Digital Preservation and Technological Obsolescence: A Risk Assessment Strategy

Suzanne LeBlanc

Suzanne has a BA in World Literature and Culture Studies from the University of New Brunswick and an MA in Theory, Culture and Politics from Trent University. She is a second year MIS student at the University of Toronto’s iSchool. Suzanne has completed a specialty in library sciences, and is working towards a second in archives and records management. She has been working in the wealth management industry as a records manager for several years, which has piqued her interest in the flow of information within corporate structures. Suzanne has more recently turned her attention towards issues of digital preservation and corporate memory.

In a digital format, the essential nature of recorded information has changed in several important respects: records can now be easily distributed and refreshed, they do not suffer a loss of fidelity when copied or used, their integrity can be secured and verified, they can be analyzed using a range of automated tools, and they can be searched with ever increasing ease and accuracy. Specific methods are needed to ensure the preservation of these digital resources. From the mass of digital records being generated, archivists need to consider how much data should be retained, and whether it would be practical or even possible to preserve all of this information for posterity.

Preservation of digital information requires active intervention, as data recovery is a poor solution to the preservation problem. It is expensive, labour intensive, and risky considering that “few data recovery methods are comprehensive or successful on every occasion” (Ross, 2000, p. 12). Specific digital preservation strategies which have been investigated include: the preservation of obsolete technologies, emulation of obsolete software and applications, migration of digital records to new environments, persistent object preservation, and bundling. This paper, rather than expanding upon any of these specific technological solutions, will address the problem of digital preservation in general as an ongoing process of
risk assessment and management where the cost of preservation, by whatever means is appropriate for the particular medium and format, is judged against the cost of failure to preserve the records in terms of the patrons who need the information. It will outline a digital preservation method for describing the responsibilities for preserving electronic records.

Scope

A risk assessment lays out a model whereby an archives is able to verify, against a set of guidelines, the archival quality of a record, whether the record can be retrieved and used, and if, how, and for how long, the record can be preserved. Specifically, this paper will only consider records which are born digital, created by an organization, and preserved by an archive which has responsibility for that particular organization’s records. This clarification is necessary, since digital preservation is generally understood to encompass two distinct preservation activities: either the digitization of physical records to preserve their contents, or the long term preservation of records which are born digital. While the process of digitization is interesting, it presents its own problems which are distinct from those issues related to preservation of digital records, and will not be addressed here. Similarly, the preservation of private records produced by individuals is also an important topic – which has its own problems that will not be addressed by the scope of this paper. Risk assessment as a preservation practice does not lend itself well to private records, so we will focus instead on organizations whose administrative capacities enable such an approach.

A record is broadly defined as information which documents an activity, such as the administrative and bureaucratic activities in which organizations engage. These activities generate various kinds of digital records, such as database contents, email, web pages, geographical information systems, and text documents. Practitioners of digital preservation seek to protect the integrity, functionality, and meaning of digital records. Risk assessment allows for the level of control which this goal requires. Where the acquisition and management of the records is controlled, contextual information is secured, and appropriate metadata are attached to ensure the interpretation of these records in the future (Ross, 2000). That being said,

… it is not possible to preserve an electronic record. It is only possible to preserve the ability to reproduce an electronic record. It is always necessary to retrieve from storage the binary digits that make up the record and process them through some software for delivery or presentation. (Ken Thibodeau, as cited in Heather MacNeil, 2000, p. 53)

The problem is one of reproduction; a digital preservation risk assessment addresses this by evaluating how feasible it is to reproduce records over time.

Risk Assessment

A risk assessment focuses on the identification of potential problems. It exposes inherent vulnerabilities that have been embedded within the digital records it seeks to
preserve. It addresses risks to the content of the digital records, not the risk to the organization or archives. It evaluates all of the records under consideration to determine which ones may be at risk and should be included in a preservation program, and what constitutes risks to those records. ‘Risk,’ though, is a relative term – what may constitute a threat in one environment may not in another. Thus, a risk assessment should be tied to the values of the organization whose records it hopes to preserve. A record’s values which may be held in highest regard by a corporation could range from requirements for law and evidence, to monetary or economic value, to even the intangible, such as history and corporate memory. A risk assessment will also include looking at significant properties of digital records – those which affect the quality, usability, appearance, functionality and behaviour of the digital objects. These properties are important because they affect the authenticity, legal admissibility, artifactual value, functionality, look and feel, and behaviour of a digital record.

