range of training opportunities to suit all needs and budgets, from online tutorials, through seminars and conferences, to longer training courses and postgraduate qualifications. Digital preservation is also becoming established as a vital professional skill within information management training courses. Couple this with a very supportive and collaboration-minded community, and no one should have cause to fear that digital preservation skills are inaccessible or difficult to acquire. The opportunities for training and professional development are discussed in detail in Chapter 4, 'Models for implementing a digital preservation service'.

Digital preservation can be left until next year to tackle

This is an issue that organizations need to address urgently, if they are to realize the enormous benefits, and avoid substantial legal, financial, operational and reputational risks, as well as the loss of information of great historical and business value. This is not to say that you must do everything at once, or that your requirements will be the same as another organization's – the maturity model introduced in Chapter 4, 'Models for implementing a digital preservation service', and expanded in Chapter 8, 'Preserving digital objects', illustrates how you can develop your capabilities over time, and to a level that suits your needs. However, now is the time to begin tackling digital preservation at a practical level.

1.5 The current situation

So what challenges do small organizations currently face? A survey in 2008 provided an interesting snapshot of the state of readiness across local authority archives in the UK to preserve digital records.[3] There is little to suggest that the situation has changed greatly since, and it is worth looking at the results of this survey in some detail, as they illustrate the challenges faced globally by smaller organizations in general.

Although most archives demonstrated a basic awareness of digital preservation, and knew (74%) about basic sources of support such as the Digital Preservation Coalition, the level of more detailed knowledge dropped off very noticeably beyond that. Around half were aware of the seminal international standard, the Open Archival Information Systems (OAIS) Reference Model, and of key initiatives run by national memory institutions such as the British Library and The National Archives (TNA). Two-thirds were unaware of other key international standards, such as PREMIS or METS, and a similar proportion were unfamiliar with projects of particular relevance to UK local archives, such as the East of England Digital Archive Regional Pilot[4] and Paradigm.[5]

Nearly half (47%) had a digital preservation policy, which conforms to the findings of other surveys before and since (see Chapter 2, 'Making the case for digital preservation'). However, relatively few had taken the next step of introducing detailed standards and working practices, such as guidelines for depositors (16%) or ingest procedures (11%).

Most archives (79%) considered themselves to be reacting to the demands of depositors, rather than proactively building their digital records capability, although almost all held some digital material, and only 5% were actually turning away digital records because of a lack of facilities. Despite their nascent digital collections, they frequently lacked even basic information about the nature of that material, such as detailed volumes or file counts. The information supplied by respondents about the file formats they held is illuminating: in addition to the ubiquitous image formats resulting from digitization initiatives, and the expected Office-type formats, there was a wide range of obsolete formats, such as Lotus 1-2-3 and Claris Filemaker, as well as specialized formats such as computer-aided design (CAD) and genealogy data. Many archives also reported holding digital audiovisual collections. Although unsurprising, given the wide-ranging collecting policies of many local authority archives, this diversity highlights some significant preservation challenges. As a result of fairly minimal information gathering activities at ingest, most archives did not have the information necessary to undertake any form of preservation planning.

The majority had some form of backed-up, server-based storage, although 87% also had some material on optical media such as CD or DVD; 42% simply stored the data on its original media, although around half did at least perform basic checks on ingest, such as testing whether the media could be read. Only a tiny proportion was undertaking more sophisticated actions, such as generating checksums or normalizing formats. Only one respondent had use of a content management system, and one was outsourcing its storage.

Access is a fundamental requirement for any archive, but two-thirds of respondents were relying on purely *ad hoc* arrangements, rather than any formal user access system. Such online delivery facilities as did exist were mainly limited to image galleries, and therefore did not support access to other types of digital material.

Interestingly, less than half of respondents reported close involvement in the implementation of electronic records management systems, even though these are likely to be one of the principal sources of digital records for such archives in the future.

A particularly noteworthy aspect of the survey was the section on barriers to digital preservation, in which respondents were asked about the main perceived obstacles. The report identified three groups of these from the responses: cultural, resource and skills. Perhaps unsurprisingly, funding was seen as the main barrier, followed jointly by IT support and skills, then political support. On the other hand, staff motivation, leadership, time and strategic partnerships were all seen as less significant barriers. While one should be cautious about drawing too many conclusions from this, it does suggest that costs and skills are at least perceived as the major obstacles – the spirit is willing but the funding is weak.

Those respondents who suggested how these barriers might be overcome were most concerned with gaining institutional buy-in, and developing and embedding policies and procedures. These essential steps are discussed in detail in Chapter 2, 'Making the case for digital preservation'.

