

Risk Assessment 2007

British Library - Digital Preservation Team

Objective: This document is an internal working document for The British Library. It serves to help prioritise The British Library's digital content based upon qualitative and quantitative analysis of risk.

This prioritisation will determine the content that will be ingested into The British Library's Digital Object Management system. It is managed and maintained within a framework of continuous improvement and the final prioritisation will reflect this change based upon the most recent information available.

Authors


Rory McLeod, Peter Bright, Paul Wheatley

Reviewers


Helen Shenton, Adam Farquhar

| | | | |
|----------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

| | | |
|----------|---|-----------|
| 1 | EXECUTIVE SUMMARY | 4 |
| 2 | INTRODUCTION | 5 |
| 2.1 | PURPOSE | 5 |
| 3 | APPROACH..... | 5 |
| 4 | COMMUNICATE AND CONSULT..... | 7 |
| 5 | ESTABLISH THE CONTEXT..... | 8 |
| 5.1.1 | <i>Internal context</i> | <i>8</i> |
| 5.1.2 | <i>External context</i> | <i>9</i> |
| 5.1.3 | <i>Capabilities/resources</i> | <i>9</i> |
| 5.2 | STRUCTURE OF THE ANALYSIS | 9 |
| 5.3 | COLLECTION AREA SAMPLES AND THEMES | 10 |
| 5.3.1 | <i>Maps</i> | <i>10</i> |
| 5.3.2 | <i>Music.....</i> | <i>11</i> |
| 5.3.3 | <i>Language areas, Japanese, Chinese and Slavic</i> | <i>11</i> |
| 5.3.4 | <i>The International Dunhuang Project</i> | <i>12</i> |
| 5.3.5 | <i>Photography</i> | <i>13</i> |
| 5.3.6 | <i>Endangered Archives.....</i> | <i>14</i> |
| 5.3.7 | <i>Newspapers (JISC-funded mass digitisation)</i> | <i>14</i> |
| 5.3.8 | <i>Sound Archive, ASR project</i> | <i>15</i> |
| 5.3.9 | <i>Modern British and Legal Deposit</i> | <i>16</i> |
| 5.3.10 | <i>Web Archiving.....</i> | <i>17</i> |
| 5.3.11 | <i>E-Manuscripts</i> | <i>18</i> |
| 5.3.12 | <i>Science, Technology, and Medicine.....</i> | <i>19</i> |
| 5.3.13 | <i>Storage of digitised masters project</i> | <i>20</i> |
| 6 | IDENTIFY RISKS..... | 21 |
| 7 | ANALYSE RISKS..... | 21 |
| 7.1 | PHYSICAL MEDIA CHARACTERISTICS | 22 |
| 7.1.1 | <i>Physical deterioration.....</i> | <i>22</i> |
| 7.1.1.1 | <i>Optical media.....</i> | <i>22</i> |
| 7.1.1.2 | <i>Magnetic media</i> | <i>24</i> |
| 7.1.1.3 | <i>Other media</i> | <i>24</i> |

| | | | |
|--|-----------------------------|-------------|---------------|
|  | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

| | | |
|-----------|--|-----------|
| 7.1.2 | <i>Physical damage</i> | 25 |
| 7.1.2.1 | Optical media..... | 25 |
| 7.1.2.2 | Magnetic media | 26 |
| 7.1.2.3 | Other media | 26 |
| 7.1.2.4 | Summary | 26 |
| 7.1.3 | <i>Environmental damage</i> | 27 |
| 7.1.4 | <i>Technical obsolescence</i> | 28 |
| 7.1.4.1 | Optical media..... | 28 |
| 7.1.4.2 | Magnetic media | 28 |
| 7.1.4.3 | Summary | 31 |
| 7.2 | FORMAT CHARACTERISTICS | 31 |
| 7.2.1 | <i>File system</i> | 32 |
| 7.2.2 | <i>File format</i> | 33 |
| 7.3 | OPERATING ENVIRONMENT..... | 34 |
| 7.3.1 | <i>Hardware requirements</i> | 34 |
| 7.3.2 | <i>Software requirements</i> | 35 |
| 7.4 | POLICY ISSUES | 37 |
| 7.5 | BRITISH LIBRARY RISK RANKING..... | 38 |
| 8 | TREAT RISKS | 39 |
| 9 | EVALUATE RISKS AND FINAL PRIORITY TABLE | 40 |
| 10 | MONITOR AND REVIEW | 44 |
| 11 | CONCLUSION | 45 |
| 11.1.1 | <i>Initial strategies to mitigate risks</i> | 45 |
| 11.1.2 | <i>Short term tasks (by July 2008)</i> | 45 |
| 11.1.3 | <i>Medium term tasks (by 2010)</i> | 45 |

| | | | |
|--|-----------------------------|-------------|---------------|
|  | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

1 Executive Summary

The Digital Preservation Team conducted a Risk Assessment exercise in order to assess the condition of the British Library's digital collections, identify strategies to mitigate those risks, and recommend and plan actions to be taken. A risk assessment methodology based on the AS/NZS 4360:2004 standard was applied in a representative manner across the BL's collections.

The Risk Assessment concluded that the BL's digital collections face an array of risks that will require action on a number of fronts. Almost all of the collections were assessed to be at high risk.

The greatest and most imminent threat of loss is from media degradation. Failure rates for discs within the BL collections have reached unacceptably high levels (3%). Substantial quantities of digital objects are stored as single copies only, on handheld media in danger of decay. This stark warning is illustrated by many examples of disc decay that have been encountered and is backed up by the considerable evidence from external research into handheld media lifetimes.

Digital content will continue to be lost unless action is taken now. This report makes a number of specific recommendations to mitigate the highest risks facing the BL's digital collections. These include:

- Secure collections that are currently stored on handheld media as a matter of urgency.
- Perform further assessment to gain a better understanding of the media failure rates across the different collections
- Address the root causes of a number of the risks facing the collections, by streamlining and enhancing standards, check-in procedures and other policy issues

2 Introduction

This document assesses the risks that confront digital assets held in The British Library's collections. The assessment will help the Library to prioritise digital preservation work and recognise the issues faced by legacy digital assets. The assessment will also trigger actions to mitigate the risks and develop prioritised plans to address them.

2.1 Purpose

The British Library's corporate strategy commits to the long-term access and preservation of digital collections. Periodic risk assessments are an important tool for achieving this goal. Yearly reassessments will ensure that the Library can track trends in the risks and the effectiveness of mitigation steps.

A detailed and accurate risk assessment of the digital material held by the Library is an essential piece of information that underlies all digital preservation activities. Without knowledge of what items are held, what information is acquired, and what information is at risk, effective planning and execution of actions to preserve material is impossible.

In 2005, the British Library released its first findings on the analysis of digital materials¹. This first risk analysis focussed on the issue of media lifetimes. It produced qualitative assessments of the risks inherent in the media types used to store digital information. Relatively little consideration was given to the factors that determine the longevity of the digital material itself.

The 2007 risk assessment reported in this document results from work that included visits to all collection areas in order to build an accurate view of the Library's digital assets. The analysis covers physical media but also examines other factors including disk- and file-formats, and software and hardware environments.

The 2007 risk assessment establishes the context through a survey of digital assets held by the Library, identifying the physical media, disk- and file-formats, software requirements, and size, number, and significant properties of the assets. It then identifies and analyses the risks that arise in this context.

The assessment will guide digital preservation activities. For example, it will influence decisions on acquisition policy (given the choice, the Library would prefer to acquire assets that are identified as low risk), guide preservation planning actions (such as migrations from one format to another, or normalization of ingested assets into preferred low-risk formats), and highlight risk mitigation strategies.

The approach used follows the seven point AS/NZS 4360:2004 Risk Management standard, which is outlined in the following section.

3 Approach

Both the 2003 and the current 2007 risk assessments followed an approach based on international standards for risk management. The 2007 risk assessment is based on the AS/NZS 4360:2004 Risk Management standard. The 2003 risk assessment was based on the predecessor, standard, AS/NZS 4360:1999. Both of these standards are widely cited and used across a range of industries.

The AS/NZS 4360:2004 Risk Management standard defines a seven-step approach to risk management:

1 Communicate and consult

¹ <http://www.imaging.org/conferences/archiving2005/program.cfm>

| | | | |
|------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

Communicate and consult with internal and external stakeholders as appropriate at each stage of the risk management process and concerning the process as a whole.

2 Establish the context

This step sets the scene for the analysis. Stakeholders are identified, and the objectives of the stakeholders and the organization as a whole are established. If possible, measurement criteria are established so that the impact risk has on these objectives can be determined.

3 Identify the risks.

In this stage, the risks—that is, *what can go wrong*—are enumerated and described.

4 Analyze the risks

This step covers the evaluation of the impact of the risks, and the likelihood of those risks. The evaluation may be qualitative (an event may be “likely”, “unlikely”, “inevitable”, etc.) or quantitative (“a hard drive failure will occur on average once every 100,000 operational hours”), or some combination of the two.

5 Evaluate the analysis

At this stage, negligible risks might be discarded (to simplify analysis), and evaluations (especially qualitative evaluations) adjusted. The risks are compared to the objectives of the organization, allowing a ranked list of risks to be constructed.

6 Treat the risks

The options to address the risks are identified, the best option chosen, and implemented. This may include “taking no action” if no risk is sufficient. The fifth step is beyond the remit of the assessment project, so the risk analysis portion of the work will be restricted to steps one to four.

7 Monitor and review

It is necessary to monitor the effectiveness of all steps of the risk management process. This is important for continuous improvement. Risks and the effectiveness of treatment measures need to be monitored to ensure changing circumstances do not alter priorities.

The assessment also uses the impact scale devised in the DRAMBORAⁱ methodology. DRAMBORA is a toolkit and process for the risk assessment of digital repositories; although the process as a whole is not applicable to this assessment, the impact scale has proved to be useful.

The first part of the analysis was to create an inventory of the digital assets. Each collection area was visited and interviewed, and a partial audit of their digital material conducted. This provided an indicative sample of the current state of play within the Library. It is likely that continued annual updating of this list will form part of the long-term maintenance of the analysis.

This inventory includes the following information:

- Physical medium
This includes the general classification of a medium (e.g., “hard disk”, “floppy disk”, “recordable CD”), as well as specific information about the technology used. For example, over the past three decades hard disks have used half a dozen mutually incompatible electronic interfaces. In the early days of the floppy disk, different platforms used different, incompatible parameters to lay out data on the disk. These variations have significant impact on accessibility; a 20-year-old 3.5” floppy disk using PC parameters is still widely usable; a disk of similar vintage using Apple parameters would be much harder to read. A similar situation is seen today, with the multiple

| | | | |
|--------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

- writable DVD formats (DVD-RAM, DVD-R, DVD+R, DVD-RW, DVD+RW). To address long-term preservation goals, it is not sufficient to know only the general classification of the medium; these specific variations must also be recorded.
- **File-system**
Information can be recorded on disk in a variety of ways; some well-documented and widely-used (e.g. FAT16), others much less so (e.g. Acorn ADFS). A significant complication is that many file-systems, especially those used on hard disks, are proprietary and undocumented.
 - **File format**
Information in a digital item is encoded in one or more files in a variety of ways. The identification of the file format must identify these variants. This includes identification of the contents of any "container" formats; for example, AVI videos can contain sound and video content using hundreds of possible compression algorithms; merely knowing that a file is an AVI video is insufficient to assess risk or plan preservation actions.
 - **Platform requirements**
Digital items need a complex environment to access. This environment includes an application to read the item, an operating system to run the application, and hardware to run the operating system. This was the most complex area to examine. Although for pure "data" files (images, videos, etc.) there is relatively little dependence on the underlying platform, some file formats have considerable platform dependencies. For example, executable files are intimately dependent on their target operating system and processor. Even with data files, there are complications; for example, HTML files routinely depend on specific behaviours of individual web browsers in order to render correctly. Neither the HTML file nor the HTML specification is sufficient to establish the appearance and properties of a given HTML file.

