Collaboration as the keystone for successful management of digital records.

Maureen Pennock
Digital Curation Centre,
UKOLN, University of Bath
BATH, UK
+44 1225 386711
M.Pennock@ukoln.ac.uk

This paper was presented at the annual seminar of the International Council of Archives' Section on University and Research Institution Archives (ICA-SUV) in Rejkyavik, Iceland, September 2006. Accompanying slides from the presentation are available at http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/#2006-09.

ABSTRACT

Proper management and curation of digital records is vital for records to remain verifiable, authentic and integrally whole over their life-time, be that five years, fifty years, a hundred years or more. Achieving this is no mean feat and involves not only technical challenges, but also financial, legal, organisational and cultural issues. These exist across several different stages of the data life-cycle. Furthermore, challenges cannot be singularly addressed by a given stakeholder group at a specific stage but require input from several stakeholder groups across different stages to ensure consistent practices and demonstrable continuity of care. This paper introduces the concept of digital curation, i.e. the active management and appraisal of digital data over its entire life-cycle as a framework in which to consistently tackle these challenges, and argues that successful digital curation requires a high level of collaboration by the range of stakeholders with roles across different stages of the curation life-cycle. The paper concludes by identifying and discussing roles and responsibilities of different stakeholder groups across six suggested life-cycle stages.

Keywords
Digital preservation, digital archiving, digital curation; digital records management.

1.INTRODUCTION

The use, management, and preservation of digital records is no longer a new concept. Digital records are created in Higher Education (HE) and other institutions in huge quantities and are used on a daily basis by a wide range of stakeholder-groups. Commercial records management solutions are now widely available and research into long-term technical preservation approaches has taken place across the world. Yet despite this, the all-too-common scenario is that electronic records are printed out to paper for management and preservation. Such an approach denies future re-users the functionality and original digital experience of the original record-creating environment. The remainder of this paper will explore the challenges in electronic records management that lead institutions to rely on print versions of electronic records, and argue that although the technical challenge is a significant one, the cultural and organisational challenges that accompany it are no less significant and can only be adequately solved by input and collaboration from all stakeholders in the record life-cycle.

2.FROM ANALOGUE TO DIGITAL

Digital records are widely replacing the active use of paper-based records, both in terms of digitized and born-digital records. Many administrative records and data are now routinely created and stored in digital form only, and organisations increasingly digitize signed paper documents that are then used for access and easy re-use. This increase in using digital files can be attributed largely to the many advantages that digital records offer over paper based records: search and retrieval is faster, records can be accessed by multiple users simultaneously, physical storage requirements are reduced, and content can more easily be re-used. As such, there is widespread recognition that the use of digital information can generally improve levels of efficiency across an entire organisation. Most organisations have been quick to recognise these benefits and implement technologies for creating, storing, and accessing digital records and information. However, techniques for the

management and preservation of digital records have not kept pace with their usage. As a result, many organisations depend on printed representations of digitally-created or stored records for record-keeping, archival, and legal accountability.

Printing to paper is still widely practised, [1] yet its popularity is gradually falling as more and more users realise that the main advantages of printing to paper – firstly that digital records can simply be incorporated into an existing paper-based system and secondly that the technical challenge of combating obsolescence becomes redundant – are significantly outweighed by the disadvantages. These disadvantages include:

- Loss of the many advantages of using records in digital form, such as simultaneous consultation and rich searching abilities;
- Possible changes in the captured content: not all digital content may be accessible in the printed version, particularly when digital data is layered and views must be altered to access certain types of content;
- Loss of functionality offered by the digital environment, particularly when active links to other records and files are broken;
- Alteration in levels of usability and re-usability;
- Risks to the legal status of the document as a 'record';
- Lack of validation for digital signatures, or failure to capture digital signatures;
- Additional costs, particularly when printed versions are created and managed without destruction of the digital versions;
- Threats to the authenticity and integrity of the record.

Finally, printing to paper fundamentally changes the nature of the record. Digital records are just that: digital. Unless circumstances absolutely necessitate otherwise, digital records should wherever possible be managed and preserved in digital form.

3.SPECIFIC CHALLENGES

Printing to paper does not address the wider set of challenges arising from use and implementation of an electronic infrastructure: digital record-keeping, particularly for the longer-term, is a new paradigm that requires a new approach for success.

