Digital Curation: The Emergence of a New Discipline

Sarah Higgins,
Lecturer in Archives Administration and Records Management,
Aberystwyth University

Abstract
In the mid 1990s UK digital preservation activity concentrated on ensuring the survival of digital material – spurred on by the US report *Preserving Digital Information* (The Task Force on Archiving of Digital Information, 1996) and developed through JISC-funded activities. Technical developments and a maturing understanding of organisational activity and workflow saw the emphasis move to ensuring the access, use and reuse of digital materials throughout their lifecycle. Digital Curation emerged as a new discipline supported through the activities of the UK’s Digital Curation Centre and a number of EU 6th Framework Projects. Digital Curation is now embedded in both practice and research; with the development of tools, and the foundation of a number of support units and academic educators offering training and furthering research.
Introduction

As digital material has become increasingly ubiquitous in the day-to-day lives of normal people, the realisation that it needs to be carefully managed to ensure its survival and continuing access has gradually grown. In the UK cultural and educational sectors, digital preservation efforts originally focussed on ensuring that material survived technical obsolescence and organisational mismanagement. Preservation implied a passive state, where material would be mothballed in an inaccessible “dark archive”, with only a few authorised users, to ensure that it retained its integrity and authenticity. Over the last few years, the focus has shifted to ensuring that digital material is managed throughout its lifecycle so that it remains accessible to those who need to use it. Metadata is used to both improve accessibility and discoverability; and to control authentication procedures, creating audit trails to ensure that material cannot be accessed or altered by those not authorised to do so. Digital material is actively preserved, used and reused for new purposes, creating new materials. This is Digital Curation: the management and preservation of digital material to ensure accessibility over the long-term (Abbott, 2008).

Preservation Beginnings

The clarion call in the US was Preserving Digital Information, the report of the US Task Force on Archiving of Digital Information (Task Force on Archiving of Digital Information, 1996), which called for the development of strategies to ensure the survival of culturally valuable digital information. It emphasised the imperative to organisations of undertaking digital preservation activities, and explored the roles and responsibilities for managing a digital archive, migrating the material to guard against obsolescence, and the costs associated with these. The report aimed to create international dialogue, and kick started the development of tools and methodologies for digital preservation activities. Margaret Hedstrom, a member of the Task Force, called for the digital library community to take on the challenge of developing and improving digital preservation research and techniques (Hedstrom, 1998).

Developing a UK Digital Preservation Infrastructure

Digital preservation came to the fore in the UK in the mid 1990s with a series of high level activities. These examined the organisational challenges involved in maintaining access to digital materials, and concentrated on building a sustainable collaborative support infrastructure. The draft of US Task Force’s report informed discussions at a workshop on the Long Term Preservation of Electronic Materials, organised by eLib¹, at the University of Warwick in 1995. This explored a number of strategic issues relating to digital preservation, including policies, responsibilities and methods for practical implementation, with 18 action points identified (Marc Fresko Consultancy, 1996). To address these, a programme of seven studies was funded by the Joint Information Systems Committee (JISC) (eLib, 2000), with the Digital Archiving Working Group created to oversee them. A synthesis of the studies concluded that there was a need for: awareness raising; collaboration between organisations; and improved strategies, guidelines, criteria and checklists (Feeney, 1999).

¹ Electronic Libraries programme: http://www.ukoln.ac.uk/services/elib/.
A second Warwick Digital Preservation Strategy Workshop in 1999 focussed on setting the agenda for the preservation of research information (JISC/NPO, 1999). The implications of failing to manage data were highlighted (Ross, 2000), and presentations were given by the National Data Centre at ULCC (University of London Computer Centre) and the Essex Data Archive (now the UK Data Archive). Final recommendations identified the need for cost analysis, research into rights issues and increased awareness raising (Cedars Project, 1999). The key recommendation was the establishment of a central body to act as a focus for leadership and collaboration in digital preservation activities. This recommendation led to the formation of two bodies. In 2000 the JISC Digital Preservation Focus was established to coordinate and disseminate research and best practice within a long-term strategy for the UK Higher Education community. This was followed in 2001 by the Digital Preservation Coalition (DPC), a cross-sectoral membership organisation, which aimed to “develop a UK preservation agenda within an international context” (Beagrie, 2001a).

