The International Journal of Digital Curation
Issue 1, Volume 5 | 2010

Designing Digital Preservation Solutions: A Risk Management-Based Approach

José Barateiro,

Instituto de Engenharia de Sistemas e Computadores Investigação e Desenvolvimento em Lisboa (INESC-ID)

Laboratório Nacional de Engenharia Civil (LNEC)

Gonçalo Antunes, Filipe Freitas, José Borbinha,

Instituto de Engenharia de Sistemas e Computadores Investigação e Desenvolvimento em Lisboa (INESC-ID)

Abstract

Digital preservation aims to keep digital objects accessible over long periods of time, ensuring the authenticity and integrity of these digital objects. In such complex environments, Risk Management is a key factor in assuring the normal behaviour of systems over time. Currently, the digital preservation arena commonly uses Risk Management concepts to assess repositories. In this paper, we intend to go further and propose a perspective where Risk Management can be used not only to assess existing solutions, but also to conceive digital preservation environments. Thus, we propose a Risk Management-based approach to design and assess digital preservation environments, including:

- the definition of context and identification of strategic objectives to determine specific requirements and characterize which consequences are acceptable within the identified context;
- the identification, analysis and evaluation of threats and vulnerabilities that may affect the normal behaviour of a specific business or the achievement of the goals and conformance to the requirements identified in the context characterization; and,
- definition of actions to deal with the risks associated with the identified threats and vulnerabilities.

We generalize and survey the main requirements, threats, vulnerabilities and techniques that can be applied in the scope of digital preservation.[1]

[1] This paper is based on the paper given by the authors at the 5th International Digital Curation Conference, December 2009; received November 2009, published June 2010.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. ISSN: 1746-8256 The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre.

Introduction

Digital preservation aims to optimize the information life-cycle management, from the creation to the dissemination and use of the information objects, for long periods of time. This issue is gaining increasing attention, and important standardization efforts, such as the Open Archival Information System Reference Model – OAIS (Consultative Committee on Space Data Systems, 2003), and the Preservation Metadata: Implementation Strategies (PREMIS)[2] data dictionary for preservation metadata, have contributed to its solution.

The Institute of Electrical and Electronics Engineers (IEEE) defines interoperability as the ability of two or more systems or components to exchange and use information (Geraci, 1991). In fact, digital preservation stresses the time dimension of interoperability, focusing on the requirement that digital objects must remain authentic and accessible to users and systems over a long period of time, thus maintaining their value.

In order to achieve the goals of digital preservation, repositories must "protect" digital objects against several threats that can affect their future interpretation. Actually, protecting digital objects against threats is equivalent to reducing the risk of those threats, which is the main goal of the broad area of *Risk Management* (Institute of Risk Management, Association of Insurance and Risk Managers & Public Risk Management Association, 2002).

This paper proposes a Risk Management-based approach to design and assess digital preservation solutions, enclosing three main phases:

- define the context and digital preservation requirements;
- identify threats and vulnerabilities that may affect the achievement of the requirements; and
- address the potential threats and vulnerabilities.

Even though this process must be optimized for any particular digital preservation scenario, in this paper we generalize this approach by surveying the main requirements to digital preservation, proposing a taxonomy for digital preservation threats and vulnerabilities, and identifying a set of techniques that can be used in digital preservation systems to reduce the consequences of the identified threats and vulnerabilities. Moreover, we evaluate the application of those techniques to digital preservation, regarding the proposed set of requirements and the taxonomy of threats and vulnerabilities. We claim that the proposed requirements, threats, vulnerabilities and techniques can then be further used as the basis to design new digital preservation environments.

[2] PREMIS: Preservation Metadata Maintenance Activity (Library of Congress): http://www.loc.gov/standards/premis

The remainder of this paper is organized as follows. First, we describe the Risk Management approach to digital preservation. Second, the main requirements for the long-term preservation of digital contents are presented. Third, we propose a taxonomy of threats and vulnerabilities to digital preservation. Fourth, we list several techniques that can be used for digital preservation. Fifth, we describe how these techniques can be used to address digital preservation threats and vulnerabilities. Finally, we list the open issues and conclude.

The Risk Management Approach

Risk Management is a continuously developing arena whose ultimate goal is to define prevention and control mechanisms to address the risk attached to specific activities and valuable assets, where risk is defined as the combination of the probability of an event and its consequences (ISO/IEC Guide 73, 2002). It is recognized that Risk Management is concerned with both the positive and negative consequences of risks.