The main challenges in managing digital records and data have until recently been perceived as mainly technical. Digital records are a combination of hardware, software, and record file(s). All three elements are required, in a functional state, before the record can be rendered onscreen in a usable manner. The fast pace of developments in ICT means that the technology used to produce digital records and data can quickly become obsolete, and if the software used to create and accurately read files is 'de-commissioned' and no longer available, then access to the information and content stored in the files is at risk. This can become a problem within as little as five years – as Jeff Rothenberg noted at the turn of the last century, 'digital records last forever - or five years, whichever comes first'. [2] This is thus not simply a problem for archivists and material designated for long-term retention: it is a problem for all records with a life-span of as little as five years or more. Addressing this challenge by simply migrating records to a newer file format can cause yet more problems, for any change to the file format can result in a change to the record when rendered onscreen that affects the authenticity and integrity of the records. A paradox thus results: action can mean a change to the record when rendered onscreen, but inaction will eventually result in loss of access to the record and its contents. Because of the seemingly conflicting nature of these last two points, technical preservation action must be planned and properly executed to ensure that records can endure throughout time without damage to their integrity and authenticity.

Similar technical risks are evident for storage media: not only do popular storage media formats change and hardware devices to read older types of storage media become unavailable, (remember the 5.25 inch floppy?) but the rate of bit degradation on digital storage media is often significantly faster than initially anticipated. Such degradation may not be uncovered until it is too late and important data lost. Significant research into these technical issues has now begun to yield usable solutions and ways to avoid obsolescence and deterioration. Despite this, there is no one size fits all solution, as different types of records and contexts have different technical, organisational and authenticity requirements. Research is thus ongoing and only time will tell if current approaches are truly sufficient.

Technical challenges are, however, only part of the problem. Massive organisational and cultural challenges persist that continue to threaten the authenticity, integrity, and long-term availability of usable and reliable digital records. These

challenges range from coping with the sheer scale and number of electronic records, to the relatively lowly position of records- and data- management within organisational infrastructure and project goals. Organisational and cultural infrastructures are frequently not geared towards digital records management, with records created, stored and managed in an ad hoc manner by different employees. Awareness of user and organisational responsibilities outside of the records management section is frequently low, leading to multiple versions of records stored in an adhoc and personal manner with little recognition of an 'authoritative' and final version. Staff can have a strong sense of personal ownership over digital files and what they perceive as the contents of their PC, and this can compound efforts to exercise organisational control. Furthermore, an absence of support and policy from top-level management means that resources and support are often lacking at ground level. Finally, a lack of effective communication and collaboration between staff with responsibilities for records throughout their life-cycle can result in inconsistent and incompatible practices that in turn impact on the sustainability of an authentic record.

In addition to these aspects, legal and financial challenges must also be addressed. Who will pay? What are the cost benefits? Where is the business model? [3] A viable business model is vital to convince policy makers and financial departments of the enduring or legal value of the materials. Steps must also be taken to ensure legal responsibilities can be met, from compliance with data protection and freedom of information legislation to control of intellectual property rights. Digital records are a core business asset and must be recognised as such; a business model and risk assessment can help ensure that resources are available for proper and ongoing digital records management and legal records management responsibilities can be met.

Many of these challenges have not been sufficiently addressed by the research community as their importance has only recently been identified, although initiatives from, for example, the DCC [4] and the University of Kansas [5] are beginning to fill the void.

4.DIGITAL CURATION

Digital Curation offers a framework in which the risks and challenges involved in managing and preserving digital information and records can be coherently and consistently addressed within in a digital environment.

The term 'Digital Curation' is increasingly being used to describe the actions that must be undertaken to maintain digital material over its life cycle so that it remains accessible and re-usable for the future. [6] In the UK, the digital curation movement is spearheaded by the UK Digital Curation Centre (DCC), which supports UK institutions that store, manage, and preserve data (especially research and record-keeping data) to help ensure its enhancement and continuing long-term use. Digital curation encompasses a number of traditional good records management practices but extends beyond the active phase of the records to address preservation, archiving, and re-use.

Digital curation, broadly interpreted, is about maintaining and adding value to a trusted body of digital information for current and future use; specifically, the active management and appraisal of data over its entire life cycle. Recent discussions of this definition have begun to integrate a risk management aspect, insofar as digital curation also involves taking organisational, technical, procedural and other uncertainties and transforming them into manageable risks. Digital curation is distinct from digital preservation and digital archiving in that curation builds upon the underlying concepts of digital preservation whilst emphasising opportunities for added value and knowledge through annotation and continuing resource management. Digital preservation and digital archiving, are activities within the curation life cycle, although all are concerned with managing digital resources with no significant (or only controlled) changes over time.

The life cycle approach is vital to securing the provenance, authenticity, and integrity of records. It ensures continuity and consistency of care by different stakeholders despite technological, organisational and contextual change, whilst maximising investments and potential and resulting in a meaningful chain of custody with compatibility of practices between different stages. However, complete life cycle control can only be truly achieved with a strong level of commitment and collaboration by those involved in the life cycle process.

