Digital Curation and the Management of Digital Library Cultural Heritage Resources.

Maureen Pennock
Research Officer
Digital Curation Centre
UKOLN, University of Bath

Email: m.pennock@ukoln.ac.uk
Web site: http://www.dcc.ac.uk/ http://www.ukoln.ac.uk/

Abstract

This article summarises a presentation given on behalf of LOCSCOT to the 'Google or Bust: Challenging the Profession' CILIPS day event in Dundee on 27 September 2006. The presentation explored certain issues faced by regional and local libraries regarding management, collection and preservation of digital cultural heritage resources i.e. websites and e-journals. The full presentation is available from the DCC and UKOLN websites.

Introduction

The availability and growing use of digital information and services online has occasioned increasing expectations of digital services from cultural heritage institutions. Many local and national libraries have already embraced the initial challenge of providing basic administrative and collection information online. Yet the provision of digital library *objects* online takes the digital library challenge to another level: objects and content must not only be collected or created, but also properly managed, stored, and preserved in order to maximise the initial investment and ensure the objects remain available for users for as long as is deemed necessary. 'Digital Curation' is the way to achieve this, with complete life cycle management to tackle these many challenges in a carefully planned, consistent, and ongoing process.

The UK Digital Curation Centre

In the UK, research and development into digital curation requirements and activities are the core activities of the Digital Curation Centre (DCC). The DCC was founded to support UK institutions which store, manage and preserve digital information, and to help them address the enhancement and continuing availability of digital information for long-term use. As a centre of expertise in digital curation and associated digital preservation, the DCC provides a national focus for research and development into curation issues and promotes expertise and good practice, both nationally and internationally, for the management of digital outputs.

The DCC is a consortium project with four main partners: the University of Edinburgh; UKOLN at

the University of Bath; the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow, and; the Council for the Central Laboratory of the Research Councils, the CCLRC in Didcot. The DCC has a strong research and development agenda and maintains numerous services and outreach activities. These include:

- Organisation of workshops and seminars to explore different aspects of curation such as policies, audit and certification, and legal issues;

- Development of a community-developed digital curation manual to assist data creators, curators and re-users to understand better the challenges they face and the roles they play in creating, managing and preserving their digital information over time;

- Production of briefing papers, technology watch papers, and standards watch papers on issues relevant to digital curation and preservation, as well as the first journal to focus exclusively on digital curation: the *International Journal of Digital Curation*;

- Organisation of the annual DCC Digital Curation conference with peer reviewed papers and posters from an international stage;

- Participation in development of an audit and certification framework for trustworthy digital repositories;

- Provision of an associates network, online forum, and helpdesk service offering on-demand support for individuals and UK HE/FE institutions doing preservation planning and solving challenging curation problems.

The DCC is funded by the Joint Information Systems Committee (JISC) and the e-Science core programme. More information about the DCC is available from the DCC website.[1]

What is Digital Curation?

As the term 'curation' has only recently begun to be applied to digital materials, we will at this point take a moment to define what we mean when talking about 'digital curation'. Digital curation, broadly interpreted, is about maintaining and adding value to a trusted body of digital information for both current and future use: in other words, it is the active management and appraisal of digital information over its entire life cycle.

---

[1]   See http://www.dcc.ac.uk
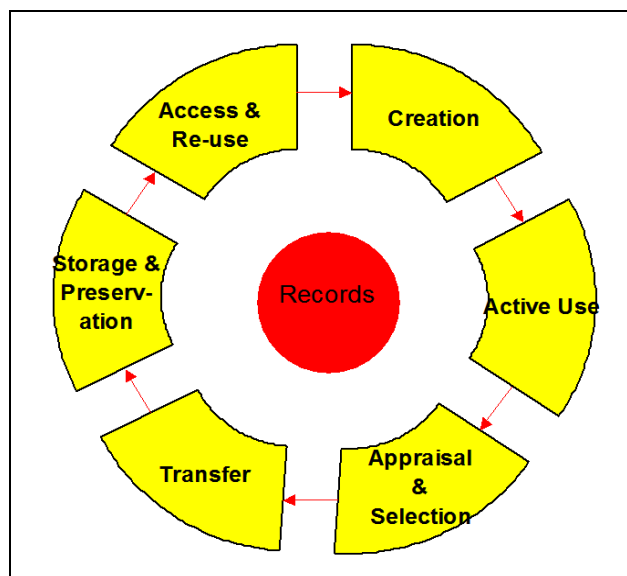
Figure 1: Sample life cycle model

The life cycle approach is necessary because:

- Digital materials are fragile and susceptible to change from technological advances throughout their life cycle, i.e. from creation onwards;

- Activities (or lack of) at each stage in the life cycle directly influence our ability to manage and preserve digital materials in subsequent stages;

- Reliable re-use of digital materials is only possible if materials are curated in such a way that their authenticity and integrity are retained.

Approaching digital information management from a life cycle perspective facilitates continuity of service. This in turn supports verification of the provenance of digital resources despite technological and organisational changes in their context, and helps to maximise the initial investment made in creating or gathering them.