The information on the holdings has come from the curators who encounter digital objects, with closer analysis and examination performed by the Digital Preservation Team. With the holdings list created, the risk assessment began.


4 Communicate and consult

Definition: Communicate and consult with internal and external stakeholders as appropriate at each stage of the risk management process and concerning the process as a whole.

In February 2007, the DPT. visited all Heads of Collection areas. They were pre-briefed with a short paper outlining the context for this work, and a presentation was given outlining the commitment and the objectives of the work.

From this initial meeting each collection Head has identified the persons best suited to play the role of stakeholder for this work.

The Collection Areas visited and the persons nominated are recorded within the wider project documentation, they are not recorded here.

| | | | |
|---|-----------------------------|-------------|---------------|
|  | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

5 Establish the Context

5.1.1 Internal context

The British Library has clearly outlined its commitment to safeguarding digital objects and to making these objects accessible. The Library's 2005-2008 strategy highlights the following points as critical to the ongoing purpose, goals and objectives responsibilities of the organisation:

- The British Library Strategy 2005-2008:
 - Strategic priority 1: Enrich the user's experience
 - Strategic priority 2: Build the digital research environment
 - Strategic priority 3: Transform search and navigation
 - Strategic priority 4: Grow and manage the national collection
- E-IS eStrategy and Information Systems strategy in full
- S&C Scholarship and Collections content strategy in full
- 10 Year Digital Preservation strategy

Additionally, the Legal Deposit Libraries Act 2003 place upon The British Library the future responsibility to maintain United Kingdom electronic legal deposit publications. This proposed extension to legal deposit to cover digital objects is being programmed so is included here as a contextual basis to be considered.


Within The British Library, there are a number of strategies that also add to the context. The e-IS strategy and the digital preservation strategy both set out clearly the responsibilities for effective stewardship of digital objects.

Ensuring the long-term accessibility of digital assets is the goal of the Digital Preservation Team. There are a number of tiers of accessibility, with each higher tier dependent on the lower tiers. Specifically:

- 1 Bit-stream preservation: The raw sequence of bits stored on a digital medium must be readable. This requires safeguarding of digital media and/or migration to more robust media as necessary.
- 2 File preservation: The bits must be interpretable as a usable digital object; this means developing or preserving suitable software/hardware to open the file, or performing migrations on the file, or some combination thereof.
- 3 Semantic preservation: The files themselves typically constitute part of a greater whole (for example, each file may represent a scanned page of a book), and to be given meaning (for example, "this is page X of book Y") requires the creation and preservation of suitable metadata. Similarly, suitable metadata must exist to allow retrieval and discovery of the objects in the first place.

In keeping with these strategic responsibilities, the recommendations from the risk assessment will take the form of

- Technical recommendations (what to do with the material we already have to safeguard it)

| | | | |
|---|-----------------------------|-------------|---------------|
|  | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

- Procedural/policy recommendations (to cover all stages of the lifecycle, from ingest through to long-term storage and preservation)
- Acquisition recommendations (given the choice the Library would prefer to acquire low-risk items)

The Digital Preservation Team has defined the depth and breadth of the risk assessment in keeping with our lifecycle approach to preservation. We will address not just the preservation aspects but also the entire lifecycle of the objects that we encounter. Due to the scale of the task it will not be possible to look at lots of objects but those that we do look at we will consider entire lifecycles in order to extrapolate all hidden technical issues.

5.1.2 External context

The methodology defines this section as addressing the business, social, regulatory, cultural, competitive, financial, and political demands placed on organization.

External stakeholders to The British Library include Department of Culture Media and Sport our parent body. In “Redefining the Library”ⁱⁱ the annual report for 2005/06 the library outlines its responsibilities these include

- Responsible to Department of Culture Media and Sport
- Other UK legal deposit libraries
- Research Community/Higher Education
- General Public

These external stakeholders expect accountability for the safekeeping of all library assets, part of this is the management of digital objects within our collections. The risk assessment is a practical example of this management.

5.1.3 Capabilities/resources

The strengths of the Digital Preservation Team are that we have a cross-directorate team, meaning we have access to both the collections and the technical resources required to undertake this work. The weakness is that we are a small team in terms of full time equivalents (FTE) and equipment. This will limit the size of the workload we are able to do but will not stop us testing the methodology or fulfilling our Lifecycle approach to preservation.

5.2 Structure of the analysis

This section contains descriptions of the material found in each collection area. These do not constitute an exhaustive audit of all collection areas, but are representative of the types of material found in each curatorial area, and hence allow the identification of the risks faced. The representative samples also serve to illustrate specific real-world problems.

From the samples, the common risk themes will be extracted and analyzed in detail in section 5. This analysis will assess both the impact of each risk factor (using the 7-point DRAMBORA scale) and where possible the likelihood of the risk factor affecting the material.

| | | | |
|--------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

The combination of impact and likelihood of these risk factors will be used to produce a ranking of the factors.

The final stage of the analysis will be the application of this risk ranking to the Library's collection, to produce a prioritization table. The risk ranking will be the primary driver of this table; secondary factors such as cost and resources will also be considered.

In section 8 and 9, we will conclude with the treatment of the risks and the plan to monitor and review the entire risk assessment.

5.3 Collection Area samples and themes


A case study approach was best for this assessment; each collection area provided representative data for analysis. For the purposes of this document, only the technical details of each collection are captured here.

5.3.1 Maps

The BL maps collection form the national collection of cartographic materials. The focal point is the Map Library, which provides access to maps, atlases, and globes of all parts of the world dating back to the 15th century. More modern material collected by this department involves a range of complex technical geospatial formats with dependencies on viewer applications and interactive software.

Maps Department Interview analysis

| Analysis | Detail |
|------------------------|--|
| Media carrier: | HDD, CDROM, DVD |
| Size of archive: | 1000 GB |
| File formats: | Raster and vector based maps and images JPG and EPS. Raw data is NTF, GML, TIFF |
| Software dependencies: | Handheld items viewable only through proprietary and customised (GIS-type) viewing applications Some database dependency Some video files Some windows 98 material already marked as "inaccessible" in the reading rooms. |
| Hardware dependencies: | British Library standard 2007 image PC is adequate |
| Technology Themes: | Limited resource to QA disks on arrival at the BL. Many disks found to be inaccessible. Lack of a central online storage area is a problem for this data, which is growing in both size and complexity. |
| Collection Themes: | Some training may be required relating to modern technology within the team. |

| | | | |
|---|-----------------------------|-------------|---------------|
|  | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

| | |
|-------------------------------|---|
| | <p>BL "digital stamp" required for ownership of files.</p> <p>Collection Management policy needs to be updated for this resource.</p> |
| Future Preservation activity: | <p>Most suited to an online management system to provide at least one level of back-up. Currently none</p> <p>Emulation rather than migration may need to be considered due to the dynamic nature of the material.</p> <p>Preservation plan needs to be written, if ingest to Digital Object Management Programme to be considered.</p> |

5.3.2 Music

The BL Music Collection includes over 100,000 pieces of manuscript music and 1.5 million separate items of printed music. It has small but growing percentage of digital data.

| Analysis | Detail |
|-------------------------------|--|
| Media carrier: | CDROM, DVD |
| Size of archive: | Maximum of 300 CD-R |
| File formats: | Standard Audio files (WAV, MP3) |
| Software dependencies: | Some database dependency Some video files |
| Hardware dependencies: | British Library standard 2007 image PC is adequate |
| Technology Themes: | Limited resource to QA disks on arrival at the BL. |
| Collection Themes: | Acquisition policy in this area could be addressed. Stricter colour management controls could be investigated to make sure that any digital scan is consistent with its siblings. |
| Future Preservation activity: | None considered |

5.3.3 Language areas, Japanese, Chinese and Slavic

The language collections of the British Library cover many different geographical areas, for this assessment we have chosen three examples and grouped them together as the technical issues faced by each are the same. The overall archive size covering these language areas is 10 terabytes.

| | | | |
|------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

| Analysis | Detail |
|-------------------------------|---|
| Media carrier: | CD-ROM, DVD plus some online content |
| Size of archive: | 4 GB |
| File formats: | PDF, TIFF, JPG |
| Software dependencies: | Non-English Operating system requirement |
| Hardware dependencies: | British Library standard 2007 image PC is adequate |
| Technology Themes: | More support required to maintain access to digital material. Both purchased and donated material is currently inaccessible through limitations of software environment |
| Collection Themes: | Policy and Procedure in this area needs to be updated. |
| Future Preservation activity: | Maintenance of non-English operating environment. Some specialist curator knowledge would be required if migration or emulation were to be considered. Priority should be to help create an online environment for this material running in its native language. |

5.3.4 The International Dunhuang Project

IDP is an international collaboration based at the British Library within the Asia, Pacific and Africa Collection (APAC) to make information and images of more than 100,000 manuscripts, paintings, textiles and artefacts from Dunhuang and other Silk Road sites freely available on the Internet. This project has a high profile international reputation and has generated a massive amount of data since its start date in 2001.

| Analysis | Detail |
|------------------------|---|
| Media carrier: | CD-ROM, DVD, HDD and RAID 5 server |
| Size of archive: | 40 TB |
| File formats: | Camera RAW (PhaseOne), TIFF, JPG |
| Software dependencies: | Proprietary camera software from PhaseOne required to access RAW files. |
| Hardware dependencies: | British Library standard 2007 image PC is adequate |

| | | | |
|------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

| | |
|-------------------------------|--|
| Technology Themes: | Lack of a central storage area to manage, store and preserve this high volume of data. There is a risk that if the project were to stop or key personnel left that this resource could be inaccessible. |
| Collection Themes: | There is a high understanding of complex technical issues within the team. IDP could play a role in educating other collection areas in the handling of data. This team could be a trainer of other collection areas. Potential legacy problem if project funding stops. |
| Future Preservation activity: | Massive amount of data, handheld disks already starting to fail based upon risk analysis. More analysis required. Recommendation is to move this resource at some future point to DOM. |

5.3.5 Photography

The Prints Drawings and Photographs Section of the British Library's Asia Pacific & Africa Collections (APAC) is one of the richest archives of images from India in the world.

| Analysis | Detail |
|-------------------------------|---|
| Media carrier: | CD-ROM, DVD, HDD |
| Size of archive: | Estimated to be 3-4 TB |
| File formats: | TIFF, JPG, some camera RAW files (various) |
| Software dependencies: | Various Camera RAW formats (Canon, Nikon etc) and proprietary image formats (e.g. Adobe Photoshop) all require non-standard proprietary software |
| Hardware dependencies: | British Library standard 2007 image PC is adequate |
| Technology Themes: | This collection is now acquiring some "original" digital content. The storage maintenance and back-up requirements therefore become more important for this collection. |
| Collection Themes: | Genuine Born digital "originals" starting to appear in this collection area but lack of a central control or storage area creates a degree of risk. |
| Future Preservation activity: | Technology watch for camera raw software changes. Recommendations to move any born digital material into DOM as a high priority. |

| | | | |
|--------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

5.3.6 Endangered Archives

As its name suggest the Endangered Archives programme has a remit to collect and archive vulnerable digital objects from a wide geographical area. It has a charitable remit sponsored by Arcadia² and is affiliated with UNESCO which has recognised the scale and urgency of the danger by establishing 'The Memory of the World' Programme to stimulate action to safeguard our documentary heritage.

| Analysis | Detail |
|-------------------------------|---|
| Media carrier: | CD-Rom, DVD, HDD |
| Size of archive: | Archive size 3-4 TB |
| File formats: | File formats- TIFF, RAW, PDF, video format/s, audio format/s, and possibly other such as excel spreadsheets and word documents. |
| Software dependencies: | Excel spreadsheets contain key project data. |
| Hardware dependencies: | British Library standard 2007 image PC is adequate |
| Technology Themes: | Lack of a central store for this data to be moved to (as it is catalogued). The archive is currently being burned to DVD-R and CD-R |
| Collection Themes: | Possible danger of introducing legacy problem if not monitored. Cataloguing and control is high quality but long-term future needs close consideration. |
| Future Preservation activity: | Deeper analysis of the content required (formats, relationships etc.). Prioritisation of the archive in anticipation of it moving into an online managed store. Some work to do for the preservation team on database preservation. |

5.3.7 Newspapers (JISC-funded mass digitisation)

British Library Newspapers is the national archive of British and overseas newspapers.