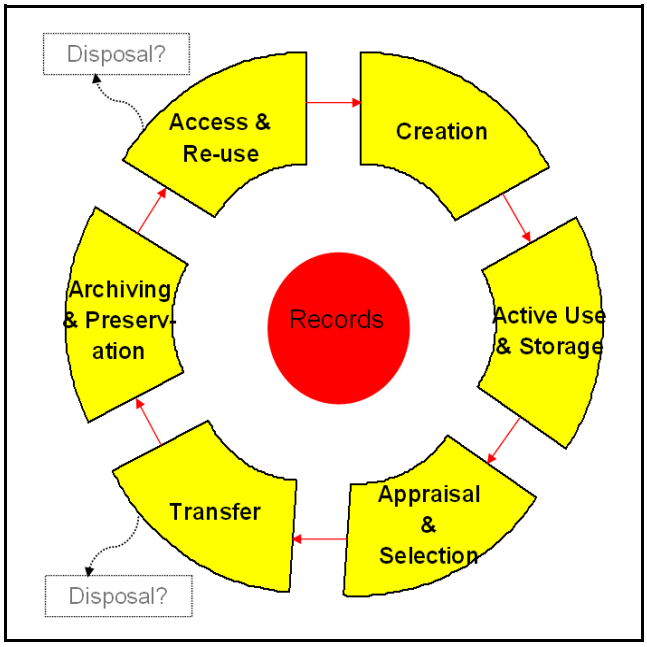

Figure 1: Generic life cycle model

The life cycle model above is a generic model that illustrates the main stages in the record life cycle. This generic model places preservation and archiving within a single and late stage. In practice, preservation activities may be required much earlier in the life cycle during the active use and storage phase, particularly if records are not transferred into an archive until many years have passed.

Different stakeholders have responsibilities for records during different stages of this life cycle. Responsibilities are often cross-stage, in that action may be required that is relevant across numerous stages or which cannot be delayed until the records enter a particular stage. For example, funding and policy development is relevant across a number of stages and requires input from numerous stakeholders. Likewise, development of an appraisal and selection approach for digital records should not be delayed until the records need appraising or selection, nor should planning for a digital archive be delayed until records are ready for transfer. In a similar way, analysis of records during appraisal and technical preservation activities can feed into recommendations for creation practices that facilitate preservation. Communication between stakeholder groups is therefore vital if continuity and compatibility across these stages is to be achieved.

5.SHARED ROLES & RESPONSIBILITIES

As explicit measures need to be taken at every stage of the life-cycle to ensure that records can be considered authentic and maintain integrity over time, communication and collaboration between the different stakeholder groups is vital if they are to be aware of, and fulfil, their responsibilities. The level of collaboration thus required to maintain the authenticity, integrity, and persistence of digital objects is hitherto unforeseen in the field of records and information management, and although this need to collaborate is itself a significant hurdle, it is not insurmountable

If we consider the basic stages of the record life-cycle and the activities within to be creation, active use, appraisal/selection, transfer, archiving & preservation, access & re-use, and possible disposal, and map them against the most likely immediate user stakeholders and their responsibilities at specific stages of the life-cycle, the table below is produced.

	Management	***Creators***	***Curators – RM/Archive***	***Curators - IT***
Creation	Policy on acceptable use	Good practice creation	Good practice creation guidelines	Policy on acceptable use
Active Use & Storage	Policy on responsibilities for personal records/data management	Good practice storage and filing	Good practice storage and filing guidelines	Policy and practice on infrastructure provisions
Appraisal/Selection		Making desktop records/ data available to curators	Policy and guidelines on appraisal/ selection; Advice and guidance to users; Training	
Transfer		Enabling transfer from a 'push' perspective	Effecting transfer from a 'pull' perspective	Providing technical transfer infrastructure
Archiving & Preservation	Policy on data/records archiving & preservation; allocation of finances for archiving/ preservation		Developing archiving/ preservation requirements	Providing archiving/ preservation infrastructure & functionality
Access & Re-use	Policy on data/records access & sharing		Developing metadata-enabled access	Providing reliable access whilst protecting archived records
Possible disposal	Policy on disposal	Desktop disposal of non-transferred records	Identifying retention schedules	Disposing of redundant records/data; deletion of backup data

Readers are referred to the presentation given at the ICA-SUV seminar for more extensive identification of roles and responsibilities at each stage. [7]

This is of course a generic appropriation of roles and responsibilities; individual institutions must apply their own organisational infrastructure and requirements to the tasks identified in the table body. The intention of the table is not to limit such tasks, but to illustrate the interdependencies between groups at different stages that make communication and collaboration between them so vital to the success of an electronic records management and curation strategy.

