

Risk Assessment; using a risk based approach to prioritise handheld digital information

Rory McLeod

The British Library, Digital Preservation Team
St Pancras, 96 Euston Road, London NW1 2DB, UK
rory.mcleod@bl.uk

Abstract

The British Library (BL) Digital Library Programme (DLP) has a broad set of objectives to achieve over the next few years, from web-archiving to the ingest of e-journals through to mass digitisation of newspapers and books. These projects are decided by the DLP programme board and are managed by the wider corporate governance structure which includes our legal deposit responsibilities. As part of this work it was identified by the Digital Preservation Team (DPT) that a significant number of handheld media (CDROM, DVD, Tape) within the BL collections may be at increased risk of obsolescence or decay due to the increased time they may spend on handheld media. The DPT and DLP agreed that an assessment should be undertaken and the results used to help prioritize future ingest.

The DPT conducted this risk assessment exercise in order to assess the condition of the BL 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 these 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 hand held (physical carrier) 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 collections have reached high levels (up to 3%).

Additionally substantial quantities of digital objects are stored as single copies only, on handheld media in danger of decay. This stark warning was illustrated by many examples of disc decay that have been encountered and is backed up by the evidence from external research into handheld media lifetimes. Digital content will continue to be lost unless action is taken now. The report made 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. (Move the collections from CD/DVD etc)
- 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

In order to achieve this a number of organisational changes have had to be undertaken that will eventually become measurable benefits.

Using a Risk Based Approach and its organisational impact

Overview

This paper will describe the organisational context within which the BL's 2007 risk assessment should be understood. It is not a technical overview of the methodology or the results as this information can already be found through the final report available at the BL's Digital Preservation website¹. What this paper will do is relate how the process of undertaking such a risk assessment informs the organisational and change management activity that is required to fundamentally shift the perception of digital preservation activity within an institution.

This paper will describe the different levels of organisational involvement required to undertake such risk based activity. The paper will also profile the awareness raising that has resulted in this piece of work becoming one of the most significant piece of analysis so far done by the BL's DPT, and how it has become a catalyst for various follow-up work scheduled for 2008/09. The aim of this paper would be to describe just how beneficial such a piece of work can become to running the business of preservation within a memory institution.

The start point

The BL's DPT is only three years young, incidentally our birthday coincides with the hosting of this years IPRES 2008 event so it is a good time for us to review past work and to think about what the next three years will hold for the vision of digital preservation at the BL and how the risk assessment forms a vital part of this work.

As part of our activities as the DPT we are determined to make sure that our work is representative of the Library's entire digital holdings. This means that as well as the broad corporate programmes (outlined in the abstract) we should focus at least some part of our effort on the material within the collections that does not have a

prioritised timeframe for ingest into our digital library system (DLS). This approach allows us to focus on our growing hand-held media collections and on digital content that may be outside the scope of current DLS work. By using the risk assessment we inform both our DLP and Collection Management strategies as well as providing a practical and measurable information source about the overall state of our collections.

In order to achieve this we first found that we had to undertake a detailed analysis of the BL collections and tie this analysis to the risk assessment following an approach based on international standards for risk management.

The 2007 risk assessment is based on the AS/NZS 4360:2004² Risk Management standard. This standard defines a seven-step approach to risk management:

Communicate and consult

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.

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.

Identify the risks.

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

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.

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.

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.

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.

With the exception of steps 3 and 6 you can see how the methodology refers you to any organisational policies and strategies that may exist and asks you to make reference to them before you proceed. This is a good sanity check before starting as your organisation may not value this activity highly or this may in itself identify a gap in the strategic plan that is worth investigation.

The Risk assessment also provides a way of tying these strategic plan to the operational objectives of the business, for example the BL has a very clear digital strategy³ so for us it was very easy to balance the effort required for this work against the strategy of the organisation. This involved a small scoping study where we worked through points 1 and 2 of the methodology to establish the context. At this point we allowed ourselves a little time to develop the idea within our team using our own department plus our steering group as a mechanism to approve (in this case) our approach.