More detailed questions about the skills gap yielded a range of development requirements, from generic management and IT skills to very specific digital preservation knowledge. These highlight the importance of access to practical and affordable training.

Another key issue highlighted was the disconnect between archivists and information and communications technology (ICT) support services, with relationships in some cases being described as poor or antagonistic. Allied to a lack of budget provision for, and experience of managing, major IT projects, this means that although most archives have access to ICT support services, including developer resources, few are in a position to take advantage of these facilities to develop digital preservation capabilities.

There was remarkably little consensus among archivists when asked what their preferences were for providing digital preservation services in future. Although an in-house repository or regional consortium was preferred by the greatest number of respondents, almost as many ranked the in-house solution their least favoured. The only point of consensus was a general rejection of outsourcing to a commercial provider, although it was unclear whether this was motivated primarily by perceived budget constraints, a paucity of plausible commercial services, or as a point of principle.

So what can we conclude about the situation faced by smaller organizations today? First, the main barriers to developing digital preservation capabilities are practical – money, skills, leveraging available resources – rather than the more fundamental obstacles of awareness and will, although the latter may still apply to parent bodies and other funders.

Second, most organizations have some of the basic building blocks of a capability in place, and are not allowing the lack of a more comprehensive capability to stop them from beginning to collect digital material. While such an approach needs to be taken cautiously – it would be irresponsible to accept material that one is fundamentally unequipped to preserve – it must also be encouraged: trying to develop a complete and perfected solution in one step can only lead to disappointment, and practical experience is essential for learning.

1.6 A very brief history of digital preservation

Interest in the longevity of digital information and curatorial approaches to its management have been evident since the early years of the digital information age, and can be traced back at least to the 1960s, when the first data archives were established. Designed to manage scientific research data, and make it accessible to the scholarly community, archives such as the Inter-University Consortium for Political and Social Research (1962)[6] and UK Data Archive (1967)[7] laid much of the groundwork for digital curation as we know it today.

The advent of personal computers and the internet triggered an explosion in the creation and use of digital information, which started in earnest in the early 1980s, gained enormous momentum in the 1990s as a result of the emergence of the world wide web, and continues unabated to the present day. Suddenly, the world was producing a plethora of new types of digital information – from office documents to multimedia, web pages to 3D models, e-mails to e-books – in hitherto unimaginable quantities. Digital information had moved from being the preserve of big business and major research institutions to a fact of everyday life for billions of people.

Concerns about the fragility of digital information crystallized in the formation in 1994 of the Task Force on Archiving of Digital Information. After two years of deliberation, this US group published a seminal report,[8] which laid the foundations for most subsequent work in the field, and continues to shape the agenda even today. Concepts and concerns such as certification of trusted digital repositories, format registries, cost models, and integrity and authenticity – which this report first articulated as a coherent set of challenges – remain the focus of daily discussion within the digital preservation community today, at conference, in blogs and on Twitter.

This should not be taken to indicate that the discipline has failed to progress since the report was published, or to find answers to the searching questions that it posed. Far from it: digital preservation today is the focus of an enormously vibrant, active and collaborative community. Indeed, it is instructive to look briefly at how far that community has come in such a short period of time. In 1996, when I first began developing a digital archiving programme at the English Heritage Centre for Archaeology,[9] it was possible to assemble and read virtually everything of note written on the subject in a few pages of bibliography;[10] 16 years later, even maintaining awareness of developments in the field is a constant challenge, reading all their published outputs an impossibility.

As with any emerging discipline, two strands of activity are required to progress: the development of strong theoretical underpinnings and standards, and the establishment of a diverse and active pool of practitioners, who can advance and expand the theory through practical application.

The publication of the OAIS Reference Model has proved to be one of the seminal moments in the development of a coherent conceptual framework for digital preservation. Originally developed by the space science community in the 1990s, and released as a draft recommendation by the Consultative Committee for Space Data Systems in 1996, it rapidly became accepted as a *de facto* standard. It was formally published as a full recommendation in 2002, before being issued as an international standard (ISO 14721: 2003), and most recently updated in 2012.[11] It sets out a detailed model of the functions and processes required of a digital repository, as well as introducing a set of terminology that has become established as the *lingua franca* of the digital preservation community.

Another key area of standardization has been in relation to metadata. Thanks to the emergence of internationally recognized schemes such as METS (2001) and PREMIS (2003), the community is well served by a range of standards tailored to the needs of digital preservation (these are discussed in detail in Chapter 7, 'Describing digital objects').