The Risk Management Standard (ISO/FDIS 31000, 2009) is currently a working version, but it is expected to be published during 2009. It intends to define the principles and implementation of Risk Management to control the behaviour of an organization with regard to risk, and is based on the principle that Risk Management is a process operating at different levels, as shown in Figure 1. The Risk Management process encloses the limitation of the context, risk assessment (identification, analysis and evaluation of risks) and risk treatment. This process requires a continuous monitor and review activity to audit the behaviour of the whole environment allowing, for instance, the identification and treatment of an unexpected vulnerability.

First, defining the context is crucial to identify strategic objectives and define criteria to determine which consequences are acceptable in a specific context. Second, today's organizations are continuously subject to several threats and vulnerabilities that may affect their normal behaviour. A process of identification, analysis and evaluation of these threats and vulnerabilities is the only way to decide on the appropriate techniques to handle them. The identification of threats, vulnerabilities and risks is based on events that may affect the achievement of goals identified in the first phase. After that, risk analysis and evaluation estimates the likelihood and impact of risks to the strategic goals, in order to be able to decide on the appropriate techniques to handle these risks (treat risks).

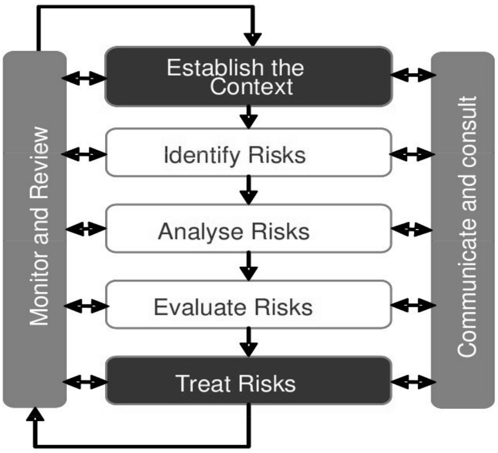

Figure 1. Risk Management Process.

Currently, the digital preservation arena just uses Risk Management concepts to assess repositories. The Trustworthy Repositories Audit and Certification (TRAC) Criteria and Checklist[3] is meant to identify potential risks to digital content held in repositories. It takes OAIS as its intellectual foundation, and as the benchmark for measuring success in terms of trustworthiness. It establishes appropriate methodologies for determining the soundness and sustainability of digital repositories.

The Digital Repository Audit Method Based on Risk Assessment – DRAMBORA (McHugh, Ruusalepp, Ross & Hofman, 2007) process focuses on risks, and their classification and evaluation according to the activities, assets and contextual constraints of individual repositories.

The *Managing Information Risk guide for Accounting Officers, Board Members and Senior Information Risk Owners*, elaborated by the UK government (HM Government, 2008) has a larger span than that of repository assessment. Although the guide deals with risk in the management of digital information in organizations, its scope is not particularly technical.

In this paper, we intend to go beyond and propose a framework based on Risk Management that can be used to assess existing solutions, but also to conceive digital preservation environments, which comprises three steps:

- establish digital preservation requirements (context and strategic objectives);
- identify digital preservation vulnerabilities and threats, and
- address digital preservation threats and vulnerabilities (treat risks).

Digital Preservation Requirements

In order to define the scope and establish the context and strategic goals of digital preservation, we will survey the main requirements of this arena. However, it is impossible to define all the requirements applicable to all digital preservation needs, since digital preservation requirements depend, for instance, on the type, size and amount of data. It also depends on the goals of each organization, regarding the reuse of data. However, there are several generic and common requirements that can be surveyed, based on what someone in the future would require from information stored today.

First, digital preservation requires that a copy (or representation) of any preserved digital object survives over the system's lifetime, which is usually unknown, but may be as long as decades or even centuries. This can be defined as a **reliability** requirement. Therefore, a digital preservation system must be designed to store data indefinitely without suffering any data losses.

Second, a future consumer should be able to decide if the accessed information is sufficiently trustworthy. Usually, this requires the **authenticity assurance** of digital objects (which is already a common requirement for artefacts). Also, the **provenance** of digital objects should be required, especially their creator or the entity responsible for it. Moreover, it is crucial to assure the **integrity** of digital objects, guaranteeing that their informational content has not been modified.