Communicate and Consult

The first stage of implementing a risk based assessment of digital content is to outline your communication plans and identify your key stakeholders. The BL is an organisation that has geographical challenges due to its multiple sites plus it has challenges of size. This is an organisation of some 2000 people and making your voice heard within such a business is a critical part of the success you can expect. As such the DPT outlined a clear communications strategy to assist, we initially took a top down approach and used our Executive steering group which involves our CEO and a number of Directors and Heads of Department, we presented our plan for the execution of the risk assessment and then allowed a period of time to address concerns raised by this group. The types of queries asked prior to our start were, who will undertake the work? What will be the time commitments in each department? And how much will this help us address or prioritise our digital content, why is this different to the 2003 study?

Our answer were

The DPT will be the primary resource and allocations for time have been given to the two key people involved. Each Department Head should support this and allocate us some time from one member of their team.

When we examined the 2003 risk assessment we concluded:

- Having the object isn't enough
- Knowing the format of the object or its content isn't enough
- You need software to use it, a computer to run it on
- The functionality and access of the object can intimately depend on the details of the environment, most of which we don't have.
- The organisation and business needs to change to support any attempt at e-collection management.

This information was presented to our own team in order to achieve good understanding of the unknowns that we were trying to address.

Internal communications mechanisms such as the Intranet and our staff publications were used to explain why the DPT were undertaking this study and what the benefits would be.

Additionally a questionnaire was compiled as part of the communication plan to be sent to the staff identified by

our Executive steering group. The questionnaire covered the areas listed below and was deliberately left broad enough so as to be easy to start the information flowing back to us.

Location, location, location

- Do you know where your digital assets are?
- If they are related to a physical (analogue) item, are they colocated with that item?
- If not, where are they?
- What conditions are they being stored in?

Retrieval

- Can we achieve easy access to them?
- Can they be sent to us?
- Are they catalogued?
- How many digital assets do you have?
- How big (MB) are the digital assets?
- Is their number of assets considered large? i.e. will we have to examine only a sample set?

Identification

- What media formats do you have?
- (CD (ROM, R, RW, Audio), floppy (various kinds), hard disk (IDE, SCSI, ESDI, etc.), magnetic tape
- What file formats do you have?
- What software environment (operating system, applications) is required to use the assets in question?
- What hardware environment is required to use the assets in question?
- Is there material that you know you have already lost access to?
- Is there material that you would deem to be at high risk?

The questionnaire was well received and alongside our internal communications and reporting structures formed the communication plan.

Establish the context

Once the communication plan was set-up and approval to proceed had been assured, the context of the study had to be drawn up. As stated in the introduction there needs to be a clear relationship between what you are trying to achieve and your corporate or institutional strategies. For the BL this was a matter of looking through our various strategic documents to find the correct measures of value to place our risk assessment with.

The BL follows a number of important legislative and strategic documents. The DPT split this responsibility into an internal (to the BL) and external (to the BL) context.

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

Other relevant BL strategies

- E-IS strategy (the BL's IT strategy)
- S&C content strategy (the BL's Collection strategy)
- 10 Year Digital Preservation strategy

Additionally, the Legal Deposit Libraries Act 2003 and the Irish Copyright Act 1963 (currently being replaced by similar provisions in the Copyright and Related Rights Bill 1999) place upon The British Library the responsibility to maintain legal deposit publications. These publications can include digital objects and, although not expressly covered under existing legislation, the stewardship of these objects must be considered. A proposed extension to legal deposit to cover digital objects is pending and is expected to pass sooner rather than later, 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:

- 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.
- 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.
- 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 were able to take the form of

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

- Organizational/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)

R02	Physical damage	Medium	General
R03	Environmental Damage	Medium	General
R04			General