The collections, which include popular magazines and periodicals, are made available in hard copy, in microform, and on CD-ROM in the Newspaper Reading Rooms in Colindale, North West London.

This particular project of 40 terabytes is representative of the range of formats and metadata expected in large-scale digitisation projects, it should not be confused with born digital newspaper content which is not considered here.

² <http://www.arcadiatrust.org/content/splash.asp>

| | | | |
|--------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

| Analysis | Detail |
|-------------------------------|--|
| Media carrier: | LTO-3 Tape |
| Size of archive: | Estimated at 70 TB |
| File formats: | TIFF, JPG, METS, ALTO |
| Software dependencies: | none |
| Hardware dependencies: | LTO-3 tape drives British Library standard 2007 image PC is adequate |
| Technology Themes: | Massive amount of data to check, authenticate, store and manage Commitment from DOM to store this information, but timeframe still being worked out. |
| Collection Themes: | There is still a lack of standardized policy in large digitisation projects (metadata, storage and preservation.) |
| Future Preservation activity: | Close monitoring of the situation. There is a time constraint on this collection based upon its project commitments METS/ALTO are an acceptable digital preservation standard for pagination and metadata information. Technology watch has been introduced by the DPT for this format. |

5.3.8 Sound Archive, ASR project

The Sound Archive holds recordings from music, drama, and literature, to oral history and wildlife sounds and is one of the largest sound archives in the world. The ASR project is a large digitisation project for sound. It is only one of a number of current projects and was selected due to its representative nature of the wider sound archive holdings.

| Analysis | Detail |
|------------------|---|
| Media carrier: | Master files on DVD, CD and on HDD (BL servers) |
| Size of archive: | Total Sound Archive activity estimated to be 200 TB growing at over 50 TB per year. The ASR project Total Estimated at 12-15 TB |
| File formats: | Master WAV, Playback WAV, MP3, TIFF, JPG, Flash |
| Software | |

| | | | |
|--------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

| | |
|-------------------------------|--|
| dependencies: | |
| Hardware dependencies: | British Library standard 2007 image PC is adequate |
| Technology Themes: | Massive data project, plans for future ingest required. Legacy issue potentially at project end. |
| Collection Themes: | Sound Archive activity growing rapidly and producing very large archives. |
| Future Preservation activity: | Technology watch activities for sound formats, MP3 and WAV. METS files are 500K each and require investigation into what information is included. Movement of this archive recommended to an online-managed store. |

5.3.9 Modern British and Legal Deposit

Modern British is the single largest collection of books and periodicals of its kind in the world. Coverage is of all subjects in the Humanities and Social sciences in publications produced in the British Isles after 1914. This collection is also the home for UK legal deposit material that is housed at the BL.

| Analysis | Detail |
|------------------|--|
| Media carrier: | Cassette, Floppy, CDROM, DVD |
| Size of archive: | <p>Uncatalogued c.1981-c.1996</p> <p>This collection comprises 57 linear metres of hand-held e-media. 1 linear meter comprised 47 CD-R (47x 700mb x 57= 1,875,300 MB. 1.8 TB</p> <p>Catalogued, c.1982-</p> <p>10,000 items x 700 MB = 7,000,000 MB therefore 7 TB</p> <p>Handheld Offline (Voluntary Deposit)</p> <p>c.900 E-books per annum; c.1000 E-Serial Issues per annum</p> <p>Based on LIFE findings each VDEP object averaged</p> <p>Digital Donation</p> <p>c.2200 e-books per annum (1600 via email; 600 via download)</p> <p>c.33000 e-serial issues per annum (31,500 by email; 1,500 via download)</p> <p>Approx 100 PDF per annum.</p> <p>Based on LIFE findings each object thought to average the same as VDEP.</p> |

| | | | |
|--------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

| | |
|-------------------------------|---|
| File formats: | Large variety of formats from 1982-2007 |
| Software dependencies: | <p>Any number of software dependencies due to the timeframe this collection occupies (history of computing)</p> <p>There is a variety of software system requirements which include IBM, Unix and Macintosh examples of specific software are DOS 3.0; Lotus 1-2-3 version 2.0 or greater, Microsoft Excel and Windows of various levels (8 bit 16 bit etc.)</p> |
| Hardware dependencies: | <p>There is a variety of hardware system requirements, which include IBM, Unix and Macintosh also more obscure types Acorn Electron, Amiga. (here the software can be supplied by computer cassette) and Windows of various levels (8 bit 16 bit.)</p> <p>e.g.: "System requirements for computer disk: IBM AT, XT, PS/2, Macintosh, or compatible; 256K memory; 2 disk drives" (YK.1993.b.8039)</p> |
| Technology Themes: | Very broad range of material, recommendation is to focus on smaller sections rather than tackle everything. |
| Collection Themes: | <p>Decisions need to be taken quickly as to what the future holds for this collection, and our legal responsibilities.</p> <p>Born digital e-books, serials, and manuscripts should be treated as original collection items and should be subject to stricter checking controls to preserve access.</p> |
| Future Preservation activity: | <p>If kept, this is a contender for emulation technology rather than any online system or migration techniques due to its size and diversity of content.</p> <p>We do have to keep this material as it is the principle of legal deposit (including voluntary schemes).</p> <p>Born digital content needs to be considered more quickly and connections with the E-media team and acquisitions in Boston Spa should be advanced.</p> |

5.3.10 Web Archiving

Web archiving at The British Library is responsible for primarily collecting sites of research value selectively from the UK web space. Additionally though, the Library also archives a selection of sites which are representative of British cultural heritage in all its diversity across the regions. It also archives where possible a small number of sites which demonstrate web innovation.

| Analysis | Detail |
|------------------|--------------------|
| Media carrier: | HDD, some CD-ROM |
| Size of archive: | 1000 plus websites |

| | | | |
|------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

| | |
|-------------------------------|---|
| | 683 GB at time of analysis |
| File formats: | text/html; image/jpeg; application/pdf; application/msword; text/plain; text/css; text/xml; application/octet-stream; image/png; application/x-javascript; audio/x-pn-realaudio, and many others |
| Software dependencies: | Varying browser versions |
| Hardware dependencies: | British Library standard 2007 image PC is adequate |
| Technology Themes: | The UKWAC UK Web Archiving Consortium archive is currently stored outside the BL by Magus, the contractor for UKWAC Team have to use the crawler HtTrack outside of the web archiving system Pandas for large crawls, this means that there is no original copy of the gather preserved at the BL. |
| Collection Themes: | Legal deposit legislative changes could bring this collection to the top of the priorities meaning additional resource would be required. |
| Future Preservation activity: | Multitudinous formats, corrupt/invalid files, complex formats, and large volumes of data are all to be found in this collection. Preservation activity should be brought forward for this collection in our prioritisation (outside of the risk assessment). |

5.3.11 E-Manuscripts

The Manuscript Collections of the British Library form part of the national repository of manuscripts, private papers, and archives. They contain material of outstanding research importance for all periods, countries, and disciplines. E-manuscripts is an area at the forefront of change and the field of writing and personal archives includes some of the most at risk material we hold.

| Analysis | Detail |
|------------------------|--|
| Media carrier: | Paper tape, punch cards, tape, Floppy, Zip, CD, Data cartridge, HDD |
| Size of archive: | 100 GB |
| File formats: | Diverse: but include the usual such as jpegs, docs, pst, PDFs, etc. Note also that most of the eMSS collections to date originate from scientists using a variety of software and programs. |
| Software dependencies: | Originally PC, UNIX and Apple in roughly equal proportions, but working towards interoperable viewing and access |

| | | | |
|--------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

| | |
|-------------------------------|---|
| Hardware dependencies: | <p>Originally PC, UNIX and Apple in roughly equal proportions, but working towards interoperable viewing and access.</p> <p>ZIP Drives, 3.5", 5.25", 8" floppy, 0.5" Tape, 0.25" data cartridge.</p> <p>Among the most urgent for capture are the HP magnetic program cards. Expert specialist has indicated a willingness to enable their examination and duplication</p> <p>British Library standard 2007 image PC is adequate for others</p> |
| Technology Themes: | <p>A commitment required to keep (or find a provider of) older drives to retain access on an as needs basis.</p> <p>Or a commitment move this information to a secure online store. Not necessarily DOM for this data due to the curatorial specialism and confidentiality required.</p> |
| Collection Themes: | <p>There are metadata files and complementary contextual files (e.g. digital images of labels, interviews of originators digitally recorded) but the listing, cataloguing and management of these ancillary files is currently in progress and under review.</p> <p>In future, some of these files can be expected to be quite large: e.g. digital video and audio recordings created by curators.</p> |
| Future Preservation activity: | <p>Techniques for migrating data include forensic tools and procedures will need to be developed.</p> <p>Some concerns over the proprietary scientific formats.</p> |

5.3.12 Science, Technology, and Medicine

The British Library houses one of the best reference collections of current science technology and business literature in the world. This collection is unique in that it houses—within one building—science/technology literature and a collection of patents from almost every issuing authority in the world. The stock is acquired from legal deposit intake, by purchase, from donated material and by exchanges. An increasingly large range of material is collected including e-journals and databases in a variety of electronic formats, it includes a mixture of material—most of it is as accompanying CD and may contain text, computer programmes, maps, images, sound as well as the more expected scientific and technology e-journals.

| Analysis | Detail |
|------------------|--|
| Media carrier: | CD-ROMs, DVDs and 'floppies' (5¼" and 3½"). |
| Size of archive: | 25,000 items mainly CD-R Estimated at 17 TB |
| File formats: | .doc, .xls, .ppt, .rtf, .wav; and equivalent for Mac OS; and equivalent for UNIX/LINUX |

| | | | |
|--------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

| | |
|-------------------------------|---|
| Software dependencies: | MacOS 8, MacOS X, UNIX/LINUX, MS-DOS, Windows 3.11, Windows NT/2000/XP, Windows 9x Currently using Dell Optiplex GX 270 (BL standard) and G4 iMac |
| Hardware dependencies: | Older versions of PCs may be required for parts of this collection British Library standard 2007 image PC is adequate for others |
| Technology Themes: | |
| Collection Themes: | All should be the have either a (TB) or (TP) prefix. Not all do. |
| Future Preservation activity: | The material is not often accessed some of it has time expiry dates on it which may be impossible to address. We do not currently have the right technical environment/equipment for some STM items. Some specific training may be required for primary data sets |