The table is largely self-explanatory, but a number of issues relating to stages in the table deserve further discussion:

Creation is a particularly important but sadly often underestimated stage of the life cycle. Curation and preservation begin at source, for the way in which records are created can have a direct impact on their sustainability. As the table illustrates, different stakeholder groups have a role to play regarding the creation of digital records. Absent from the table is the role all stakeholder groups play in the availability and take up of training in using records creating software. Basic software training is often overlooked because users already appear competent in using available software. Yet in many cases they have never been properly trained and have simply learned as they worked. This results in bad and inconsistent records creation practice by different staff within the same organisation, practices which adversely impact on preservation activities (especially when preserving large numbers of records). Studies have estimated, for example, the spreadsheet errors in some organisations exist in as many as 80% of the sampled records. [7] Training must be financed, provided, encouraged, and taken-up. This can require input from all four stakeholder groups identified above. Training remains an important aspect in later stages: if a records management system is implemented, for example, then users must be trained to understand the necessity of the system and their role in using it.

Good communication is required between all stakeholders, particularly IT staff, records managers and archivists, to ensure compatibility of practice. Email management is a good example of this: a well thought-out retention policy on email management can be seriously disadvantaged by a corresponding IT policy to delete inbox messages over sixty days old or mailboxes over a certain size. Good communication and training on deletion and disposal can also help ensure that records with no enduring value are properly destroyed at the end of their valuable life. This demands more than simply clicking 'delete' and moving files into the recycle bin: the bin must be properly emptied and all other copies of the records destroyed, including those on backup tapes and those residing on local disks or mobile storage devices. Complete destruction is vital for compliance with records schedules and in legal discovery cases.

Finally, channels for communication between organisations and records re-users must exist so that appropriate re-use facilities can be provided, based on input from users on the types of information they wish to access, how they wish to access it, and what they wish to do with it. Re-use is, after all, one of the main drivers in preserving and curating information in the first place.

6.SUMMARY & CONCLUSIONS

Digital records management is complex and success requires facing up to challenges on a number of different levels, not just technical. Yet that is no reason to rely instead on print-outs. The 'print-to-paper' approach is not an acceptable solution for managing and preserving digital records: it is detrimental to the integrity and usability of digital records and should only be used as a last resort. Digital records are just that: digital. They should therefore be managed and preserved in a digital environment.

The particular vulnerabilities of digital records means that care must be continuously taken to protect the authenticity and integrity of the records, which should be assured and demonstrable from the outset. Any solution should therefore, wherever possible, take a digital and curatorial life-cycle approach that facilitates continuity of care and consistency between activities undertaken at each stage.

The final key to the solution is the development of collaborative relationships between all parties with responsibilities for creation, management, and preservation of digital records and data. Given the complexity of the relationships and the interdependency between activities for which different groups have responsibility, it is clear that only by collaborating together on the tasks can a successful solution be implemented. Finally, collaboration need not be restricted to stakeholders within a single organisation. The challenges of successfully managing and preserving digital information are shared across numerous sectors, including not only the research and academic communities but also government, NGO's, private enterprise, cultural heritage communities and scientific data centres. Although each may have specific sectoral or data-type requirements, the common base elements of the challenge remain the same. Wider cross-sectoral collaboration and knowledge sharing will minimise the effort each organisation and sector must expend and ultimately benefit all involved.

7.REFERENCES

[1] As evidenced in a series of case studies carried out for the ERPANET project, 2001 – 2003.

[2] Rothenberg, Jeff, *Ensuring the Longevity of Digital Documents*. Scientific American, January 1995, pp. 42-47; revised 1999 and re-published by CLIR.

[3] The JISC-funded eSPIDA project has developed a sustainable business-focussed model for digital preservation and asset management, see http://www.gla.ac.uk/espida/index.shtml

[4] The DCC is researching the specific organisational and cultural challenges of digital curation as part of its ongoing research programme. Further information about the DCC research agenda is available at http://www.dcc.ac.uk/research/

[5] Staff from the University of Kansas have published at least two ECAR research bulletins that address organisational and cultural issues arising from their attempt to develop a holistic approach to digital preservation. See . http://www.educause.edu/ResearchBulletins/1007

[6] Beagrie, Neil, *Digital Curation for Science, Digital Libraries, and Individuals,* The International Journal of Digital Curation, Autumn 2006 (Issue 1 Volume 1), pp 3 – 16. http://www.ijdc.net/ijdc/article/view/6/5

[7] Pennock, Maureen *Collaboration for Success in Digital Records Management and Curation* (September 2006), available from http://www.ukoln.ac.uk/ukoln/staff/m.pennock/presentations/#2006-09

[8] Pryor, Louise, *Spreadsheet error rates* http://www.louisepryor.com/showTopic.do?code=errorRates