Exemplar Projects

The UK infrastructure was now in place to move the preservation agenda forward. This was reinforced by a number of ongoing exemplar projects, and the developing data repositories at ULCC, the UKDA and AHDS. The influential CEDARS Project tested the recently published OAIS Reference Model (ISO 14721, 2003) to establish its applicability as a common framework for digital preservation applications, and created the first coherent metadata set specifically for digital preservation activities. The CAMiLEON Project evaluated emulation as a long-term strategy, with the rescue of the BBC Domesday Project as its main proof of concept. Meanwhile JISC was one of the funders of a series of research projects examining different aspects of the practicalities of digital preservation, such as web archiving (Day, 2003), legal implications (Charlesworth, 2003) and e-prints (James, Ruusalepp, Anderson & Pinfield, 2003). The establishment of the international Digital Preservation Award by the DPC in 2004 celebrated the achievements of “those people and organisations that have made a significant contribution to ensuring that we can have long term access to digital data” (DPC, 2010).

The Curation Agenda

By this time the wider remit of maintaining persistence and access to digital material was being investigated by a number of sectors. The concept of digital continuity – “the ability to use your information in the way that you need, for as long as you need” (The National Archives, 2011) – for electronic journals and archival material was being discussed in archives and libraries, particularly in The Antipodes. This was typified by the formation of the Australasian Digital Continuity Forum in 2001 (Swinburne University of Technology, 2001). In 2002, JISC’s Continuing Access and Digital Preservation Strategy (Beagrie, 2002) set out a major funding stream to cover the next three years. The curation of data was a major discussion point, with the first of the resulting nine key initiatives being the establishment of a Digital Curation

---

3 University of London Computer Centre: [http://www.ulcc.ac.uk/](http://www.ulcc.ac.uk/).
4 UK Data Archive: [http://www.data-archive.ac.uk/](http://www.data-archive.ac.uk/).
5 Arts and Humanities Data Service: [http://www.ahds.ac.uk/](http://www.ahds.ac.uk/). Funding for AHDS ceased in April 2008.
6 The CEDARS Project ran from 1998-2002. It was funded by JISC as part of the eLib Programme and managed by CURL (the Consortium of University Research libraries).
7 The CAMiLEON Project: [http://www2.si.umich.edu/CAMILEON/](http://www2.si.umich.edu/CAMILEON/).
Centre (DCC) to provide a central focus, and develop tools and services for the curation of research data.

The establishment of a digital curation centre was endorsed in a 2003 report on the curation needs for e-science (Lord & Macdonald, 2003), which identified this as one of the main players for achieving their recommendations. Following a tendering exercise, the Digital Curation Centre was launched in 2004 as a collaborative distributed service (Digital Curation Centre, n.d.a).

In 2005 the DCC was one of the organisers of the third Warwick meeting, Digital Curation and Preservation: Defining the Research Agenda for the Next Decade (Pothen, 2006), which looked at technological, process and policy issues for both curation and preservation. The meeting highlighted the need for greater understanding of curation processes, better funding for initiatives, and education and training for curation. These recommendations fed into the activities for the first three year phase of the DCC, who embarked on a varied programme of research, advice, technical development and community development.

Curation Projects

Around the same time the EU 6th Framework Programme funded a number of high profile projects to develop tools and methodologies, with the core aim of maintaining access to digital material. Partners of the DCC fed into a variety of the work packages for DigitalPreservationEurope (DPE), the PLANETS Project (Preservation and Long-term Access through NETworked Services) and the CASPAR Project (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval)8. Along with the DCC, these served to both raise the profile of digital curation and extend the UK skill set.