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 library⁴ 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, this risk assessment goes some way to illuminating how this management can be done in a digital environment

R04a	Technical Obsolescence	Medium	CD-ROM/CD-R/CD-RW
R04b			DVD-ROM/DVD-R/DVD+R/DVD-RW/DVD+RW
R04c			Floppy disk (e.g. 8", 5.25")
R04d			Floppy disk (3.5")
R04e			Hard disk
R04f			Magnetic tape (e.g. IBM 3480)
R04g			Magnetic tape (e.g. LTO3)
R04h			Other magnetic media
R04i			Paper tape/punch card
R04j			Other
R05		General	

Identify the risks

The e-collections analysis 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

R05a	Technical Obsolescence	File system	FAT
R05b			NTFS
R05c			HFS
R05d			ISO 9660
R05e			UDF
R05f			ADFS
R05g			OFS
R05h			FFS
R05i			(Other obsolete/legacy file system)
R06			General

Reference	Risk	Type	Subtype	
R01			General	
R01a			CD-ROM	
R01b			CD-R	
R01c			CD-RW	
R01d			DVD-ROM	
R01e	Physical deterioration	Medium	DVD-R/DVD+R	
R01f			DVD-RW/DVD+RW/DVD-RAM	
R01g			Floppy disk	
R01h			Hard disk (online)	
R01i			Hard disk (array)	
R01j			Hard disk (offline)	
R01k			Magnetic tape (e.g. IBM 3480)	
R01l			Magnetic tape (e.g. LTO3)	
R01m			Paper tape/punch card	
R01n				Other

R06a	Technical Obsolescence	File format	JPEG
R06b			GIF
R06c			TIFF
R06d			JPEG 2000
R06e			Broadcast Wave
R06f			NTF
R06g			Word .doc
R06h			Excel .xls
R06i			Photoshop .psd
R06j			Wordstar (etc.; legacy software)
R06k		"Programs"	
R07		General	
R07a	Technical Obsolescence	Hardwar environ.	PC
R07b			Amiga
R07c			Atari
R07d			Acorn
R07e			Apple Mac
R07f			Sun
R07g			Other

R08			General
R08a			
R08b			
R08c			
R08d			DOS
R08e			Windows 3.x
R08f			Windows 9x
R08g			Windows NT
R08h			Windows 2000/XP
R08i			Windows XP non-Latin
R08j	Technical Obsolescence	Software environs	MacOS X
R08k			MacOS 9/below
R08l			AmigaOS
R08m			Atari TOS
R08n			Acorn RISC OS
R08o			Linux
R08p			Solaris
R08q			Niche obsolete operating system
R08r			Word
R08s			Excel
R08t			Acrobat Reader
R08u			Photoshop
R08v			NTF software
R08w			Broadcast Wave software
R08x			Wordstar (etc.; legacy software)
R09			Complex process for digital acquisitions that discourages material from being collected
	Acquisition		
R10			Insufficient up-front planning of storage and handling requirements
R11			No standardized verification of acquired media
R12	Ingest		No standardized analysis of acquired media
R13			No standard handling of acquired media
R14			Inadequate cataloguing of digital assets
R15	Metadata	Policy	Insufficient creation of metadata
R16			Limited usage statistics collected
R17			Little up-front consideration of who will access material and how they will do it
R18	Access		Internal IT policy causing premature loss of access
R19			DOM not ready to use
R20	Storage		Project-based funding does not always address storage
R21			Lack of digital curators
R22	Preservation		Lack of developed digital preservation tools
R23			Limited DPT resources

Analyse the risks

The 23 identified risks were then analysed using a combination of the DRAMBORA trusted repository

impact scale and industry analysis of the characteristics and the deterioration rates of physical media. 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. Media types are split into optical, magnetic and all others and the types of damage were identified as physical, and environmental. Additionally obsolescence of hardware and software, arguably the most pressing concerns from a digital preservation point of view were used to evaluate the risks at this point. Physical and environmental damage was useful to identify the people and organisational risks. This means that we are able to document and recommend future activity to reduce risk in this area alongside the technical obsolescence thereby covering not just what is at risk today but addressing what might be at risk tomorrow.