[3] The TRAC checklist is available at http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

Third, digital preservation requires that future consumers are able to obtain the preserved information as its creators intended, **dealing with obsolescence** threats. This requirement involves several challenges, since a digital object, to be explored, requires a technological context defined by specific software and, in some cases, even by specific hardware.

Finally, dynamic collections and environments for digital preservation require technical **scalability** to face technology evolution allowing, for instance, the addition of new components through incremental updates (Baker, Shah, Rosenthal, Roussopoulos, Maniatis, Giuli & Bungale, 2006). Existing static collections (with a fixed size) like, for instance, a digitized historical archive, where no new items will be added, will have a fixed data size. Although it will not be necessary to add new components to increase the storage capacity, it may be necessary to replace components by others with more recent technology (in order to achieve lower maintenance costs or simply because the initial technology was disrupted). This also implies a requirement for supporting **heterogeneity** (which is reinforced by the requirements for scalability).

Fortunately, some typical requirements of normal storage systems are not crucial in digital preservation. For instance, data updates are uncommon because, usually, objects in digital preservation systems are supposed to remain unchanged. Almost all write accesses to the repository are to either ingest new objects or re-write the existing objects in new migrated formats.

Digital Preservation Threats and Vulnerabilities

In this section we present a revision of the taxonomy of threats to digital preservation presented in Barateiro, Antunes, Cabral, Borbinha & Rodrigues (2008), which was based on papers that point out different threats (Baker et al., 2006; Rosenthal, Robertson, Lipkis, Reich & Morabito, 2005).

Table 1 presents our revised taxonomy, which is based on the Risk Management terminology, considering vulnerabilities[4] and threats[5] to digital preservation. Thus, vulnerabilities are weaknesses (potential points of failure) in the environment and threats are events that affect normal behaviour. For instance, a natural disaster threat may exploit several vulnerabilities in the preservation environment.

[4] The existence of a weakness, design, or implementation error that can lead to an unexpected, undesirable event compromising the security of the computer system, network, application, or protocol involved (ISO/IEC Guide 73, 2002).

[5] Any circumstance or event with the potential to adversely impact an asset through unauthorized access, destruction, disclosure, modification of data, and/or denial of service (ISO/IEC Guide 73, 2002).

Vulnerabilities	Process	Software faults Software obsolescence
	Data	Media faults Media obsolescence
	Infrastructure	Hardware faults Hardware obsolescence Communication faults Network service failures
Threats	Disasters	Natural disasters Human operational errors
	Attacks	Internal attacks External attacks
	Management	Economic failures Organizational failures
	Legislation	Legislative changes Legal requirements

Table 1. Taxonomy of vulnerabilities and threats to digital preservation.

Like common information system's architectures, we consider a preservation environment as the aggregation of different components, namely:

- the information entities, including preserved objects and metadata;
- processes controlling the information entities (can be supported by computational services); and
- the technological infrastructure that supports the preservation environment.

Based on that assumption, each of these components may present several vulnerabilities. Thus, we propose a classification of vulnerabilities in:

- process vulnerabilities, affecting the execution of processes (manual or supported by computational services) that control information entities;
- data vulnerabilities, affecting the information entities; and
- infrastructure vulnerabilities, including the technical problems in the infrastructure's components.

Processes supported by software services can be affected by software faults, usually known as bugs, that can cause abrupt failures in the system. For instance, a firmware migration error can cause an unexpected data loss. Software obsolescence is a different kind of failure that occurs when a software component becomes obsolete and unable to run. This type of failure can limit the execution of processes due to the failure to interact with other components (infrastructure or data).

Data vulnerabilities include media faults that occur when a storage medium fails partially or totally, losing data through disk crashes or "bit rot". Again, media obsolescence occurs when the representation format becomes obsolete and cannot to be rendered, even if the "bit stream" survives over time.

Infrastructure components can suffer hardware faults through transient recoverable failures, like power loss, or irrecoverable failures, such as a power supply unit burning out. Similar to media formats, hardware components can become obsolete and unable to communicate with other components (hardware obsolescence). Communication faults occur in packet transmission, including detected errors (e.g., IP packet error) and undetected checksum errors. Other network services failures, such as DNS problems, can compromise system availability.