5.3.13 Storage of digitised masters project

SDM is given here as a collection that is unique. The SDM project was started in 2002 in order to deal with high value digital projects identified as "at risk". It includes 70 separate digitisation projects. It includes a broad cross section of digitised content representative of digital projects through the nineties and early 21st century. The range crosses most collection areas and in many cases can involve highly valuable (both culturally and commercially) content. The project's vary in quality which was largely determined by the projects scope or funding so must be treated on a case by case basis to prepare for future use.

| Analysis | Detail |
|------------------------|---|
| Media carrier: | Mainly DVD, some CD, DLT, and LTO tape. |
| Size of archive: | No. of DVD-Rs burned = 1,670 No. of master files burned to DVD-R = 65,550 Volume of master files burned to DVD-R = 3,640 GB (not IDP) Total 23TB |
| File formats: | Image files, TIFF and JPG also PDF. Metadata in a number of types and quality. |
| Software dependencies: | None |
| Hardware dependencies: | British Library standard 2007 image PC is adequate LTO drives are available but perhaps not in enough numbers |

| | | | |
|------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

| | |
|-------------------------------|--|
| Technology Themes: | Media reliability- DVD have only one copy and are already starting to fail. DOM system has not been delivered in enough time for this project |
| Collection Themes: | Valuable collection but no clear owners due to project staff moving on. |
| Future Preservation activity: | Data needs to move from DVD to an online store asap. This is a high value collection of projects. Deeper analysis required of formats and relationships to metadata objects. |

Conclusion of the collection samples: At this point it was felt by the Digital Preservation Team that this cross-section of activities represented the issues faced by the collection areas and enabled us to proceed to the identification phase of the work. Further analysis of collection areas not covered and indeed the wider business areas of the BL is encouraged. Operations and Services, Marketing and Product development and even Corporate and Board records could be considered strong contenders for future risk assessment work.

6 Identify Risks

The collection area-based approach has identified a number risks to digital objects across the collections. This enables us to group together the common themes and pull out the risks in groupings in order to rank them.

The identification and analysis of the collection area material has given us around 23 numbered risks, these risks are numbered from R01 through to R023. These risks once can usefully be grouped into 23 key risks to the collections. These 23 risks are as follows




Microsoft Office
Excel Worksheet

7 Analyse Risks

The 23 categories of risk will now be assessed and scored using the impact table devised in DRAMBORA. DRAMBORA describes the impact to digital information in a numerated manner based upon the loss of understanding and/or authenticity based on a scale from zero loss through to cataclysmic loss.

| Score | Impact |
|-------|---|
| 0 | Zero impact, results in zero loss of digital object authenticity and understandability |
| 1 | Negligible impact, results in isolated but fully recoverable loss of digital object authenticity and understandability |
| 2 | Superficial impact, results in widespread but fully recoverable loss of digital object authenticity and understandability |

| | | | |
|---|-----------------------------|-------------|---------------|
|  | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

| | |
|---|--|
| 3 | Medium impact, results in total but fully recoverable loss of digital object authenticity and understandability |
| 4 | High impact, results in isolated loss, including unrecoverable loss or loss that is recoverable only by third party of digital object authenticity and understandability |
| 5 | Considerable impact, results in widespread loss, including unrecoverable loss or loss that is recoverable only by third party of digital object authenticity and understandability |
| 6 | Cataclysmic impact, results in total and unrecoverable loss or loss that is recoverable only by third party of digital object authenticity and understandability |

This scale will now be used to classify the 23 identified British Library risk groups for digital objects.

7.1 Physical media characteristics

Physical damage to media originates from three principle sources; there is deterioration due to intrinsic properties of the media (e.g. the use of unstable chemical dyes in recordable optical discs), deterioration due to handling and storage of the media (e.g. scratching of optical discs when used), and deterioration due to environmental factors (e.g. exposure of optical discs to sunlight).

Estimates on media longevity vary greatly. The mostly widely used archival and distribution media are presently optical discs (the many CD and DVD variants), and so it is optical discs that have seen the most research into expected lifetimes. This industry trend is reflected in the Library's own holdings; the bulk of the Library's optical discs are either read-only ("silver") discs or write once, read many recordable discs.

7.1.1 Physical deterioration

Physical media all undergo a certain amount of deterioration "naturally"; even if kept in ideal circumstances, their lifetimes are finite due to unavoidable decay of their components.

7.1.1.1 Optical media

Generally, read-only discs are more robust than recordable ones, though manufacturing defects can make even read-only discs short-lived. A particularly notable example is certain discs pressed between 1988 and 1993; either improper lacquer or an improper metal layer was used, with the result that audio CDs pressed at certain plants could deteriorate rapidly. Detailed analysis of pressed discsⁱⁱⁱ tends to confirm that—when correctly manufactured—they are robust and should prove reliable for decades.

Limited sampling of the Library's own content did reveal a number of unreadable read-only CDs; the nature of the problems with these discs is not known (due to a lack of equipment to perform such an analysis), and it is plausible that the discs were simply manufactured improperly; no consistent verification of ingested media is performed. For properly manufactured discs, an annual failure rate of 3% would seem reasonable.

| | | | |
|----------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

Recordable discs are less consistent than pressed discs. Recordable discs consist of a metal (silver or gold) reflective layer and an organic dye recording layer. To an extent, recordable discs are deliberately unstable; their recording layer is necessarily light sensitive, as they depend on (laser) light to allow data to be recorded in the first place. A number of different chemical dyes are used for the recordable layers, the first dyes used (cyanine dyes) were chemically unstable, and could fail within a few years. They were also highly susceptible to damage by light. Newer dyes (phthalocyanine and azo) and manufacturing techniques have greatly enhanced the chemical stability, and reduced the light sensitivity; original cyanine discs could fail after as little as a few days exposure to sunlight; azo discs should survive many weeks of sunlight without significant errors. The nature of the reflective layer is also influential; discs using gold layers tend to be more resistant to environmental mistreatment than those using silver, although silver discs have largely displaced gold from the marketplace, due to their lower cost.

Recordable DVDs use similar technology to CDs, although the use of gold reflective layers is significantly less common, and restricted to "archival" grade discs.

Unfortunately, knowing the technology used in a recordable disc is insufficient to make estimates of its robustness; it provides a ballpark estimate, but there are many proprietary variations used by manufacturers, and these all significantly influence lifetimes. High-grade "archival" discs using phthalocyanine dyes and gold reflective layers show expected lifetimes of several tens of years (in accelerated aging tests), but cheaper media can fail within two or three years^{iv}. Another general trend is that recordable CDs have lower failure rates than recordable DVDs; this is probably due to the greater data density that DVDs have; a given level of damage is more likely to cause a problem on a DVD than a CD.

Recordable discs, as with pressed discs, are also susceptible to manufacturing defects, and in addition can suffer problems from being burned improperly. Marginally out-of-specification burners, media, or readers can result in discs that are readable in some drives but not others, and interruptions during the write process can produce discs that are unreadable from the outset.

In the sampling of the Library's own media, a significant number of marginal or unreadable recordable discs were found (10 of 22 recordable discs examined). Sixth of these appeared to be minor defects incurred during burning; whether the remainder of the defects are due to improper manufacturing (or burning), or subsequent physical deterioration is impossible to tell. There are, however, examples that have clearly failed due to degradation of the media. The IDP's International Dunhuang Project's collection contains a number of CD-Rs that have suffered visible corrosion of the metal layer; a search on the Internet reveals that other users of the same brand of discs have experienced identical failures.

Given the variety of disc manufacturers and recording environments used to create the Library's recorded optical media, it is impossible to produce an accurate estimate of failure rates. It would require considerably broader sampling and more detailed analysis than has thus far been possible, along with more upfront testing to distinguish between disks that are faulty as received, and discs that have developed faults through deterioration. Assuming that the discs recorded have been at the lower end of the quality spectrum, and that the failures seen thus far are representative, an estimated failure rate of 3% per year would seem a reasonable worst-case.

| | | | |
|------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

7.1.1.2 Magnetic media

A broad variety of magnetic media is also found in the Library's collection. Life expectancies of magnetic media are much less widely studied by the industry at large, and the Digital Preservation Team presently lacks the hardware (multi-format floppy drives, tape drives, etc.) to perform any meaningful analysis of magnetic media.

The magnetic media found were representative of the full history of magnetic data storage; there are small amounts of reel-to-reel tape, proprietary magnetic cards, 3", 3.5", 5.25", and 8" floppy disks, "superflopsies" (e.g. Zip disks), sundry backup tape cartridge formats, and hard disks of all ages.

What limited industry data there is indicates that, depending on the specific technology being used, backup tapes^v have lifetimes ranging from "years" to "decades" (or even "centuries"). Anecdotal evidence would suggest that floppy disks are doing well to last anything longer than a couple of years before a high probability of failure.

Hard drive manufacturers have long cited MTBF (mean time between/before failure) for their drives; typical new hard drives have MTBFs ranging from 100,000 to upward of 1,500,000 hours (more than 150 years), although these statistics are often felt to be barely meaningful. MTBFs of that order of magnitude would suggest nominal annual failure rates of under 1%; real-world experience^{viii} suggests annual failure rates of 4% (or even more) are more likely. These figures are all for online (i.e. powered on, in-use) hard drives; little data exists for powered down drives.

Online hard drive storage has some unique points in its favour. Although a nominal failure rate of 4% would be relatively high (higher than that of high quality optical media, for example) this is mitigated by the data storage strategies used with hard drives; redundant arrays are commonplace, and these allow individual hard drive spindles to fail (and be replaced) without impacting on the integrity of the data being stored. In addition, online storage can be easily (and continuously) verified and tested to warn of impending disk failures, without the need to perform manual sampling and analysis. Thus, a properly maintained hard drive array (using some suitable storage policy such as RAID 5) has an effective annual failure rate considerably lower than its component parts.

7.1.1.3 Other media

Limited amounts of paper tape and punch cards are also held by the Library. Although obviously obsolete, these media are physically quite robust and stable. Paper will decay over time, but the tapes and cards the Library has are in essentially the same condition that they were in 40 years ago when they were created.

| | | | |
|------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

7.1.2 Physical damage

| Medium | Expected annual failure rate ³ | Impact (0 = low, 6 = high) |
|----------------------|---|-------------------------------|
| CD-ROM | 2% | 4-6 |
| CD-R | 3% | 4-6 |
| CD-RW | >3% | 4-6 |
| DVD-ROM | 1% | 4-6 |
| DVD +/-R | >3% | 4-6 |
| Floppy disk | >10% | 4-6 |
| Tape (e.g. IBM 3480) | >20% | 4-6 |
| Tape (e.g. LTO3) | 5% | 4-6 |
| Hard disk | 4% | 4-6 |
| Hard disk array | <1% | 4-6 |

In addition to the deterioration due to intrinsic properties of the media, there is also further damage caused by the use and storage of physical objects. In many areas of the Library, it was found that original media are provided to users in the reading rooms, which places those media at considerable risk. Further, the limited sampling has revealed one instance of damage due to the way that discs have been stored.


7.1.2.1 Optical media

Proper handling of optical media is essential to ensuring their longevity. Although sturdier than the vinyl records that they were designed to replace, they are nonetheless susceptible to damage through mishandling.