Evaluate the risks

At this point we Evaluate and compare to the organisations objectives. This evaluation has allowed us to compare using our LIFE⁵ methodology, the procedural and organisational gaps that have enabled us to plan for future work. Using a lifecycle methodology we are able to track the digital objects whether they are CDROM or DVD and use the methodology to streamline or make recommendations to tighten existing systems. This follow-up work is called the Acquisition and Handling study which will focus in part on training needs and system requirements to reduce the overall risk of the collections. They are;

Creation

- The digitisation approvals process does not cover all projects within the BL. Many projects are still co-ordinated from the Business areas of the BL. (currently now being addressed)

Acquisition

- 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. 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. (now being addressed)
- No BL standard (minimum implementable amount) of metadata for digital projects

- No comprehensive recording procedure of what disks have come from what source
- No extractions of available metadata, no tools on ingest to help.
- No good usage statistics are collated for digital objects.

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 can be policy.
- Some collection area content may only be accessible on previous versions of operating environments.

Storage

- DOM (now DLS) central storage is available but not ready to help with this. There is a need for a service to help mitigate the risks. (now underway)
- 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 Planets⁶ project. However, there is still a time gap between this risk assessment and the end of the Planets project.
- 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.

This assessment of the policy issues surrounding the technical issues have brought to our attention the areas in most need of follow-up consultation

From this combination of media type, risk faced and policy and organisational objectives it is now possible to group the 23 risks into 8 categories and rank them in order to start to mitigate the risk faced by those most pressing.

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

Treat the risks

In terms of the risk assessment itself, treating the risks was considered to be outside the scope. However it is very important to note that the treatment of the risks identified has formed the major part of a funding bid within the BL to address the needs identified. Up until this point it was thought to be the case that hand-held media had a shelf-life that was in keeping with the timeframes to ingest this material. It was actually the case that urgent action has had to be done and so treatment for the risks now falls to the DPT under the name of content stabilisation, this work is currently in progress and is expected to form a vital part in the overall National Digital Library Programme for the UK in coming years. The facility is now installed within the BL's centre for digital preservation and is currently conducting analysis of 120TB of digitised newspaper content.

Monitor and review

Risk assessment requires a continuous improvement approach to be effective. The 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 prioritisation table, representing an overall reduction in risk to the collection. This performance will be monitored and reviewed by the Digital Preservation Team so reduction in risk will 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. This 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).
- Business change functions are being monitored.

Concluding statement

Digital Preservation is much more than a technological problem, its management and measurement requires it to be embedded throughout any organisation. The risk assessment carried out at the BL could be expected to return a similar result regardless of where an institution is geographically or what the organisations function is. The common denominators to this work are what value is placed upon the digital content and what resource is available to do something about it. The ability to use risk as a catalyst for change is a powerful argument and one which has proven beneficial to not just our understanding of the content but our understanding of the organisation and the policies that govern its existence. It is expected by the BL DPT that this work and its subsequent follow on exercises in Acquisition and Handling and content stabilisation will form an important part in our future efforts to preserve e-collections.

British Library digital preservation strategy
www.bl.uk/dp

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

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

McLeod, R and Wheatley, P and Ayris, P. (2006) Lifecycle information for e-literature: full report from the LIFE project. Research report. LIFE project, London, UK. 122p. <http://eprints.ucl.ac.uk/archive/00001854/>

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

Kings of all we survey
<http://www.guardian.co.uk/saturday/story/0,,2041775,00.html>

British Library's Content Strategy: Meeting the Needs of the Nation
<http://www.bl.uk/about/strategic/pdf/contentstrategy.pdf>

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

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

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

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

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

Planets project www.planets-project.eu