We propose the classification of threats to digital preservation into disasters, attacks, management and legislation. Management failures are the consequences of wrong decisions that produce several threats to the preservation environment. Disasters and attacks correspond, respectively, to non-deliberate and deliberate actions affecting the system or its components. Finally, legislative threats occur when digital preservation processes or preserved data violate new or revised legislation.

An organization responsible for a preservation system may find itself unable to continue operating at the desired level due to sudden financial limitations (economic failure), political changes or any other unpredictable reason (organizational failure). Furthermore, failures can also occur due to incompetent management.

Natural disasters, such as earthquakes or fires can cause failures in many components simultaneously. For example, an earthquake may cause a data centre to be destroyed, producing, for instance, hardware faults and media faults. Human operational errors may introduce irrecoverable errors. For instance, people often delete data by mistake. Additionally, humans can cause failures in other components such as hardware (e.g., accidentally disconnecting a power cable) or software (e.g., uninstalling a necessary library).

Attacks may encompass deliberate data destruction, denial of service, theft, as well as modification of data or component destruction, occasioned by criminal, political or military activity, such as fraud, revenge or malicious intent. Systems connected to public networks are especially exposed to external attacks, such as viruses or worms. Similarly, internal attacks might be performed by internal actors (e.g., disaffected employees) with privileged access to the organization and to the physical location of the components.

Some vulnerabilities and threats cannot be detected immediately, remaining unnoticed for a long time. For instance, a damaged hard disk sector can remain undetected until a data integrity validation or hard disk check is performed. Moreover, we can not assume threat and/or vulnerability independence.

Digital Preservation Techniques

In this section we present the most relevant techniques and strategies that can be used to handle digital preservation vulnerabilities and threats, increasing the probability of matching the digital preservation requirements.

Redundancy

If data are stored in a single component, they will be lost when that component fails, which is very likely to happen in the long term. Therefore, digital preservation systems can take advantage of a basic attribute of digital information: it can be copied without any loss of information. This means that several copies of the data can be stored across many components.

One of the major problems of systems with replication capabilities is the maintenance of coherence between copies. Several techniques have been developed to handle this issue providing a trade-off between consistency (how many clients can see the digital object they are supposed to see) and availability (how long it takes for clients to access the digital object). However, it is not common that digital contents stored in digital preservation systems need to be updated. Also the availability of preserved contents is not critical (it can be accessed later).

Migration

The goal of migration is to keep digital objects in recent media formats. Lossless migrations of data maintain exactly the same contents as the original version, while loss migrations might imply the loss of some information in the process. Several techniques can be used in migration processes:

- **analogue media**: converting digital media back to analogue formats, such as paper or microfilm. This ensures that the data will be accessible as long as there is a reader for that analogue format. The problems with this technique are the loss of information during the conversion of data from a digital format to an analogue format, and the limitation to data that can be represented in analogue media, such as images or text (interactive or dynamic objects such as software are not supported);
- **version update**: converting data from an old format version to a new one. This is useful since newer versions of the software which use a specific format frequently fail to read very old versions of it;
- **conversion to other formats**: some formats are closed, making conversion only possible when software vendors support the conversion between the format's versions. Therefore, it may be better to convert files in such closed formats to open formats, which are independent of any particular application, or even another vendor's, thereby reducing dependency on the original vendor; and
- **normalization**: in the scope of digital preservation, normalization consists of reducing the number of different formats, in order to reduce the complexity of migration tasks. Normalization can be achieved by limiting the accepted formats or converting the ingested files, on-the-fly, to specific formats. An example of normalization is the conversion of image files in different formats (e.g., BMP, JPG or GIF) to uncompressed TIFF.

Emulation

Emulation is the simulation of the original hardware and/or software conditions of execution for which the information objects were initially conceived (production environment) in more recent systems. It can be a very complex strategy to implement, since it requires not only the preservation of the original objects but also detailed knowledge of the original systems.

Refreshing

The goal of the refreshing technique is to keep the system infrastructure updated with the most recent technology, consisting of the replacement of components by more recent ones. The refreshing of components can be used to prevent failures and obsolescence in the infrastructure's components. Media refreshing for more reliable, durable and less expensive technology can also be attractive.