Each archives should define its own “worry radius”, which is the context that provides definitions of perceived risk and acceptable loss (Kenney et al., 2002). The worry radius of an organization takes into account the scope and value of its digital records. It determines how much loss is acceptable and what properties of the record are most important, as loss is inevitable when preserving digital information. A good program should also balance flexible access to digital records while ensuring the security needed to protect them.

A risk assessment is laid out in stages beginning with data gathering and characterization, simple risk declaration and detection, contextualized risk declaration and detection, and ending with risk management implementation (Kenney et al., 2002). Data gathering and characterization involves collecting a range of data using applicable collection and manipulation tools and techniques, the selection (or creation) of which will depend on the nature of the digital records in question. Simple risk declaration and detection is the process of developing a model to categorize risk, into which observable, potentially risky characteristics of records can be entered. Qualitative and quantitative methods are used to characterize and classify the risks. Contextualized risk declaration and detection involve defining relevant scenarios or events which could result in damage or loss to a digital record, including the probability of these events. It considers questions such as: What are the possible threats to the records? What are the known vulnerabilities? What is the likelihood of loss? What are some potential safeguards? An example of risk declaration and detection would be a scenario in which a provider of proprietary software stops supporting the formats of digital information it produces. This section of the risk assessment would ask and determine what the likelihood of this would be and what danger this would present to the continuing preservation of the digital record. Lastly, risk management implementation involves creating policies, procedures and mechanisms to manage and respond to identified risks. The value of the
records should be balanced against the cost of preventing or recovering the records from damage or loss. Automatic methods will most likely be needed to deal with the sheer volume of records in a cost effective manner, as well as to obtain consistent results less prone to human error.

Characteristics of Digital Records

As mentioned above, a risk assessment lays out a model in which an archives is able to verify the archival quality of a record. For this strategy to work, it must be based on a shared understanding of the characteristics of digital records. If one looks first at the theoretical characteristics of a record, one can see that the archival quality of a record can be broken down into five characteristics: naturalness, interrelatedness, uniqueness, impartiality, and authenticity. The naturalness of a record is confirmed once it has been determined that it was not collected for some purpose outside the administrative needs generating it, and it was not put together with other records according to some scheme to serve other than those needs. The interrelatedness of a record refers to its archival bond – that is, its relationship to the records around it or connected to it in some digital manner. The characteristic of uniqueness may first appear to be an outdated method of archival analysis, since in the instance of digital records, the viewable record is always a copy. However, the context of a digital record renders it unique. The authenticity of a record is verified through the determination of the identity and integrity of a record.

Of the five characteristics, the last (authenticity) deserves special attention, as it establishes an important connection between the public and the archives as a trusted institution, as will be described below. An authentic record is one that can be proven to be what it claims to be and is free of falsification or inappropriate modification (MacNeil, 2000).

Digital records possess a fixed documentary form, content (which is not necessarily stable), an archival bond with other records, and context. These properties manifest themselves in explicit and implicit ways. For example, an explicit way would include the information contained within metadata linked to the record. Implicit information about a record, on the other hand, would most likely be found through an investigation of the content of a record, or through an analysis of its archival bond.

Next is a consideration of the physical characteristics of digital records. Digital records are stored on a variety of media, from magnetic or optical disks, to magnetic tape and solid state memory storage. Except for files that are simply data streams, most digital records contain two basic components, namely structural elements and data elements. A file format “represents the arrangement of structural and data elements in a unique and specific manner” (Lawrence, 2000, p. 2). Metadata should also be considered as a physical characteristic of digital records, as it exists as bits on a disk just as much as the contents of a digital record do. In any preservation strategy, metadata plays an important role – it is used to capture the context of a digital resource, including the processes that
Digital records are also mutable – that is, they are interactive and can be rearranged in respect to other records. Consideration of the physical characteristics of a digital record should also include the platforms, interfaces, standards and coding necessary to read and contextualize it.