The Digital Curation Centre’s First Phase

During the first funding phase (March 2004 – February 2007) the DCC’s outreach activities concentrated on building an understanding of the discipline and creating a community which straddled the higher education, commercial and public sectors. It took a holistic approach to data, considering all material created digitally to be within its remit. This included both digital copies of analogue material, text-based materials and scientific research data. Information days explaining the basic concepts of digital curation were held at venues around the country9 and an Associates Network was established for discussion, information and support, “providing a forum for cross-sectoral communication on important problems” (Digital Curation Centre, n.d.a). A programme of publications included the Digital Curation Manual10, and a series of Briefing Papers explaining the basic activities and responsibilities for successful curation11. The International Digital Curation Conference (IDCC) was established. The first conference, held in September 2005, was concerned with establishing definitions of Digital Curation and possible approaches for undertaking it (Digital Curation Centre, 2005). By the next year increased community understanding of the discipline and practical activity allowed a conference theme which considered “Digital Data Curation in Practice” focussing on case studies from both the library and archives and the research data domains. This theme continued into the third conference, in

---

8 WePreserve: http://www.wepreserve.eu/about/.
9 Other DCC Events: http://www.dcc.ac.uk/events/other-dcc-events.
11 DCC Briefing Papers: http://www.dcc.ac.uk/resources/briefing-papers.
December 2007, where the emergent research data emphasis of the discipline was reflected in the programme (Digital Curation Centre, 2007). The foundation of the peer-reviewed International Journal of Digital Curation (IJDC) in 2006 firmly established digital curation as an academic discipline.

**The Digital Curation Centre’s Second Phase**

The second phase of the DCC (March 2007 – February 2010) concentrated more firmly on the curation of research data, and the unique challenges associated with this. The new focus was reflected in the new strapline: “Because good research needs good data.” In the UK this complimented the archives, libraries and museums focus of the Digital Preservation Coalition. The DCC’s eScience liaison function was instituted and case study work on research data curation activities was undertaken with both the SCARP Project and in conjunction with the Research Information Network (Research Information Network & British Library, 2009). The activities for successful curation were articulated in the DCC Curation Lifecycle Model (Higgins, 2008) and work to define model research data management plans (Donnelly & Jones, 2009). These formed the basis of the Digital Curation 101 training activities, which have been undertaken worldwide over the last few years.

During this phase of the DCC, the UK Arts and Humanities Research Council gave ten months notice that funding would be withdrawn from the AHDS, a distributed repository which existed to “collect, preserve and promote the electronic resources which result from research and teaching in the arts and humanities” (AHDS, 2008). This threw the digital sustainability of research data into the spotlight, underlining the need for a wider understanding of long-term commitment to curation activities across the research community. In response, research libraries identified themselves as trustworthy organisations with a remit to preserve material and make it available. The feasibility of developing a UK Research Data Service (UKRDS) began to be investigated as a coherent approach to research data management, which would build on the DCC’s expertise (UKRDS, 2008 & 2010). This was proceeded by two JISC-sponsored investigations into the costs and activities required, and the benefits which might accrue from an orchestrated approach (Beagrie, Chruszcz, & Lavoie, 2008; Beagrie, Lavoie & Woollard, 2010).

**Curation and Preservation Tools**

The maturation of a number of tools provided methodologies for undertaking curation. The DCC and the Humanities Advanced Technology and Information Institute (HATII) published the Data Asset Framework (DAF) which provided a method for enumerating and auditing data holdings. This was trialled in four successful UK pilot audits. The DCC and Digital Preservation Europe (DPE) collaborated to produce DRAMBORA a tool for self-assessment of possible risks to the sustainability and continuity of digital repositories, which was trialled in a number of international institutions (Digital Curation Centre, n.d b). This fed into ongoing international work to develop a standard for the audit of digital repositories and their certification, along with TRAC, a methodology developed by the Online Computer Library Centre (OCLC) and the Center for Research Libraries (CLR), to certify the

---

15 ISO 16363: Space data and information transfer systems – audit and certification of trustworthy digital repositories. The standard is still under development.
effectiveness of digital repositories (OCLC & CRL, 2007). The Digital Preservation Suite developed by the PLANETS Project provided tools to support the implementation of preservation plans, a testbed environment and an Interoperability Framework. Meanwhile, the National Archives won the Digital Preservation Award for DROID, a tool which identifies file formats for preservation, against the master PRONOM database (Digital Preservation Coalition, 2007).