While OAIS provides a conceptual model for what digital repositories should do, the widespread development of operational digital preservation services has led to much discussion about the detailed standards to which they should adhere in practice. From this has emerged the concept of 'trusted digital repositories'; this trend is examined in Chapter 4, 'Models for implementing a digital preservation service'.

The more practical development of the discipline has been driven equally by the efforts of individual institutions in building their own preservation solutions, and through collaborative research. Projects such as CEDARS (CURL Exemplars in Digital ARchiveS) in 1998,[12] and the Dutch Nationaal Archief's Digital Preservation Testbed (2000),[13] were highly influential, applying rigorous scientific principles to the development and testing of practical digital preservation methods.

The first major digital preservation repositories were built by national cultural memory institutions, such as the National Library of Australia (2001), the Koninklijke Bibliotheek, the National Library of the Netherlands (2002) and the UK National Archives (2003). Today, they are no longer the exclusive province of such institutions, with repositories proliferating among many other types and scales of organization, including university libraries, local archives and business archives.

This has been enabled by the emergence of production-quality digital repository systems, which provide the technological platforms on which to build digital preservation services. A number of open-source solutions have emerged, of which Fedora (1997), EPrints (2000) and DSpace (2002) are the most widely adopted examples today. In parallel, commercial products such as Safety Deposit Box (2003) and Rosetta (2008) have been brought to market, often borne out of initial funding from national memory institutions. Most recently, cloud-based services such as DuraCloud (2011) and Preservica (2012) offer a new paradigm for providing digital Preservation-as-a-Service (PraaS), which may be of particular interest to smaller organizations. These technologies and the options for building digital repositories are discussed in detail in Chapter 4, 'Models for implementing a digital preservation service', and Appendix 3.

Alongside repository software, the emergence of widely available, practical preservation tools and services such as the PRONOM technical registry (2002), JHOVE characterization tool (2003) and DROID format identification tool (2005) has played an essential role in making digital preservation a practical proposition for many organizations.

The specialized discipline of web archiving has a history almost as long as the Web itself. From the foundation of the Internet Archive (1996) and the Nordic Web Archive (1997) to the wealth of local, national and international web archiving programmes we see today, the huge volumes of data acquired have spurred the development of digital repositories capable of managing and preserving them.[14]

Since the early 2000s, many advances have come as a result of major research projects, such as those funded through the Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP) (2000),[15] and the European Commission's various research funding programmes.[16] While space does not permit a detailed account of these, projects such as ERPANET (2001), DELOS (2004), DigitalPreservationEurope (2006) and Planets (2006) have all had a huge impact on the development of the state of the art, and this momentum is being carried forward in the current crop of projects, which are discussed in Chapter 10, 'Future trends'. Similarly, NDIIPP has funded the development of major tools and services, including JHOVE2, LOCKSS and the MetaArchive.

We have also begun to see the emergence of organizations dedicated to the advancement of the discipline, such as the UK's Digital Preservation Coalition (2002)[17] and Digital Curation Centre (2004),[18] the Dutch Nationale Coalitie Digitale Duurzaamheid (2008)[19] and the international Open Planets Foundation.[20] This last, together with projects such as SPRUCE (Sustainable PReservation Using Community Engagement),[21] signals a growing movement towards the development of nationally and internationally based practitioner communities. Agile and enthusiastic, and centred more around community activities such as hackathons, rather than traditional project and institutional structures, these have the potential to advance the discipline in new and exciting ways (as discussed in Chapter 10, 'Future trends').

An excellent visual overview of the history of digital preservation, alongside key IT developments, is provided by the timeline developed by Cornell University Library, as part of its online tutorial on digital preservation management.[22]

1.7 A note on terminology

Digital preservation provides a fertile breeding ground for new terminology, as well as finding new uses for that which is established. As a young discipline, its specialist nomenclature has yet to mature and settle – in some cases, a number of alternative terms have been applied to the same, or similar, concepts. Furthermore, it bridges a number of long-established fields, each with their own unique vocabularies. All of this can appear calculated to confuse newcomers and seasoned practitioners alike. I have therefore attempted to be clear and consistent in my own use of terminology, and have provided a glossary in part to clarify the sense in which I have chosen to use it.

Two terms in particular appear constantly throughout the book, referring to the subject and means of preservation respectively: *digital object* and *digital repository*. These are so fundamental as to justify exploring them in a little more detail now.

What is a digital object?

In this book I am using the term 'digital object' to signify the thing that we are seeking to preserve, but what does this phrase really mean? It is worth taking a moment to consider the nature of these 'digital objects', to really understand what they are, and how they compare with their analogue counterparts.