CDs are constructed such that the metal layer containing the data is closer to the top of the disc (the side with the label) than to the bottom (the shiny side). DVDs in contrast are symmetrical, with the metal layer in the middle of the disc. This has the somewhat counterintuitive result that damage to the shiny side of a CD is often less fatal than damage to the label side. Scratches on the shiny side of the disc are unlikely to reach as far as the data layer, and although they may prevent reading of the disc (due to making it impossible to focus the laser on the metal surface), these problems are recoverable using widely available "CD Repair" products. Conversely, damage to the label side is much more likely to reach the metal layer, and damage the data irreparably. This has unfortunate consequences; common practice for handling CDs is to place them label side down (shiny side up) to protect the shiny side.

A related issue is that occasionally sticky labels are used to identify discs. Removal of these labels can often result in removal of the lacquer coating and metal layer as well.

³ Presuming the media to be stored in good environmental conditions

| | | | |
|---|-----------------------------|-------------|---------------|
|  | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

A secondary problem with the shiny side up practice is that it exposes the light-sensitive surface of recordable media to the light. Although brief exposure is unlikely to be harmful, discs kept this way for prolonged periods (especially older discs with their less stable dye layers) are placed at great risk. Placing an old cyanine disc in the light for as little as a week could be sufficient to render it unreadable.

The scale of this problem is largely unknown; the Library does not appear to have any readily accessible usage statistics that would allow estimation of the impact of handling on the media held. Not all areas of the Library use originals in this way. There is a procedure for acquiring software and then providing it via a networked file share, which obviously greatly limits the vulnerability of media to mishandling.

Damage can also be caused by the storage. Although most of the discs examined are stored in jewel cases, there are many in plastic wallets. This was especially prevalent in the Legal Deposit collection, where the discs are found either as inserts in books or on the covers of magazines. Discs kept in this way are prone to damaging the book or magazine that they are kept with. The Sound Archive has many discs kept on spindles; this makes retrieval and use of individual discs more difficult

One example found in Science, Technology and Medicine was of a magazine cover disc kept in a plastic wallet; the wallet also contained the original paper insert describing the contents of the disc and a second paper insert with details of an ISP promotion—all as originally acquired. The print from the Internet Service Provider promotional insert had transferred onto the shiny surface of the CD, rendering the CD unusable (and as a side effect, the insert unreadable).

Again, estimation of the scale of this problem is difficult. There are likely to be many thousands of discs in plastic wallets and many more as inserts in books, and the level of damage to both books and digital media alike is hard to quantify.

7.1.2.2 Magnetic media

Similarly, magnetic media (specifically floppy disks) are also used in Reading Rooms. This leaves them susceptible to damage through proximity to magnets (e.g. in CRT monitors) and (especially in the case of 5.25" disks) physical damage to their magnetic surfaces.

Online hard disks are again worth special mention; access to material stored on such media through a network means that the media themselves are at no further risk due to usage.

7.1.2.3 Other media

Nothing worth considering at the time of analysis 2007.

7.1.2.4 Summary

There are neither reliable usage statistics nor measures of damage caused by usage. In the analogue world, we have been able to find Reading Room figures for delivery (approximately 3 million items) and a figure for preservation intervention where an item that is damaged is brought to the attention of preservation staff to view and, if required take action (260 requests per year).

In the digital world, it would be useful to try to gather these metrics using a similar method. Although requests are very low for these objects, the impact of damage is

| | | | |
|------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

arguably higher (the loss of access due to a scratch on a CD will tend to be greater than that of a torn page, for example). The risk assessment has identified the collection of usage figures for reading room digital objects as a good step forward. This data is collected in Reading Rooms although it is in a manual way and difficult to extract. It is expected that a digital preservation equivalent of the analogue system will be a future initiative from the Digital Preservation Team.

Given the available figures the risk for damage due to handling is given as

| Usage | Likelihood | Impact (0 = low, 6 = high) |
|---------------------|-------------------------------------|-------------------------------|
| Reading room access | 1 damaged item per 100,000 requests | 4-6 |

7.1.3 Environmental damage

Most digital storage media are optimally stored at 12-15 °C with a Relative Humidity of 30±5 %. Excessive heat and humidity generally tend to accelerate the degradation of storage media. On top of this general issue, there can be specific problems caused by temperature and humidity. Thermal cycling of CDs can cause catastrophic failure due to delamination; the layers of the disc can separate from each other. Further, the polycarbonate used in CDs and DVDs tends to absorb moisture from the environment, causing premature failure. Excessive humidity can also encourage mould growth, which is also harmful. No such conditions were encountered for this assessment.

Recordable CDs are damaged by light (especially UV light) and atmospheric pollutants (especially sulphur compounds). Within the Library atmospheric pollutants tend to be negligible, and light levels tend to be low throughout.

| Site | Temperature | | Humidity | | Light | Pollutants |
|-----------------------|-------------|----------------|-----------|----------------|------------|------------|
| | Specified | Measured | Specified | Measured | | |
| St Pancras (basement) | 16-19 °C | 17-21 °C | 45-60% | 43-60% | Acceptable | Acceptable |
| St Pancras (offices) | 18-22 °C | (not recorded) | 45-60% | (not recorded) | Poor | Acceptable |
| Boston Spa | 16-19 °C | 16-24 °C | 45-60% | 30-73% | Acceptable | Acceptable |
| Boston Spa (ASP) | 16 °C | | 50% | | Acceptable | Acceptable |
| Woolwich | 16-19 °C | 14-27 °C | 45-60% | 27-62% | Acceptable | Acceptable |
| Micawber Street | 18-20 °C | 14-31 °C | 35-45% | 27-68% | Acceptable | Acceptable |
| Colindale | 16-18 °C | 5-30 °C | 30-40% | 18-64% | Acceptable | Acceptable |

| | | | |
|------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

7.1.4 Technical obsolescence

Obsolescence is a long-standing problem in the world of technology. The computing industry has undergone considerable changes over its life, with older products displaced by new—and frequently incompatible—successors. The Library has suffered obsolescence issues in the analogue realm and responding to such problems is time-consuming and difficult.

7.1.4.1 Optical media

Optical media are in the fortunate position of being widely available in the marketplace and well standardized. The CD has undergone certain changes since its invention—though it started as an audio format (Red Book)⁴, extensions were made first to allow storage of data (Yellow Book), and later to allow recordable and rewritable storage (Orange Book)—but these changes have consistently been made in such a way as to retain backwards compatibility.

The CDs from the Library's collection that have been examined have all been CD-ROM or CD-R. Although occasional exceptions may be found, almost all of these CDs will be readable on contemporary equipment, and due to their prevalence we anticipated will be usable for the foreseeable future.

The situation with DVDs is somewhat less positive. Although the core read-only DVD specification (used for both video and data) is stable and widely compatible, the picture for writable DVDs is murkier. A plethora of conflicting specifications from competing industry partnerships has resulted in no fewer than five recordable DVD formats; the "standard" DVD-RAM, DVD-R and DVD-RW and the "non-standard" DVD+R and DVD+RW. Though DVD-RAM is at this point a niche player, the ± R and RW formats are all widely used. Although any current drive should be able to read all current disc variants, older drives and standalone DVD players may have difficulty reading recordable discs. Looking into the future, it does not appear that any clear winner will emerge, and that the whole set of formats may continue to co-exist.

The Library's discs examined thus far have been DVD-ROM and DVD-R, and any new or future equipment should be able to read them without difficulty. It is not known whether there is any DVD+R media; although less "standard" than DVD-R, there are reasonable technical arguments in favour of DVD+R, so acquisitions in this format are not out of the question (especially from e-manuscripts, where there is little control over the formats used).

The next generation of optical media does not appear to be enjoying the same standardization that CD and DVD have experienced. Two competing and incompatible "high definition" optical formats (HD DVD and Blu-ray) have been developed and presently no equipment is available that can fully read both competing formats. It may be that neither of these formats gains the prevalence that their predecessors have achieved, but if they do both take off, a future compatibility issue could emerge.

7.1.4.2 Magnetic media

Magnetic media are unfortunately both less standardized than optical media and have seen greater obsolescence. Contemporary computers increasingly ship with

⁴ <http://www.jhepple.com/AudioCD/redbook.htm>

| | | | |
|----------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

no facility to read any removable magnetic media, and even those that can read removable magnetic media can read nothing more than the 3.5" floppy disk.

The worst kind of material in this area is probably the kind found in e-manuscripts namely reel-to-reel tape, magnetic cards, etc. Not only are these media not used by any contemporary hardware, they are also not even manufactured any more making the sourcing of new read hardware difficult if not impossible. Moreover, they were not necessarily widely used even when new. This material is essentially unusable without considerable effort to acquire suitable reading devices, typically salvaged/reconditioned original equipment. Finding the right hardware is not the end of the problem; there may be more material than the hardware can reliably read (magnetic tape read heads, for example, can only be used for a limited number of hours before they need to be replaced). In addition, even if the material can all be read, migration to a more recent medium may still be difficult. Transferring data from a 1970s computer to a more modern machine may prove difficult due to a lack of e.g. compatible networking protocols. Old hardware may represent a dead end, onto which data can be read but not migrated elsewhere.

When the media were more widely used, the salvaged/refurbished hardware route is more appealing; for example, there are a number of vendors selling refurbished 8" floppy drives, such that they are relatively simple to acquire and use. Although the form factor of the 8" floppy remained consistent, the physical disk format varied over the life of the medium. Some 9 different variations of 8" floppy were created, with capacities ranging from 79.5 KiB to 1,184 KiB. The larger sizes were achieved by increasing the number of tracks on the disk, recording on both sides of the disk, and changing the encoding used to write data. This provides great scope for incompatibility; the many different disk variants are not readily identifiable, and a disk drive that can read some variants need not read others.

The late 1970s and 1980s were arguably the heyday of the floppy disk; during that period, first the 5.25" disk, and then the 3.5" disk, became ubiquitous. The 5.25" disk, being a direct descendent of the 8" disk, has a similar problem to that format; there are many variants of 5.25" disk, using a variety of track densities and encoding formats, and again with no ready means of identification. Newer drives may be able to read older disks, but this is by no means guaranteed; a PC-era 5.25" drive may be able to read 160 KB single-sided disks but equally may well be unable to read the original hard-sectored 98.5 KB disk.

The late 1980s and early 1990s saw the rise of the 3.5" disk. 3.5" disk drives are still manufactured, and as such represent the oldest magnetic disks that contemporary PCs are likely to be able to use. As with the older disk formats, variations in encoding mean that there will be some disks that look like "regular" 3.5" disks but cannot be read in conventional drives. The widely used Apple Macintosh (which was in fact arguably responsible for the success of the 3.5" format) wrote disks that were wholly unreadable by the 3.5" drives used in PCs; the encoding of the data on-disk cannot be understood by PC floppy controllers. This is perhaps a lesser problem than with the older formats. Although the 3.5" disk variants are more widely found amongst the Library's collection than the 5.25" and 8" variants (and so represent a more significant amount of data), their greater usage works to their advantage. The success of the Apple Macintosh, the Amiga 500, the Atari ST, and other platforms also using 3.5" variants, has meant that custom disk controllers can be purchased that allow a regular 3.5" drive to read all of these disks. These controllers also provide some facility to read the 5.25" disk variants.

| | | | |
|------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

Though the 3.5" and 5.25" disks were the most common media during the 1980s, a number of other competing formats were devised. The most significant of these was probably the 3" disk; this format found use in the widely used Amstrad PCW range of word processors/computers, and consequently small numbers of 3" disks are found in the Library's holdings.