**Preservation Methods**

In this way, metadata is crucial to the preservation of digital records. In order to best capture the context of the processes and use of a record, metadata should include information about the source of the data, the how, why, and when it was created, its intended function or purpose, guidelines about how to open and read the record, terms of access, the migration history of the record (if it is migrated), information about how it interrelates to other records within the organization and broadly, how it relates to external organizations. Metadata, though, is a digital record itself and as such the same set of problems applies. For instance, there are many different models, which have not been tested for interoperability, and the use of fields to represent various business processes have not been standardized.

Further, metadata is associated with documentation, which may be lacking and which also takes time and labour to produce. Metadata is derived from an analysis of an organization’s structure and functions and its fields need to be derived from an analysis of business processes to ensure that they reflect the functional uses of the data to which they are linked. The purpose of this is to ensure that future users of the records will be able to understand the role that the information played in the organization. Metadata should also include evidence of processes, which gives context to the records. Finally, there are difficulties associated with ensuring that future systems will be able to read and use the metadata.

Metadata serves an important role in the preservation method of migration. Migration is “the process of rearranging the original sequence of structural and data elements [...] to conform to another configuration” (Lawrence, 2000, 2). In this instance, metadata needs to be linked to the resource, but held apart in order to support the migration of records through multiple generations of hardware and software by supporting the reconstruction of decision-making in regards to migration. The issue with migration as a method of preservation, which forms a category of risk within a digital preservation risk assessment, is that something about the original record will be lost – some property (or properties) of the original record will not make the transition. They will either be cut, or rearranged from the original order. Thus acceptable levels of loss and ensuring the functional relationship between records, before and after migration, need to be determined, a process which can be facilitated by the creation of good metadata.

Another, less popular, method of digital preservation is emulation. A piece of software is created which will run on contemporary hardware, but will mimic obsolete hardware, so that the digital record can be accessed and read in the original environment in which it was created. Emulation requires future access to multiple data objects held in a cluster: the data...
file to be preserved, the application software that generate the file, the operating system in which the application functioned, and the hardware environment which is emulated through the software (Lawrence, 2000). As this is a complicated procedure, it is both an expensive method and one which has an increased chance of ultimately failing.

Categories of Risk

Keeping in mind that different organizations’ realities vary, archivists should develop their own situation-specific risk assessment and develop their strategy accordingly. Each must determine what level of risk and loss they are willing to assume. Conceptually, any loss is not acceptable. Practically speaking, however, it is difficult (perhaps even impossible) to employ preservation methods which are entirely loss-free. The loss must be kept to a minimum in order for future users to be able to identify, retrieve, render, manipulate and use records and the properties of context, content, integrity and functionality must be preserved (Ross, 2000). Within this category the time-frame for which the archives hopes to preserve the records should be considered, and should be done so in terms of what is viable and what is necessary. The risk assessment must also take into account the extent of the preservation measures for a record – nothing lasts forever and the value of the record may not warrant the preservation measures required.

An assessment of the kinds of procedural controls which were used on a creator’s system and the records it contains is one way risk should be assessed. This will help determine the authenticity of those records from the creator’s point of view. Consideration of the metadata fields which are currently attached to a digital record, such as name of the author, chronological date, and place of origin, form part of the record’s juridical and administrative context and also help to determine the authenticity of the record, especially if one of those fields marks that the content is especially at risk of deterioration. This might be the case with the date field as records older than a certain date might be at risk due to changes in software since their creation. A preservation method such as migration might have to be applied as soon as possible in such a situation.

The environmental conditions under which digital information is deposited influences its survival, recovery, and study (Ross, 2000). An assessment of the risk associated with particular media is thus an important component of the proposed method. For instance, there are problems associated with the physical breakdown of media, records being inaccessible due to technological advances, and records rendered meaningless due to a lack of contextual evidence. For digital records to survive more than ten years (on more than good luck) digital preservation must be planned, as digital storage media “give little clue to the presence or format of the data they contain, and only occasionally an indication of what devices might be needed to access them” (Ross, 2000, p. 6). Peripheral devices and the software to operate them come into widespread use and are made obsolete often within a fifteen year span. Though it
would not be impossible to read a 5¼” floppy disk, it would also not be easy or cheap, and not all archives would have the means to do so. In fact, no archival organization can realistically hope to maintain a comprehensive collection of old computer hardware – integrated circuits, thin film heads, and laser diodes cannot be repaired or easily fabricated, except in multi-billion dollar facilities. The ability to read tapes or disks in the future without the appropriate peripheral device needs to be taken into consideration under its own category of risk.