Diversity

System failures are far from independent. Diversifying the properties of the components can limit the number of simultaneous failures in the system and can be used to design a replication strategy which is more likely to survive a large correlated failure, as in the case of a worm outbreak. Important properties that can be diversified are:

- **physical location**: different geographic locations can limit the number of simultaneous failures in the case of natural disasters or attacks to system components;
- **software**: diversifying operating systems and other software disperses the vulnerabilities to worms or viruses and also prevents vendor lock-in;
- **hardware**: reducing the probability of related component destruction under specific conditions and also prevents vendor lock-in;
- **administration**: independent administration prevents a single person from compromising the entire system in the instance of internal attack or human error;
- **storage**: using heterogeneous media for storage minimizes the impact of failures due to technological defects; and
- **funding:** diversifying the sources of funding for digital preservation systems prevents economic disruption.

Inertia

A system that works quickly also fails quickly, especially if subject to deliverate attack. Since digital preservation systems usually do not require speed, they can be designed not to change rapidly. Consequently, the system is less likely to fail abruptly. For example, when the system is attacked, administrators have more time to take corrective action and prevent total failure. Obviously, there are some types of attack and failure that cannot be slowed down.

Moreover, digital preservation systems should sometimes operate quickly, such as when recovering from a failure. Still, this principle can be observed in parts of the system. For example, limiting the rate at which files can be deleted by an administrator. The LOCKSS (Maniatis, Roussopoulos, Giuli, Rosenthal & Baker, 2005) system implements this strategy through rate-limiting techniques, which allows it to limit the rate at which an attacker can make progress compromising the rest of the system after he is able to compromise a node.

Metadata

We can simply define metadata as "data about data". Therefore, metadata are an added value, which is usually required to interpret data. Metadata is not only a digital preservation technique per se, but are also required to correctly apply other techniques. For instance, emulation and migration require highly detailed metadata.

With respect to digital preservation, we can find different classes of metadata. We propose the following classification:

- **descriptive metadata**: are information describing the content of a specific digital object. In domains like digital libraries and archives, descriptive metadata standards are broadly used as, for instance, bibliographic descriptions using MARC[6] schemas;
- **technical metadata**: focus on the characterization of the technological context (specific software and hardware) used in the generation of digital objects describing, for instance, the format, format-specific technical characteristics, and so forth;
- **structural metadata**: provide information to establish relationships between different digital objects to create a logical unit. The Metadata Encoding and Transmission Standard - METS (Digital Library Federation, 2007) is a metadata specification specially created for "the management of objects within a repository"and "the exchange of objects between repositories", that includes structural mappings and links;
- **preservation metadata**: are metadata elements that could be used explicitly for preservation. The PREMIS dictionary of preservation metadata relies upon the concepts of Intellectual Entity, Object, Rights, Agent and Event, to prove authenticity and integrity of digital content;
- **rights metadata** are used to characterize and define rights of digital contents. Some standards have been developed, like copyrightMD[7] and METSrights[8].

Note that the categories of the proposed classification may overlap and there are some standards that address issues in different categories.

We would like to stress that a metadata standard is not metadata itself, but a specific schema to represent metadata. Consequently, it is possible to have different instances and implementations of the same metadata standard.

Auditing

Auditing supports the detection of latent faults, allowing the system to recover faster and reducing the chance of losses. For example, faults that cause data loss may only be detected when the data are accessed. This can be done by auditing the system periodically.

Auditing is especially important in digital preservation systems with few data accesses. When data ingestion is performed by a third party, it may also be crucial to audit these systems, in order to check if the data are properly ingested. The most prominent techniques usually involve challenge-response protocols using digests such as MD5. The main disadvantage of this approach is that digest algorithms can be cracked and usually rely on a lack of computing power of an attacker for security. This means that a digest algorithm may cease to be secure while the digital preservation system is operating. Therefore, if a system is not flexible enough to allow the auditing system to be changed during all of its lifetime, it is likely that the ability to securely perform audits will be lost.

[6] Library of Congress – Network Development and MARC Standards Office: http://www.loc.gov/marc/
[7] California Digital Library: http://www.cdlib.org/inside/projects/rights/schema
[8] Library of Congress Standards: http://www.loc.gov/standards/mets/news080503.html

Addressing Digital Preservation Threats and Vulnerabilities

Following the Risk Management process, this section shows how the list of techniques presented in the previous section can be used to treat the risks associated with digital preservation threats and vulnerabilities identified in this paper (see Table 2).