Indeed, the analogue world is a good place to start. We have little difficulty identifying and understanding the nature of physical collection objects, whether they be printed books, parchment rolls or stone sculptures. Their very physicality provides a natural structure for describing and arranging them. For example, it is easy to see that there is a different relationship between the individual pages of a book and the book as a whole, as opposed to that between two different books by the same author; the volume provides a natural atomic point of reference. Of course, even in the physical world it is important to acknowledge differences in approach between the curatorial disciplines. The hierarchical nature of archival description, for example, is very different from the more discrete, unitary world of the library catalogue. However, the physical nature of the material does impose structures that are much more clearly and rigidly defined than in the digital realm.

This may not be immediately apparent. If we consider a PDF version of a paper publication, there is a straightforward one-to-one correspondence between the digital and physical object. However, this represents the simplest possible case, and conceals a frequently overlooked complication. At one level, the digital world has a very obvious and simple atomic unit: the file, but in reality the file is a purely technological artefact, having no direct relationship with the structure or nature of the information content.

This can be illustrated by considering the varying ways in which the same information object might be technically represented. We can simplify this by taking an example that is analogous to an object in the physical world – a book. An electronic book could exist in a plethora of forms (I deliberately avoid referring to 'formats', for reasons that should become apparent). There might be the author's finished 'manuscript' version, in Microsoft Word 2000 format. Depending on authorial practice, this might comprise a single Word file, or multiple Word files – one for each chapter. The Word 2000 files might subsequently be updated to Word 2007 format. This is a fundamentally different creature – each file is actually a container format, comprising a series of separate XML documents. The printed version of the book might be digitized, resulting in a set of TIFF image files, one for each page. These might then be amalgamated into a single PDF file, for ease of access. An e-book version could be created for use on devices such as the Kindle, in specialized formats such as EPUB or Amazon's KF8. Finally, we might envisage a web version of the book, where each page or chapter of the book becomes a separate web page. In this case, the book is represented as a series of HTML files, together with a range of additional files, such as cascading stylesheets and images, which are required to render the pages in a web browser. These representations are summarized in Table 1.1.

Table 1.1 Alternative representations of a book

Version	Technical representation
Physical	1 printed volume (comprising 12 chapters and 700 pages)
Word 2000	12 DOC files
Word 2007	12 DOCX files (each containing various XML files)
Digitized masters	700 TIFF files
Digitized access copy	1 PDF file
e-Book	1 EPUB container file (containing various XML, XHTML and image files)
Web	12 HTML files, 1 CSS file and 15 GIF images

We can therefore see that our digital object is much more complex and variable than its physical analogue, which has a very clear cut, discrete existence. It can comprise one or many files, in the same or different formats; it can comprise files contained within other files; even the relationship between the constituent files varies – in some cases, such as with individual Word documents for chapters, each file serves an equivalent function; in others, such as the website, a single stylesheet file might be used by every HTML file, and has a very different function.

And this represents the simpler end of the spectrum; something like a

Geographic Information System (GIS) is a very complex entity, with many component parts in sophisticated and dynamic relationships, and no real-world counterpart.

My use of the term 'digital object' therefore serves as shorthand to cut through some of this complexity. Chapter 8, 'Preserving digital objects', delves deeper into the fascinating implications of the digital information environment, and examines how we can manage these complexities through the separation of digital objects into information objects (representing the underlying entity, such as a book) and data objects (the technical components of that entity, such as files), and the use of concepts such as multiple manifestations.

What is a digital repository?

The term 'digital repository' conjures visions of vast, complex, expensive and forbidding IT systems, only viable for major institutions to consider building. This is very far from the case: as this book will demonstrate, a digital repository is a concept, capable of being realized in many different forms, to suit all levels of budget and expertise. For the purposes of this book, the following definition will be applied:

> A digital repository is a combination of people, processes, and technologies which together provide the means to capture, preserve, and provide access to digital objects.

In general, the term is therefore used in this book to refer to the body providing the digital repository function, rather than just the systems employed at a given point in time to help realize this. In the cases where it is employed in the narrower sense, this should be apparent from the context.

As previously mentioned, there is a detailed, formal definition of what is required to provide those means: the OAIS Reference Model. However, although widely cited, and undoubtedly of great value, especially in providing a common vocabulary for expressing these concepts, the complexity and terminology of OAIS can be off-putting. Fundamentally, the core functions of a digital repository are the same as any memory organization, and can be expressed very simply: it must be able to acquire control of new content, make that content available to its designated user community, and perform the various preservation and management activities required to continue doing so for as long as required. This is illustrated in Figure 1.1.