Complete life cycle control is an ideal scenario. However, it is not always possible. Many cultural heritage organisations manage and host resources created by external parties; in such instances it is not always possible to exercise influence over creation and external resource management practices. The active area of life cycle management for such heritage organisations is therefore reduced and can begin only at the point of appraisal or transfer. This scenario exemplifies how responsibility for curation is shared between different stakeholders, particularly when dealing with transfer of ownership. Communication between these stakeholders and documentation of activities is key to achieving life cycle continuity in such a scenario.

Issues in developing a digital library collection

The types of digital information held by digital libraries can vary considerably. It may include, for example, relatively straightforward images and textual documents, or more complex audio-visual material and websites. Such items may be digitised from an existing original analogue source, or are 'born-digital'. Whatever the nature of the collection item, it should always be accompanied by metadata that describes the resource, helps to manage and preserve it, and facilitates its re-use. Both collection items and metadata need to be properly managed across the entire life cycle.

This is not a simple task; nor merely a technical one. Whilst certain aspects are certainly technical, others are cultural and organisational, whilst others are of a legal or financial nature. For example:

- Organisational and cultural infrastructure may not be geared towards digital longevity: resources are often created for the here and now, but such practices can be incompatible with the sustainability of authentic resources. Addressing this issue requires not just the proper technical storage infrastructure, but also education, training, managerial support, and dedicated funding;

- Curation and preservation is an ongoing commitment, but what of funding? What are the cost benefits in proper management of digital information? Is there a business model, and what are the options for revenue generation?

- The copying and provision of information online must comply with legal requirements, including Intellectual Property legislation and the Data Protection Act. Which legislation affects the creation of digital surrogates, and how? What solutions would support legal compliance whilst still providing access to the materials in digital form?

- What tools are available to ease the processes of harvesting, acquisition, ingest, indexing, and preservation? Which storage mechanism is most suitable, what metadata are needed, and how can the threat of technological obsolescence be addressed?

The active management and appraisal of digital information over its life cycle – in other words, digital curation – requires that these risks are addressed. Forward planning and an understanding of the life cycle elements is key to achieving this, and a proactive approach is better than a reactive approach - although in practice elements of both are usually required. Each stage in the life cycle must be broken down so that the challenges relevant for that stage are identified and tackled. To complicate matters, challenges cannot necessarily be mapped to a single stage of the life cycle but may persist across different stages: for example, preservation is not simply a storage challenge and may require attention during earlier stages of the life cycle, depending upon time periods and resource availability. Funding and organisational issues such as infrastructure and training can also require attention at several different stages.

Useful projects and initiatives

Developing and implementing a curation strategy is not an easy task; yet neither is it one that each organisation must tackle alone. The library, archives, curation and preservation communities have already invested a great deal of time and effort in tacking these solutions. National institutions are a very useful starting point for organisations starting to address curation, and many national libraries have undertaken initiatives for preservation and accessibility upon which local and regional libraries can draw. A number of further national and international initiatives can also prove useful, including:

The Internet Archive, a non-profit organisation based in the US that crawls the Internet to collect websites and make them available to the public through their Internet site. Harvesting takes place at regular intervals and results in a snapshot of site evolution over time.

The International Internet Preservation Consortium (IIPC), building cooperation between the Internet Archive and national and research libraries. The IIPC is developing an open source web archiving toolkit to support acquisition of websites, collection management, storage & maintenance, and access & finding aids.

The UK Web Archiving Consortium (UKWAC), which collects selective web content relevant to the UK.

The [LOCKSS Project](#) (Lots of Copies Keeps Stuff Safe), an open source solution, harvesting and preserving subscription e-journals through duplication of copies within a prescribed network.

Finally, the [Digital Curation Centre](#) is the UK's premiere source of information on curation and preservation. In addition to developing guidance and advice of its own, the DCC resource centre also includes one of the largest virtual collections of international curation and preservation resources.

Addressing the issues

Recommendations for dealing with some of the challenges introduced above are listed below. This is not an exhaustive list, but is intended to provide a starting point for preservation and curation planning:

- Develop a written policy and strategy to support activities and help secure resources

- Take a life cycle approach to support curation and preservation planning

- If creating resources, provide good practice guidance for sustainability (eg when digitising or accepting digitised resources)

- Assess collection/selection criteria – are they still valid? Do they need expanding? Identify possible resources. (Remember that digital resources can complement and enhance physical ones and do not have to act as replacements.)

- Be aware of externally produced digital resources (eg websites); check other heritage collections before gathering!

- Identify legal restraints in collection/management/access

- Consider whether value can be added to resources during acquisition?

- Store objects in a secure environment

- Plan for preservation activities to maintain access to authentic resources over time and avoid incurring extra costs

- Determine access and user requirements

- Implement an integrated approach to collection accessibility

- Adapt and learn from national and other leading activities

Conclusions

Digital curation is vital if resources created now are to be properly preserved and remain accessible for future generations. This is more than simply a technical challenge and requires commitment at a multitude of levels. This brief article has, I hope, provided some insight into the types of challenges involved and some ways in which they may begin to be addressed. The most useful of these is based on that most fundamental aspects of the human condition: communication. Digital curation and preservation is an issue for almost any organisation with a responsibility for managing digital information. In the age of the computer, that excludes very few! Whilst these organisations currently have varying degrees of experience, they often have experience in different areas. Communicating to exchange and share different experiences across a network of curation and preservation organisations can therefore be one of the most fruitful and effective ways to move the practice of digital curation and preservation forwards.