First, we would like to point out that auditing is used to quickly detect vulnerabilities and threats to the preservation environment, allowing the rapid execution of corrective techniques. Thus, the auditing technique is equivalent to the monitor and review activity of the Risk Management process.

The risk of media faults can be reduced by refreshing the media supports used (e.g., replace hard disks). Moreover, redundancy, metadata and auditing support the recovery from media faults, in the sense that auditing allows the quick identification of undetected media faults. Metadata are also required, for instance to verify the integrity of corrupted objects and finally, redundancy is crucial to obtain an undamaged copy.

The extra information required to be able to deal with obsolescence of media formats is usually the technical metadata. In the simplest scenario, the requirement of storing the object along with its technical metadata (encapsulation) may suffice. In this scenario the preservation system is not required to do anything special with that data, except give it back when required. In a more complex scenario, the system might be required to support specific input and output formats, to make it possible to ingest objects in one or more formats and retrieve them in other different ones (whether the transformation inside the system is done in advance or in real time is a question of implementation to be dictated by other, for now irrelevant, requirements).

Threats and vulnerabilities			Techniques							
			Redundancy	Migration	Emulation	Refreshing	Diversity	Inertia	Metadata	Auditing
Vulnerabilities	Data	Media faults	R	-	-	r	-	-	R	R
		Media obsolescence	-	r	r	-	-	-	R	R
	Infrastructure	Hardware faults	-	-	-	r	r	-	-	R
		Hardware obsolescence	-	-	-	r	r	-	-	R
		Communication faults	-	-	-	r	r	-	-	R
		Network service failures	-	-	-	r	r	-	-	R
	Process	Software faults	-	-	-	r	r	-	-	R
		Software obsolescence	-	-	-	r	r	-	-	R
Threats	Disasters	Natural disasters	R	-	-	-	r	-	-	-
		Human operational errors	R	-	-	-	r	r	R	R
	Attacks	Internal attack	R	-	-	-	r	r	R	R
		External attacks	R	-	-	-	r	r	R	R
	Management	Economic failures	-	-	-	-	r	-	-	R
		Organizational failures	-	-	-	-	r	-	-	R
	Legislation	Legislative changes	-	-	-	-	r	-	r	-

Table 2. Addressing digital preservation threats and vulnerabilities. r=: reduces the risk of the threat/vulnerability; R=: required for recovery; -: does not fit.

Each of these scenarios will be adequate to different preservation strategies to deal with format obsolescence. For example, the first is the adequate for emulation, while the second might be required for migration. Note that usually, emulation is required for dynamic objects (like games or other software), while migration is widely used in static objects (e.g., images, text).

The risk of obsolescence and faults in components of the preservation infrastructure can be reduced by diversifying the infrastructure components (reducing the probability of correlated failures) and by refreshing the components with recent and robust ones. Similarly, the risk of vulnerabilities that affect processes supported by computational services (software) can be reduced by diversifying and refreshing these computational services, since common software components present the same vulnerabilities to viruses, bugs, and so on.

With respect to disaster threats, diversity, especially of the physical location, is the technique that can be adopted to reduce the risk of natural disasters like earthquakes or fires, since natural disasters may affect several components in a limited area. Human operational errors can produce the same effects as deliberate internal and external attacks. Thus, diversifying the administration and physical location limits the effect of malicious actions that can be performed by attackers. Moreover, if inertia is also applied, the speed of destruction is also decreased. Furthermore, physical location diversity also reduces the risk of terrorist attacks to the infrastructure, and auditing, metadata and redundancy are crucial to detect the effects of operational errors and attacks, also supporting recovery to a normal state.

Management failures occur when an organization responsible for running a preservation environment becomes unable to continue operating, due to financial limitations, political reasons or other organizational problems. From the surveyed techniques, diversifying funding or even organizations involved in the digital preservation environment are techniques to reduce the risk of this threat.

Sudden changes in legislation may impose changes in object rights, modification of preservation processes, and so on. The risk generated by such legislative changes can be reduced by the diversifying of responsible bodies (being located in different legislative jurisdictions) and the adequate cataloguing of rights metadata.

Conclusions and Open Issues

This paper presents a Risk Management based approach to digital preservation, consisting of three different phases: establishing digital preservation requirements, identifying digital preservation threats and vulnerabilities, and treating the risks associated with the threats. We surveyed the main requirements to digital preservation and classified the threats and vulnerabilities that might endanger preservation using a taxonomy of threats/vulnerabilities. A description of common techniques used in preservation was also presented. Finally, we proposed a mapping between the techniques and the associated threats/vulnerabilities they address.