As well as these mainstream removable media, there have been numerous attempts to provide (proprietary) alternatives to the floppy disk, typically to increase capacity. The 1980s saw the Iomega Bernoulli Box and SyQuest cartridges, and the 1990s saw closely related products such as the Zip and Jaz drives (also from Iomega) and the LS-120 from 3M. All of these pose a problem, in that the hardware is no longer manufactured (although fortunately any reconditioned hardware should be usable, as the drives typically used standard interfaces). The most notable of these is the Zip disk Introduced in 1994, these 100 MB disks were at one time expected to be the next "standard" for removable storage, and for a couple of years in the mid-1990s became quite widely used. A combination of factors (the drastic reduction in the cost of CD recorders along with questionable reliability due to manufacturing defects) saw the rapid demise of the Zip drive. Because of this (brief) period of widespread usage, the Library holds a small number of Zip (and similar) disks. In spite of the relative modernity of these formats, their brief existence in the marketplace means that acquiring suitable hardware to read them is now relatively difficult. New equipment is unavailable, leaving salvaged/reconditioned equipment as the only recourse.

Modernity thus represents no safeguard against obsolescence. The CD and 3.5" floppy are both considerably older than the Zip disk, yet both CDs and 3.5" floppies are "current" technology, and the Zip is quite obsolete.

Hard disks show a similar pattern to floppy disks, albeit with fewer variations. The earliest hard disks use encodings and interfaces incompatible with modern PCs, with the result that custom controllers (or old computers) would be needed to use them. IDE (introduced in 1986) and SCSI (also in 1986) represented a significant change in how hard drives were developed. These two interfaces (in fact, two families of interfaces) standardized the electrical and electronic interfaces used by hard drives, with the result that any IDE or SCSI disk should be usable with a modern IDE or SCSI controller. Recent versions of IDE and SCSI have seen incompatible changes made (in particular to the electrical interface), but newly manufactured compatible controllers should be available for a number of years yet.

A few other examples of obsolete magnetic media have been found that would require reclaimed equipment to use. e-manuscripts has some magnetic cards used in some HP calculating machines. Little other than the original hardware can read these (which is a problem in itself) and moving the data into a newer environment is similarly difficult due to the age of the machines themselves. Slightly more up-to-date, audio tapes containing Commodore 64 programs were found in the legal deposit collection. Although audio tape was at one time a common storage format in low-end computer equipment, it was never reliable even when new, and was largely replaced by the floppy disk. Again, there are few options to use these media aside from finding suitable original equipment.

Digital tape is still used quite widely, predominantly as a backup medium. The tapes found so far have been LTO-3 and DAT. Both of these can be read using new tape drives and contemporary PCs. The mainstream tape formats (DDS/DAT, LTO, DLT) are generally backwards compatible and still supported, and updated although DLT recently ceased development. Media in these formats should not be to much of an issue although there are limits to this; LTO only guarantees compatibility with

| | | | |
|----------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

the last two generations, for example). There have been many other tape formats in the past (e.g. Travan, QIC) and availability of suitable hardware for these formats is much lower, especially as new drives may be unable to read the earliest tapes. It is likely that the Library has some amount of data in one or more obsolete tape formats (e-manuscripts has some 0.25" data cartridges, although the particular tape technology used is not known at this time).

7.1.4.3 Summary


When considering the risk of obsolescence, media (and the hardware required to use them) fall into three categories, from lowest to highest risk:

- 1 Media that are widely available, newly manufactured, and current/contemporary
- 2 Media that are available new, but being phased out
- 3 Media that are not available new (which typically require use of salvaged/reclaimed hardware)

| Medium | Risk factor due to obsolescence (1 = low, 3 = high) | Impact (0 = low, 6 = high) |
|----------------------------------|--|-------------------------------|
| CD-ROM | 1 | 0 |
| CD-R | 1 | 0 |
| CD-RW | 1 | 0 |
| DVD-ROM | 1 | 0 |
| DVD±R | 1 | 0 |
| Floppy disk (e.g. 8", 5.25", 3") | 3 | 4-6 |
| Floppy disk (e.g. 3.5") | 2 | 3 |
| Tape (e.g. IBM 3480) | 3 | 4-6 |
| Tape (e.g. LTO3) | 1 | 0 |
| Hard disk | 1 | 0 |
| Hard disk array | 1 | 0 |
| Other magnetic disk | 2-3 | 3 |
| Paper tape | 3 | 4-6 |

7.2 Format characteristics

The physical characteristics of the media control the ability to read the stream of bits from a given object. Reading the bit-stream is only the first part of the problem; the stream must then be interpreted. That interpretation has two levels; the file system used on a physical medium must be understood, to allow extraction of individual files, and these files themselves must equally be understood.

| | | | |
|---|-----------------------------|-------------|---------------|
|  | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

7.2.1 File system

Traditionally, each operating system that has been developed has had its own corresponding file system; DOS introduced FAT, Windows NT NTFS, Linux ext2, MacOS HFS, and so on. Common file systems (in particular the FAT variants) have broad support from multiple operating systems. These broadly-used file systems represent no great technical hurdle. Many others are, however, much more limited, with first-class support typically restricted to their “native” OS. Thus, interpreting a bit stream from a disk may depend on being able to run a suitable OS, such as the OS that created the disk in the first place, or at least a closely related one.

Amongst the legal deposit collection, a broad variety of disk formats are found. notable ones include Acorn ADFS, Amiga OFS and FFS, Apple HFS and PC FAT. HFS and FAT are still readily usable, as their vendors (Apple and Microsoft, respectively) are still producing widely-used operating systems that support these file systems. The others are more problematic. Emulators are commonly used to read these file systems, as suitable original hardware is largely unavailable.

Certain storage media also specify their own file systems. Notably, most CDs follow the ISO 9660 file system, and most DVDs use some version of the UDF file system. This is of great benefit, as it means that support for these file systems is broad and, due to the wide use of the media themselves, found in all contemporary platforms.

When considering the risk of obsolescence, file systems (and the operating systems required to use them) fall into three categories, from lowest to highest risk:

- 1 File systems that are widely used, and supported by current/contemporary software
- 2 File systems that are not widely used, and supported by limited current/contemporary software
- 3 File systems that are not supported by current/contemporary software

| File system | Risk factor due to obsolescence (1 = low, 3 = high) | Impact (0 = low, 6 = high) |
|-------------------------------------|--|-------------------------------|
| FAT | 1 | 0 |
| NTFS | 1 | 0 |
| HFS | 1-2 | 0 |
| ISO 9660 | 1 | 0 |
| UDF | 1 | 0 |
| ADFS | 2-3 | 3 |
| OFS | 3 | 3 |
| FFS | 2-3 | 3 |
| (Other obsolete/legacy file system) | 3 | 3 |

| | | | |
|-----------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

7.2.2 File format

The content examined broadly fell into two main classes. Some content was largely “data-oriented”, the remainder “program-oriented”. For example, the sound archive, IDP full, and newspaper archive contained predominantly data files in various formats; Broadcast Wave, TIFF, etc. Conversely, legal deposit and STM full contain large numbers of complex programs and arbitrary non-standard data types.

The “data-oriented” collections are generally in a good position with respect to their file formats. The data from digitisation projects is all quite modern, and was created in contemporary, standard file formats.

The major image formats—TIFF, JPEG, JPEG 2000—are stable, well-documented, and (for TIFF and JPEG, at least) popular and widely used. There is a wealth of current software for using these formats, so their accessibility is currently unproblematic, and due to their entrenchment, this situation is not expected to change in the near future. These formats cover a large portion of the digital data held. Projects such as the Microsoft digitisation is producing JPEG 2000 and PDF, the JISC Newspaper Digitisation projects are using TIFF, and these alone constitute tens of terabytes.

The Sound Archive similarly is using contemporary formats. Although Broadcast Wave is not as widely supported as the image formats mentioned above (as it is not generally an end-user format), it is widely used in the digital audio processing market, and correspondingly has a good level of software support.

The Maps Collection contains a significant amount of “data” material. The discs received from Ordnance Survey (which make up the largest part of the collection) contain only NTF and GML data.

Throughout these collections are found smaller quantities of less standard formats. For example, many of the Endangered Archive discs contain manifests in Excel spreadsheets, and e-manuscripts holds some Word documents. All proprietary, undocumented formats area risk to some extent. the ability to accurately and reliably read the data is dependent on the continued support of the vendor behind the format. Where those vendors are still in business and still producing suitable software, access is usually not a problem. When the vendor is not in business (or has dropped support for a format) the difficulties are much greater. The Word and Excel files are not presently at great risk of inaccessibility. Microsoft still supports these formats and so their obsolescence is not likely in the near future. The picture for some of the similar data formats that are found is not so rosy; some WordStar documents were found on an original IBM PC XT, and this word processor has been defunct for many years. Access to these documents is conditional on either finding suitable third-party software or reverse engineering the documents themselves; for popular, widely used formats third-party software may be an option, but for the niche formats where the original software is unavailable, reverse engineering may be the only alternative.

The “program-oriented” parts of the collection are harder to predict. Programs are more closely coupled to their operating environment than data. A JPEG image can be easily viewed on any current platform, but a 68000 MacOS 7 program can only be run on a 68000 or PowerPC Mac hardware, running a suitable version of MacOS. The legal deposit and (to a lesser extent) STM collections hold a considerable volume of this more complex material. Magazine cover disks, for example, spanning the entire history of personal computing. These disks contain large amounts of both proprietary data and programs to use that data. The breadth and

| | | | |
|--------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

variety of formats and programs makes large-scale reverse engineering untenable, and though the more recent material will tend to be compatible with contemporary systems, the older material is not. Providing suitable environments to run these programs—and hence accessing this material—is therefore extremely difficult.

When considering the risk of obsolescence, file formats fall into three categories, from lowest to highest risk:

- 1 File formats that are widely used, supported by current/contemporary software, and open.
- 2 File formats that are widely used, supported by current/contemporary software, but proprietary
- 3 File formats that are not supported by current/contemporary software

| File format | Risk factor due to obsolescence (1 = low, 3 = high) | Impact (0 = low, 6 = high) |
|-----------------|--|-------------------------------|
| JPEG | 1 | 0 |
| GIF | 1 | 0 |
| TIFF | 1 | 0 |
| JPEG 2000 | 1 | 0 |
| Broadcast Wave | 1 | 0 |
| NTF | 1 | 0 |
| Word .doc | 2 | 0 |
| Excel .xls | 2 | 0 |
| Photoshop .psd | 2 | 0 |
| Wordstar (etc.) | 3 | 3-6 |
| "Programs" | 2-3 | 3-6 |

7.3 Operating environment

As implied above, "data" largely has little specific dependence on the operating environment; "programs", on the other hand, are closely coupled to particular hardware/operating system/software combinations.

7.3.1 Hardware requirements

The general pattern for hardware requirements is similar to that for file and file system formats. Each target platform generally requires specific hardware to run the necessary software. Although many platforms share common traits—for example, many of the 1980s personal computers (Ataris, Amigas, early Apple Macs) used Motorola 68000-family processors—there are usually sufficient differences (such as custom chips and firmware) that each platform must be examined on its own. There are many minor hardware platforms that are all essentially unique; known examples demanded by the Library's collection include

| | | | |
|------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

the Atari ST, the Amiga 500, the Apple II, the Commodore 64, and the Amstrad PCW 9512. These are driven predominantly by the demands of the legal deposit collection. Many of these platforms are unavailable today, obtaining access to material that demands them is difficult. Emulation provides one possible solution for many of these obsolete platforms, although it is somewhat dependent on the popularity of the platform. The major emulation projects for minority platforms are driven by enthusiasts, with a consequence that if no-one is sufficiently enthusiastic about a particular platform, no emulator is likely.

x86-compatible PCs are a key exception to this. The x86-compatible PC architecture is dominant in the personal computing industry, and has been since the 1980s. One of the major characteristics of this architecture is strong backwards compatibility. As a general rule, any software written for a PC—no matter how old—will run correctly on a modern machine. This is not without its restrictions—suitable device drivers can be hard to come by as support for the older OSes is dropped—but the level of support is high, and not found for any other architecture.