Determining the authenticity of a digital record comprises another category of risk, which is assessed in relation to the identity and integrity of a digital record (MacNeil, 2000). That is, it should be determined whether the record has been altered in any way since its creation that may have changed its essential character, which change was not documented through the use of accompanying metadata. For example, a digital record may be altered in the process of migrating it to a new environment. Any changes to its essential character on this basis would have a different, less significant, impact on its integrity rather than changes to a record made by a user. Documentation, often through the use of metadata, could serve to preserve a record’s identity and integrity over time by determining its history.

A determination of authenticity should also take into account the endurance and stability of the record over time. The stability of the content plays a role in a record’s claims to authenticity. The point in time at which the record is considered to be complete, stable, and unchangeable should be determined, though there may be no such point. Documentation must be created to that effect, citing how it can be changed, who has the authority to make changes, and how and what extent changes are tracked by the system (MacNeil, 2000). The production of authentic copies of electronic records also needs to be supported, once they are in the preserver’s care, and should be a part of the risk assessment. For instance, can the presentation features be reproduced? This includes the overall visible configuration or representation of the content, in particular “aspects of the record’s formal presentation that are necessary for it to achieve the purpose for which it was created” (MacNeil, 2000, p. 59). One must consider to what extent these features are crucial in communicating the content, and based upon that assessment whether one can afford to lose them.

Records that are relied upon by the creator for purposes of carrying out its business are presumptively authentic (MacNeil, 2000). The presumption of trustworthiness becomes less supportable once the records are no longer being actively used, as the motivation to maintain accurate and authentic records no longer exists. It is at this point that records should be considered for accessioning to the archives. A risk assessment should consider at what point in a record’s life cycle this is done. It is important to verify that the digital records upon which the creator relied are “clearly identifiable and of demonstrable integrity and that accidental corruption or purposeful tampering have not occurred after the records
are no longer in active use by the creator.” (MacNeil, 2000, p. 69). Foundation requirements for the authenticity of digital records kept by the creator of the records are essential to the archives in order to verify authenticity in this instance (MacNeil, 2000).

The authenticity of a record can be broken down into two component parts: its identity and its integrity. The identity of a record “refers to its provenance, author, addressee, writer, date, action or matter, and archival bond,” (MacNeil, 2000, p. 69) while integrity refers to its condition as unimpaired and its completeness. By this definition, the integrity of a record could be compromised upon migrating the record to a new environment. Thus an assessment of integrity is not made in an absolute sense but instead is made “in relation to the purpose the record serves in the environment in which it has been created, maintained, and used” (MacNeil, 2000, p. 70). Further, the elements of a record which are needed to convey its identity should be intelligible and the message that it is meant to convey should be able to meet the same goal as before. The precise number of bits of a record “need not be replicated in a copy, provided that the articulation of the content and its required formal elements remain the same” (MacNeil, 2000, p. 70).

To evaluate the authenticity (and hence integrity and identity) of a record, the archives should consider the following when performing their risk assessment: who implemented and monitored access privileges in the electronic system, whether a profile was designed to annotate (and be linked to) each record (essentially containing metadata to verify the record’s identity and any changes to the record), whether there were established audit procedures of the administration of the records and use of access privileges, whether procedures were established to prevent loss or corruption of records, and finally whether methods for transferring inactive records to the archive were followed (MacNeil, 2000). Despite this long list of assessment criteria, only the authenticity, identity and integrity of the record as a record can be judged. The content of a record is not submitted to a test of its truthfulness and an authentic record may faithfully document a falsehood. A record whose integrity or identity has been compromised by its creator cannot be rescued from its tarnished reputation by the risk assessment. Rather, the risk assessment is used to determine whether it is worth the effort to preserve a record whose authenticity is under suspicion.