The Digital Curation Centre – Start of the Third Phase

By the start of the DCC’s third phase in 2010, the discipline of Digital Curation was clearly defined and an active international community of practitioners had coalesced. This was demonstrated by demand for the 6th International Digital Curation Conference (IDCC) in December 2010 where attendance increased from an average of around 150 delegates to 250.

Curation Goes Mainstream

The framework of support, tools and research developed, since the first Warwick Workshop in 1995, has enabled the discipline to grow over the last few years through the establishment of a number of other digital curation support organisations. These have different focuses and different motivations. Some have an internal agenda of maintaining digital assets for their host institution, while others focus on education, research and development. The driving forces are varied, including: the requirement for back-ups of print publications; building the reputation of the institution which hosts them; the ability to maintain records as evidence; educational re-use of digital resources or the aggregated value of maintaining institutional datasets (Ashley, 2010).

Notable digital curation support organizations that have been established in the last four years include:

- The Greek Digital Curation Unit (DCU) at the Athena Research Centre was established in 2007 to “act as a national focus point in the field of digital curation”. It serves a diverse constituency, publishes widely and is a partner in the EU-funded DARIAH Project (Digital Research Infrastructure for the Arts and Humanities), which seeks to support the digitisation of arts and humanities data across Europe. It also provides the MOPSEUS Digital Library Service, a digital repository for small scale organisations who don’t have the infrastructure to develop their own.

- The University of California Curation Center (UC3) was formed in 2010 as a partnership supporting the ten University of California campuses. Its remit is one of inward looking academic support, for “campus constituencies – museums, libraries, archives, academic departments, research units, and individual researchers – to have direct control over the management, curation, and preservation of the information resources underpinning their scholarly activities.”

- The Digital Research and Curation Center at The Johns Hopkins University’s Sheridan Libraries undertakes internal research and development regarding automated tools and effective workflow to ensure

---

18 University of California Curation Center: [http://www.cdlib.org/services/uc3/](http://www.cdlib.org/services/uc3/).
19 Digital Research and Curation Center: [http://ldp.library.jhu.edu/dkc](http://ldp.library.jhu.edu/dkc).
The University of Toronto’s iSchool established The Digital Curation Institute with a broad information sciences research agenda. It launched with a conference in June 2010, with research papers from a mixture of guests, staff and students.

Purdue University Library’s Distributed Data Curation Center (D2C2) researches curation solutions for complex research data. It is undertaking the Data Curation Profiles Project, which provides a toolkit for structuring diagnostic case studies prior to curation in an academic library (Witt, Carlson, Brandt & Cragin, 2009).

**Curation Education**

The skill base of the digital curation community continues to be developed through training and higher education programmes worldwide. A number of these have formed the International Digital Curation Education and Action (IDEA) Working Group – an international alliance examining and advising on curriculum needs to continue building the skill base. This originated in the DigCCurr Project at the University of North Carolina at Chapel Hill, which examined the curriculum requirements for digital curation training and held conferences and symposia concerning their development. Chapel Hill students can now study for a Digital Curation Postgraduate Certificate. Masters-level study in Digital Curation can be undertaken at Luleå University of Technology in Sweden, and options on Digital Curation are available as part of a number of higher education information sciences degrees worldwide. The subject is now becoming a mainstream part of an information science education and the 2011 International Federation of Library Associations (IFLA) Conference will have an open session on ‘Education for Digital Curation’ (IFLA, 2011).

**Conclusions**

Digital Curation emerged as a new discipline through the iterative workshop and agenda setting process. In the UK the strategic emphasis for long-term management of digital material gradually moved from passive preservation to active curation. After a period of definition and consolidation, the subject now boasts a growing international professional base, a developing research agenda, practical tools and collaborative projects and a workforce trained to Higher Education level.

**References**


20 Digital Curation Institute University of Toronto: http://dei.ischool.utoronto.ca/index.html.
21 Distributed Data Curation Center: http://d2c2.lib.purdue.edu/index.php.
23 DigCCurr: http://ils.unc.edu/digccurr/.
24 Masters Programme in Digital Curation, Luleå University of Technology: http://www.ltu.se/edu/program/FMDBA/?l=en.