7.3.2 Software requirements

The older hardware platforms typically had a one-to-one correspondence with their operating system. Ataris ran TOS, Amigas ran AmigaOS, Acorns ran RISC OS, and so on and so forth. The PC architecture is again an exception; the open platform has seen multiple operating systems over its lifetime. The major operating system family (Windows) offers reasonable backwards compatibility with most Windows and DOS applications (although not all, and this compatibility is slowly diminishing over time). Examples have been found within the Library of applications which fail to run under current Windows versions. There is software within the Maps collection that requires Windows 98 and is incompatible with the Windows XP standard desktop image.

Related to this, and a problem predominantly for contemporary applications, is the need for language support. Non-Latin scripts need fonts and keyboard maps/input method editors not installed in the standard US English Windows version. This is found particularly in the APAC collections; the many non-Latin scripts used in the Far East pose particular problems.

Beyond the operating system, suitable application software is needed to use data files. For widely used formats this is not usually a problem. Software for using TIFF and JPEG images, reading PDF documents, etc. is readily available. For less common formats, the problems are as described in the file format section.

In examining the sample media from the collection areas, a number of more esoteric issues were found. STM created a couple of notable examples. STM contains a number of discs containing time-limited databases. These databases are published each year, and to encourage the use of only the current database, they expire after a certain amount of time. This makes the use of historic versions of the database problematic; without bypassing or circumventing the expiration, the database application will not run.


The second example was a disc from a computing magazine. This disc contained numerous commercial software packages in an encrypted form. The idea was that the decryption keys could be purchased over the telephone and used to unlock the software. However, the company selling the keys has long since gone out of business, meaning that the CD is essentially unusable.

When considering the risk of obsolescence, software requirements formats fall into four categories, from lowest to highest risk:

| | | | |
|------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

- 1 Software that is readily available from the manufacturer and compatible with contemporary hardware
- 2 Software that is readily available from the manufacturer but with compatibility issues with contemporary hardware
- 3 Software that is unavailable from the manufacturer but compatible with contemporary hardware
- 4 Software that is unavailable from the manufacturer and incompatible with contemporary hardware


| Software | Risk factor due to obsolescence (1 = low, 4 = high) | Availability in Library | Impact (0 = low, 6 = high) |
|---------------------------------|--|-------------------------|-------------------------------|
| DOS | 1 | Unavailable | 0 |
| Windows 3.x | 2 | Unavailable | 1 |
| Windows 9x | 2 | Unavailable | 0 |
| Windows NT | 2 | Unavailable | 0 |
| Windows 2000/XP | 1 | Standard | 0 |
| Windows XP non-Latin | 1 | Unavailable | 1-3 |
| MacOS X | 1-2 | Mostly unavailable | 0 |
| MacOS 9/below | 4 | Unavailable | 3 |
| AmigaOS | 2-3 | Unavailable | 3 |
| Atari TOS | 4 | Unavailable | 3 |
| Acorn RISC OS | 2 | Unavailable | 3 |
| Linux | 1 | Mostly unavailable | 3 |
| Solaris | 1 | Unavailable | 3 |
| Niche obsolete operating system | 4 | Unavailable | 4-6 |
| Word | 1 | Available | 0 |
| Excel | 1 | Available | 0 |
| Acrobat Reader | 1 | Available | 0 |
| Photoshop | 1 | Mostly unavailable | 3 |
| NTF software | 1 | Unavailable | 0 |
| Broadcast Wave software | 1 | Available | 0 |
| Wordstar | 4 | Unavailable | 4-6 |

| | | | |
|---|-----------------------------|-------------|---------------|
|  | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

7.4 Policy Issues

As described in the opening sections, The British Library uses a lifecycle approach to collection management. This approach can be used to analyse collection-based content or it can be used to highlight the policy and procedures within collection stewardship. Policy covers a number of risk factors; however, it is felt that these are less of a direct threat to digital material than the factors described above. Therefore, this section serves only to illustrate policy decisions that will help in the stewardship of this material. The Digital Preservation Team should tackle these issues concurrent with addressing the more direct risks.

- Creation
 - The digitisation approvals process does not cover all projects within the BL. Many projects are still co-ordinated from Collection or Business areas of the BL.
- Acquisition
 - There exists a questionnaire that the collection areas (particularly non-English) must fill in and send to overseas publishers. The difficulty for some countries to fill this in to a high standard has resulted in delays in acquisition and a reluctance to develop this part of the collection.
 - Inadequate planning and consideration of what to do with large-scale digitisation output—nowhere centrally to put acquired content.
 - There is not enough up-front consideration of digital preservation needs
- Ingest
 - No standard verification of received media, and in many places there is no checking procedures for digital objects at all
 - No standard analysis of received media (i.e. the specific nature of the digital acquisitions is unknown)
 - No standard handling/storage of received media. In most cases, the digital object is treated as a lesser priority, with the result that many digital objects are stored in suboptimal conditions.
- Metadata
 - No standard cataloguing of received media, meaning that there is no real understanding or knowledge of what it is we hold. Cataloguing, Acquisition and Development have some good procedures; however, these are not mirrored in many other places. This could form part of a training programme.
 - No BL standard (minimum implementable amount) of metadata for digital projects
 - No comprehensive recording procedure of what disks have come from what source
 - No extraction of available metadata, no tools on ingest to help.
 - No good usage statistics are collated for digital objects.

| | | | |
|---|-----------------------------|-------------|---------------|
|  | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

- Access
 - Not considered at point of entry—who is the target market for the acquired material and how will they gain access.
 - Some technical problems (especially software compatibility—unavailability of non-Latin Windows for example) are not ultimately technical (the software is widely and readily available) but policy.
 - Some collection area content may only be accessible on previous versions of operating environments.
- Storage
 - Digital Object Management Programme central storage is available but not ready to help with this. There is a need for a bit-bucket service (from within DOM) to help mitigate the risks.
 - Project-based funding does not always address storage concerns
- Preservation
 - Lack of widespread digital stewards within the collection areas
 - Lack of developed tools and services to aid preservation. There is work being done in this area by the Planets⁵ project. However, there is still a three-year gap between this risk assessment and the end of the Planets project in 2010.
 - DPT resources limit what we can do to help. A separate resource plan needs to be worked out so that the identified risks can be given a timetable for rescue.

Although these issues are not in and of themselves high impact—they will not directly jeopardize access to any material (unless it is simply lost)—a number of them have impact on our ability to do our job. This means that material is at greater risk than it ought to be. For example, the lack of up-front verification and analysis means that we have little knowledge of the true nature of the collection. We cannot distinguish between media failures that have occurred in storage and media that were faulty as delivered, because there is no verification of media on ingest. Though this causes no direct loss of data, it makes estimation of problems much harder.

7.5 British Library Risk Ranking

Media degradation is the biggest direct risk to the British Library's collections. The analysis for this risk assessment has taken into account the variety of software and hardware dependencies within our collections and concluded that although there is some risk, it is with the media itself that the single biggest issue resides. The risk assessment has also looked at policy and procedure issues and, although in need of improvement, they themselves do not pose a direct a risk as the media. We have therefore reached the conclusion that if we address the issue of media degradation than we immediately start to mitigate the level of risk faced by the BL digital collections.

⁵ <http://www.planets-project.eu/>

| | | | |
|------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

There is no way to tackle this wholesale without simultaneously tackling the underlying policy and procedural problems so although degradation is the key driver other factors will weigh our final decision. We cannot perform, for example, wholesale migration from CD-R to hard disk without knowing how many CD-Rs we have, having hard disks to move them to, and manpower to perform the action.

The risk due to improper handling is currently considered to be too low to include in our ranking; although the impact of improper handling can be very high (up to 6), the frequency of usage is (we believe) so low as to render the risk negligible.

As long as the bit-stream is safe (meaning that we have a suitable medium/hardware combination to read it) then further risk factors tend to have impacts of 1-3 (in general, what we hold reflects what is popular, and so emulators etc. are broadly available; this means that we are unlikely to have to write them ourselves). This means that if we move the digital objects from the media and secure the bit stream, then the reduction in risk is significant. Twelve out of the 13 collection areas were found to involve handheld media that was of concern.

Beyond this, "high level" (file and semantic) access should be demand-driven. Specific solutions can be provided to address specific pieces of inaccessible material, but the volume of this material is too great to allow any kind of wholesale provision of high-level access.

These rankings are based on consideration of (a) the impact (b) the likelihood of each risk factor.

| Risk ranking | Risk | Access type jeopardized |
|--------------|--|--------------------------|
| 8 | Media degradation | Bit-stream |
| 7 | Media obsolescence | |
| 6 | File format obsolescence | File/Semantic |
| 5 | Hardware obsolescence | |
| 4 | Operating system + file system obsolescence | |
| 3 | Software obsolescence | |
| 2 | Poor policy (improper cataloguing, metadata) | Semantic |
| 1 | Poor policy (other) | Semantic/File/Bit-stream |

8 Treat Risks

A number of strategies can be devised to treat and mitigate the risks identified, including:

- Create and implement a policy that deals with all digital content in a consistent fashion—acquisitions, digital donations, legal deposit, and content development should treat digital data in a uniform fashion. For example, digital material should be systematically verified on ingest. As part of this, it is expected that there will be training of digital curators.
 - This reduces the variations seen in how digital material is cared for

| | | | |
|----------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

- To perform more detailed investigation and analysis. There is a need for specialist equipment and software; the risk assessment has demonstrated that exhaustive and in-depth analysis is beyond the human and material resources currently available to the Digital Preservation Team.
 - This would facilitate greater accuracy in future risk assessments and allow better management of the digital assets
- Implement a cataloguing policy that creates records for all digital media, and some equivalent for born-digital material
 - This would facilitate greater accuracy in future risk assessments and allow better management of the digital assets
- Move from at-risk physical media to online hard disk-based managed storage
 - This addresses media deterioration, physical damage, environmental damage, and media obsolescence, and is believed to be the best long-term storage mechanism option available
 - This also enhances manageability of the digital collection
- Where migration to hard disk is not immediately possible, move to cold storage to ensure that the physical media last as long as possible
 - This reduces the problems due to media deterioration, physical damage, and environmental damage
- Investigate emulation to retain use of at-risk hardware/software environments and provide the infrastructure to use complex file formats
 - This addresses file system obsolescence, file format obsolescence, software obsolescence, hardware obsolescence
- Investigate migration to retain use of at-risk file formats.
 - This addresses file format obsolescence, software obsolescence, hardware obsolescence

9 Evaluate Risks and final priority table

The Digital Preservation Team is using this risk assessment and a sampling approach to determine which collection areas are in need of most urgent attention. To that end, a prioritized list of collection areas has been created.