A category of the risk assessment should broadly consider the impact of technological watersheds on digital record integrity. Major technological change can necessitate a complete reconceptualization of an organization’s method of operation, including the archives’. Technological change induces risk through several mechanisms. First, through an organization’s failure to maintain a current software operating environment, then a failure to upgrade file formats and encoding schemas, and finally through a failure to use modern tools (Kenney et al., 2002). Such failures can make taking in records from an organization very problematic for an archives. At the same time, if
an archives similarly lapses, its ability to digitally preserve materials is also compromised. While major technological shifts cannot be predicted, a retrospective analysis of technological change can help provide guidance on what threats to content should be taken most seriously. Studying periods of intense and rapid growth, when many new standards and features were introduced, may offer some guidance (Kenney et al., 2002).

**Risk Mitigated**

Authenticity of records over the long term depends on the existence of trusted recordkeeping systems that are associated with digital records (Kenney et al., 2002). The archives itself helps with this, as it is the job of the archive to protect records from being altered or tampered with. To paraphrase the *Draft Requirements for Authenticity*, any digital copy of an authentic digital record is authentic if it is declared to be so by the official preserver, namely the archives (MacNeil, 2000). Indeed, the relationship between archives and the value of their records “is rooted in Western traditions and in the purposes of ancient and medieval archives.” (Dirks, 2004). The trustworthiness of documents was conferred by their preservation within an archive, which was the place where documents were kept by their creator (Duranti, 1994). Faith was placed in the unbroken custody provided by reliable hands (as only those with sovereign power could have an archive, and were by definition reliable) (Duranti, 1994).

Professional ethics in regard to digital preservation will help foster a common understanding about questions of morality, concepts of good and bad, and justice which archivists must face when applying their judgment to the preservation of records. In addition to tradition, contemporary ethics play an important part in establishing the trust of the patrons of the archives and the public at large. They help to sustain the public’s faith in the “stewardship of our collective past” (Berger, 2009, p. 60). The professional ethics guiding the risk assessment should include preservation motivations, the responsibilities of the archives, and the guidelines it follows. If archives are to be ethically responsible for collecting and making accessible records of enduring value, and preservation measures must be employed to help them meet this goal, then preservation is a moral responsibility of these institutions (Berger, 2009). Since digital artifacts will deteriorate over time, and because there are not adequate resources to provide the fullest preservation measures possible, the risk assessment of our digital resources should be governed by an overarching moral imperative. Judgments and decisions will have to be made and the risk assessment, grounded in professional ethics, will provide support for those judgments and decisions.

Adhering to these professional ethics will help establish the integrity of the archives, which in turn will instill confidence in the public that the records the archives preserve are authentic. In the risk assessment, the archivist

---

1 See the Association of Canadian Archivist’s *Code of Ethics* at [http://archivists.ca/content/code-ethics](http://archivists.ca/content/code-ethics).
must consider whether the following responsibilities can be met by the archives: can the activity of reproduction be fully documented (that is, how are the copies produced from the record obtained from the creator and what is the impact of the preservation technology); can the identity of the record be ensured and maintained (through for example the preservation of its metadata)?; can the archives ensure that the context of the record is accessible and understandable (such as its provenance, structural and procedural contexts)?; can the archives maintain and demonstrate unbroken custody of the record?; and finally can the archives implement and monitor security controls and procedures? If the answers are yes, then the archives should feel free to proceed with its chosen electronic records management strategy.

As this paper has demonstrated, a fully formulated risk assessment lays out a model whereby an archives is able to verify, against a set of guidelines, the archival quality of a record, whether the record can be retrieved and used, and if, how, and for how long the record can be preserved. The sheer quantity of digital records makes their management an imperative, for records contextualize life in the past and tell the stories of entire societies. Digital preservation helps to ensure the good health of the roots of future memory. Risk assessment is about using past experience to manage the uncontrollable and serving the need to prophesy about the unpredictable. This approach to digital preservation would try to help, in its own small way, to make the process slightly more manageable.
References