In a digital preservation system, components may fail in a correlated manner, since some threats may cause the failure of multiple components with similar configurations. Moreover, each preservation scenario has its own specificities, making it impossible to determine which technique is better suited to all the scenarios.

Moreover, even if one can specify that a specific technique is the most appropriate, there are several potential applications of this technique (e.g., which format to migrate, where to put copies and how many in redundancy strategies). To effectively assess and measure adequate risk treatment for digital preservation scenarios, we proposed a simulator (Antunes, Barateiro, Cabral, Borbinha & Rodrigues, 2009; Barateiro, Antunes, Freitas, & Borbinha, 2009), that can be used to evaluate the risk of threats (natural disasters) and infrastructure failures, on a preservation environment using redundancy and diversity techniques. We plan to extend this simulator, currently under development, to follow this Risk Management based approach, developing an effective risk analysis tool for preservation using the proposed taxonomy of threats/vulnerabilities and digital preservation techniques.

Acknowledgements

This work is partially supported by the projects GRITO (FCT, GRID/GRI/ 81872/2006) and SHAMAN (European Commission, ICT-216736), and by the grant from FCT (SFRH/BD/23405/2005) and LNEC to José Barateiro.

References

Antunes, G., Barateiro, J., Cabral, M., Borbinha, J., & Rodrigues, R. (2009). Preserving digital data in heterogeneous environments. *2009 Joint international Conference on Digital Libraries. Austin, TX, USA.*

Baker, M., Shah, M., Rosenthal, D., Roussopoulos, M., Maniatis, P., Giuli, T., et al. (2006). A fresh look at the reliability of long-term digital storage. *1st EuroSys Conference. Leuven, Belgium.*

Barateiro, J., Antunes, G., Cabral, M., Borbinha, J., & Rodrigues, R. (2008). Using a GRID for digital preservation. *11th International Conference on Asian-Pacific Digital Libraries. Bali, Indonesia.*

Barateiro, J. Antunes, G., Freitas, F., & Borbinha, J. (2009). Challenges on preserving scientific data with data grids. *1st ACM Workshop on Data Grids For E-Science*, *Ischia, Italy.*

Consultative Committee on Space Data Systems. (2003). *Reference model for an open archival information system* ISO 14721:2003. Retrieved from http://public.ccsds.org/publications/archive/650x0b1.pdf

Digital Library Federation. (2007). *Metadata encoding and transmission standard. Primer and reference manual*. Retrieved from http://www.loc.gov/standards/mets/METS Documentation final 070930 msw.pdf

Geraci, A. (1991). IEEE Standard Computer Dictionary: Compilation of IEEE Standard Computer Glossaries. *IEEE Press*, Piscataway, NJ, USA.

Institute of Electrical and Electronics Engineers. (1991). *IEEE standard computer dictionary: Compilation of IEEE standard computer glossaries*.

HM Government. (2008). *Managing information risk: A guide for accounting officers, board members and senior information risk owners*. Retrieved from http://www.nationalarchives.gov.uk/services/publications/information-risk.pdf

Institute of Risk Management, Association of Insurance and Risk Managers & Public Risk Management Association (2002). *A risk management standard*. Retrieved from http://www.theirm.org/publications/documents/Risk_Management_Standard_030820.pdf

ISO/FDIS 31000. (2009). *Risk Management principles and guidelines*.

ISO/IEC Guide 73. (2002). *Risk management. Vocabulary. Guidelines for use in standards.*

Maniatis, P., Roussopoulos, M., Giuli, T., Rosenthal, D., & Baker, M. (2005). The LOCKSS peer-to-peer digital preservation system. *ACM Trans. Comput. Syst. 23,(1):2-50*.

McHugh, A., Ruusalepp, R. Ross, S. & Hofman, H. (2007). The Digital Repository Audit Method Based on Risk Assessment (DRAMBORA). *DCC and DPE, Edinburgh*, 2007.

Rosenthal, D., Robertson, T., Lipkis, T., Reich, V. & Morabito S. (2005). Requirements for digital preservation systems: A bottom-up approach. *D-Lib Magazine 11*,(11). Retrieved from http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html