| Priority | Collection | Media degradation (8) | Media obsolescence (7) | File format obsolescence (6) | Hardware obsolescence (5) | Operating system/file system obsolescence (4) | Software obsolescence (3) | Poor policy (improper cataloguing, metadata) (2) | Poor policy (other) (1) | Highest applicable risk ranking |
|----------|------------------------------|-----------------------|------------------------|------------------------------|---------------------------|---|---------------------------|--|-------------------------|---------------------------------|
| 1 | SDM | x | | | | | x | x | x | 8 |
| | Endangered Archives | x ⁶ | | | | | x | x | x | 8 |
| | Newspapers | x ⁷ | | | | | x | | x | 8 |
| | Modern British/Legal Deposit | x | x | x | x | x | x | x | x | 8 |
| | Maps | x | | | | | x | | x | 8 |
| | e-manuscripts | x | x | x | x | x | x | | x | 8 |
| | STM | x | | x | x | x | x | | x | 8 |
| | European/APAC | x | | | | | x | | x | 8 |
| | Photography | x | | | | | x | x | x | 8 |
| | Sound ASR project | x ⁸ | | | | | x | | x | 8 |
| | IDP | x ⁹ | | | | | x | | x | 8 |
| Music | x | | | | | x | | x | 8 | |
| 2 | Web Archiving ¹⁰ | | | x | x | x | x | | x | 6 |
| 3 | VDEP | | | | | | x | x | x | 3 |

⁶ Not currently considered a British Library collection

⁷ Although tape may not be as vulnerable as other storage media, there is a 1 year window to verify the media are correct, after which the company which performed the scanning will delete their copy of the material

⁸ The sound archive has begun migration into DOM

⁹ IDP are undertaking to migrate from CD-R and DVD-R to hard disk on their own

¹⁰ Could change quickly with legislative approval to harvest the UK domain

| | | | |
|--------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.1 | Date 26/07/07 |
| | Digital Preservation Team | | |

As stated the highest impact and the most likely risk is that of media degradation. Of the 13 collection areas, only web archiving and limited amounts of the sound archive were not stored on hand-held media. Consequentially, a purely risk-based prioritization is very blunt. The majority of material is faced with a similar level of risk, and no one collection stands far above any other.

In order to split this further in to more manageable work packages the Digital Preservation Team have devised a value scale specific to our team strategy. This value scale includes 8 points and was discussed by the team until a final prioritisation list was established:

- 1 Our obligation to preserve the material
- 2 Estimates of the cost/effort to mitigate the risks
- 3 Estimates of the resource available to the Digital Preservation Team
- 4 Estimates of the cultural significance and value of the collection
- 5 The commercial significance and value of the collection
- 6 The need for further analysis of the collection to inform future preservation activities
- 7 Reader and researcher needs
- 8 Interest and demand

Our obligation to the material is not uniform. For example, current legal deposit legislation requires the long-term preservation of all the discs that accompany books and magazines. Conversely, a JISC-funded digitisation project might explicitly demand the preservation of/provision of access to the digitised data but probably not in perpetuity.

The resources of the Digital Preservation Team are limited. It is therefore practically desirable to prefer those projects with lower resource requirements than those with higher resource requirements. Detailed resourcing estimates are beyond the scope of this document; however, rough relative appraisals have been made for the three levels of access described previously.

The first of these—bit-stream access—is determined principally by the number of physical objects within each collection, and the ease with which those objects can be handled (for example, the Library has good resources for handling CD/DVD media, but lacks sufficient equipment for large-scale handling of floppy disks).

The second of these—file access—is appraised according to the variation found within each collection; a collection with relatively few “data” formats, using formats that are well known, is easier than a collection with a broad variety of formats or containing significant numbers of programs.

The third—semantic access—is judged by the level of cataloguing information and metadata known to exist already, or readily extractable. Some data is quite structured and amenable to the extraction of metadata; web site archives, for example, generally contain reasonable amounts of automatically extractable metadata embedded into the HTML itself. On the other hand, other collections constitute little more than discs full of images; to consolidate these into semantically meaningful groups would require considerable manual effort.

Substantial appraisal of “value” or “significance” is beyond the remit of the Digital Preservation Team, but must be a factor in prioritization. However, there are some

| | | | |
|------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

criteria that would be broadly useful in making such valuations, such as whether the material could easily be recreated, or whether the Library's copy is the only extant copy. Digitisation might tend to be lower value if the original material is in good condition—future digitisations would be possible, and may in fact be higher quality than current digitisation. If the original material is in poor condition, however, then future digitisation may not be possible, and the current digitised copy may represent a digital surrogate—and accordingly be much higher in value.

Commercial value is similarly hard to appraise. The assets created by Imaging Services, for example, may be of commercial interest if suitably collated and presented, and monetization of such assets may become a priority to offset the impact of future funding cuts.

There is value in more thorough examination of the assets the Library holds. The appraisals made in this assessment are best-effort, based on limited samplings of material. The accuracy of these assessments could be improved by conducting exhaustive examinations of collections. These would allow media failure rates to be better ascertained (in particular, it would give some indication of how well industry data on media failures correlates to the Library's own assets). This assessment was also limited in terms of the media types it examined (due to a lack of suitable equipment); this is something that should be addressed, both to improve assessment accuracy and to gain experience in handling a broader range of material and issues.

There is little or no information about reader demand for this material, so at this stage no real appraisal can be made on this basis.

One specific example was identified where the value to readers of the material was likely to be very low; specifically the data in the Music collection. The view was expressed that this data would be of little use to readers as they cannot take it home to listen to—copies of the discs at lending libraries were more appropriate to the way the material would be used.

With these points considered, the resulting prioritisation from a digital preservation assessment is as follows.

| Priority for action | Collection |
|---------------------|------------------------------|
| 1 | SDM |
| | Endangered Archives |
| | Newspapers |
| | Modern British/Legal Deposit |
| 2 | Maps |
| | e-manuscripts |
| | STM |
| | European/APAC |
| | Photography |
| | Sound ASR project |
| | IDP |

| | | | |
|------------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

| | |
|----------|--|
| | Music |
| 3 | Web Archiving* |
| | Voluntary deposit of electronic publications |

Note: *Web-archiving is expected to move rapidly up the prioritisation list based upon our future legislative obligations.

10 Monitor and review

Risk assessment requires a continuous improvement approach to be effective. This document is a tool for digital preservation activity and has prioritised the most at risk parts of The British Library's digital collections. From this list, action can be taken to reduce the risk and to preserve the content in a continuous manner. In order to achieve this, the assessment will be re-evaluated each year.

The purpose of this re-evaluation will be to reduce the numbers in the prioritization table, representing an overall reduction in risk to the collection. This performance should be monitored and reviewed by the Digital Preservation Steering Group. The reduction in risk should become a key performance indicator for the Digital Preservation Team.

The key performance indicator and prioritisation table will become the overriding driver for future digital preservation activity in the area of collection based electronic content. The Digital Preservation Team's activity in this area will provide a continuous assessment of technical obsolescence, the viability of format migration, and availability of emulation technology. This may result in changing priorities or the development new mitigation strategies, where these occur updates will be added to the prioritisation table.

From the prioritisation table it has been agreed by the Digital Preservation Team that all collection content identified as category 8 risk will be addressed first. In order to do this a resource plan will be created separately from this assessment document, which will outline the time, cost, and effort required to tackle all objects within the highest risk category. If the cost is felt to be within the capabilities of the current Scholarship and Collections/Electronic Information Services budget the Digital Preservation Team will take the management of these risks to the next stage of mitigation, actively moving the data to a more stable environment. At this point, the resource plan will become part of the monitoring process.

Summary of monitoring action points:

- Annual update to the risk assessment to continuously improve the condition of the collection based digital objects
- Annual identification of resulting actions to mitigate risks
- Management of the digital preservation prioritisation table
- Key performance indicators to be drawn from the risk factors within the prioritisation table, to be monitored by the digital preservation steering group. (Ideally all risk factors should be in a continuous process of reduction)

| | | | |
|--------------------|-----------------------------|-------------|---------------|
| BRITISH LIBRARY | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

11 Conclusion

The results of the Risk Assessment summarised in Table 8, reveal the considerable challenges faced in securing the British Library's digital collections for the long term. The majority of the collections were rated at the maximum level of risk. While this is a considerable concern, it is not surprising given that:

- Nearly all digital objects currently reside on handheld media
- There is only one copy of the majority of digital objects
- Most digital objects do not have an associated plan for storage, back-up, or preservation

Following from this risk assessment, the DPT will create resource plans to address the risks faced by the collections deemed to be the highest priority. These resource plans will make specific recommendations based on the options enumerated above.

11.1.1 Initial strategies to mitigate risks

This Risk Assessment report recommends that The Digital Preservation Team pursues the following initial strategies as the initial response to this assessment:

- 1 Perform further assessment to gain a better understanding of the media failure rates across the collections
- 2 Address the root causes of a number of the risks facing the collections
- 3 Secure collections that are currently stored on handheld media.

11.1.2 Short term tasks (by July 2008)

To address the collections rated at a category 8 risk it is recommended that the DPT plans, costs, and executes the following tasks:


- 1 Sample and assess the condition of the Legal Deposit collection further.
- 2 Move the SDM collection to more secure storage
- 3 Move the Endangered Archives collection to more secure storage
- 4 Ensure effective QA of the JISC Newspaper project. Prepare a Preservation Plan and provide assistance to the DOM Programme and other Newspaper digitisation projects to ensure successful ingest into DOM

The collections selected for this work provide a useful cross section of condition, content profile, and organisational and procedural process. The BL will gain a deeper understanding of the condition and treatment of these collections by successfully completing these tasks. This will enable more accurate identification, planning and prioritisation of the other category 8 materials.

11.1.3 Medium term tasks (by 2010)

Over the medium term, it is recommended that the DPT pursues the following activities to enhance staff skills, address organisational issues, and improve collection procedures:

- 1 Engage with curatorial staff working with the digital collections in order to:

| | | | |
|---|-----------------------------|-------------|---------------|
|  | Risk Assessment 2007 | Version 2.2 | Date 22/11/07 |
| | Digital Preservation Team | | |

- a. Streamline, enhance, and standardise procedures for acquiring, storing, and preserving digital acquisitions.
 - b. Train curatorial staff, and embed knowledge and understanding of digital preservation best practice across the BL
- 2 Engage with the Digitisation Programme and appropriate digitisation projects to ensure that digital content created at the BL has a low preservation risk. Activities will include:
- a. Review and enhancement of the Digitisation Approvals Process and Digitisation Standards
 - b. Design of a procedure for digitisation project closure and handover of responsibility for the digitised content
 - c. Application of the preservation lifecycle methodology^{viii} to improve resource planning for digitisation projects

These activities are focused on addressing the root causes of the creation and acquisition of at risk content as well as ensuring that curatorial staff have the support, training and guidance necessary to act as stewards for the BL's growing collections of digital content.

ⁱ Digital Repository Audit Method Based on Risk Assessment
<http://www.repositoryaudit.eu/>

ⁱⁱ Redefining the Library: The British Library's strategy 2005-2008
<http://www.bl.uk/about/strategy.html>

ⁱⁱⁱ Longevity of CD Media <http://www.loc.gov/preserv/studyofCDlongevity.pdf>

^{iv} Stability Comparison of Recordable Optical Discs
<http://nvl.nist.gov/pub/nistpubs/jres/109/5/j95sla.pdf>

^v Longevity of high density magnetic media
<http://www.thic.org/pdf/Nov02/nara.vnavale.021106.pdf>

^{vi} Disk failures in the real world
http://www.usenix.org/events/fast07/tech/schroeder/schroeder_html/index.html

^{vii} Failure Trends in a Large Disk Drive Population
http://209.85.163.132/papers/disk_failures.pdf

^{viii} The LIFE Project <http://www.life.ac.uk/>