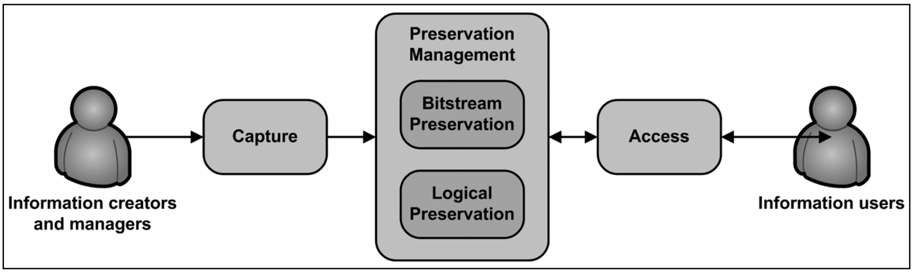

Figure 1.1 Functions of a digital repository

This book describes in detail how smaller organizations can develop the practical means to perform each of these functions, with relevant case studies throughout. It begins by looking at what is involved in building a digital preservation capability, from making the case and securing the necessary mandate and resources (Chapter 2, 'Making the case for digital preservation'), to defining your requirements (Chapter 3, 'Understanding your requirements'), and identifying an appropriate model for turning them into reality (Chapter 4, 'Models for implementing a digital preservation service'). It then examines in detail the core repository functions:

- **Capture**: A repository must have a means to capture new content, and bring it within its control. This is discussed in Chapters 5, 'Selecting and acquiring digital objects'; 6, 'Accessioning and ingesting digital objects'; and 7, 'Describing digital objects'.
- **Preservation management**: The repository must be able to manage its content so that it remains available in an accessible and authentic form. This is addressed in Chapter 8, 'Preserving digital objects'.
- **Access**: Any repository must provide a means for its users to discover and access its content. Chapter 9, 'Providing access to users', covers this.

Digital preservation is a fast-moving world, with practitioners and researchers continually evolving new ideas, techniques and tools. The final chapter therefore takes a look at how some of these may develop over the next few years (Chapter 10, 'Future trends'). Lastly, the appendices include a number of useful templates, as well as examples of a wide range of tools and services which may be of value to digital archivists, with links to further information.

1.8 Getting the most from this book

Even when focusing on smaller organizations, the diversity of resources, skills, needs and organizational contexts represented there make it very difficult to offer practical guidance useful to all: what might seem simplistic or familiar for one may be overly technical or simply irrelevant to others. I have therefore tried to provide guidance which is sufficiently detailed to provide genuine substance for the more technically minded, but which can also be dipped into by those requiring an overview. At the end of each chapter, a series of key points summarizes the main recommendations.

No single book can hope to offer a comprehensive account of such a vast and varied subject. The present volume is no exception, and claims to be no more than a starting point, an initial guide to the strange, compelling and rewarding world of digital preservation. However, it includes pointers to further information at every turn, with links to online sources wherever possible, so that readers can explore particular subjects in much greater depth, according to their inclination.

I have also included a large number of case studies throughout, for two reasons: first, I firmly believe that practical exposition is the best form of explanation, and second, I hope that demonstrating how smaller organizations of all kinds have built practical digital preservation solutions will reinforce my central thesis – digital preservation is a practical proposition for all.

1.9 Notes

1 Waller and Sharpe (2006) provide further information about these and other examples.

2 US Government Accountability Office (2010).

3 Boyle, Eveleigh and Needham (2009).

4 MLA East of England and East of England Regional Archive Council (2006) and MLA East of England (2008).

5 See www.paradigm.ac.uk/.

6 See www.icpsr.umich.edu/icpsrweb/landing.jsp.

7 See www.data-archive.ac.uk/.

8 Garrett and Waters (1996).

9 Brown (2000).

10 See, for example, the bibliography in Brown (2002a).

11 Consultative Committee on Space Data Systems (2012).

12 Two snapshots of the project website are preserved in the UK Web Archive at www.webarchive.org.uk/ukwa/target/99695/.

13 See, for example, Potter (2002) and the Testbed website, as archived by the Internet Archive at http://wayback.archive.org/web/*/http://www.digitaleduurzaamheid.nl.

14 For an overview of the history of web archiving, see Brown (2006), 8–21.

15 See www.digitalpreservation.gov/.

16 See Strodl, Petrov and Rauber (2011) for a detailed history of EC-funded digital preservation research.

17 See www.dpconline.org/.

18 See www.dcc.ac.uk/.

19 See www.ncdd.nl/en/index.php.

20 See www.openplanetsfoundation.org/.

21 See www.dpconline.org/advocacy/spruce and http://wiki.opf-labs.org/display/SPR/Home.

22 See www.dpworkshop.org/dpm-eng/timeline/popuptest.html.