Digital Curation 101

WHAT IS DIGITAL CURATION?

Topics:
- What is digital curation?
  - The background
- What digital curation is not
- Digital curation defined

What is digital curation?

The Digital Curation Centre (DCC) defines digital curation briefly in this way:

> Digital curation, broadly interpreted, is about maintaining and adding value to a trusted body of digital information for current and future use.

The background

It is useful to understand some of the background to digital curation. There were two drivers to the establishment of the Digital Curation Centre:

- e-Science, the 'data deluge' and continuing access to the data sets generated
- digital preservation, particularly the realisation that digital preservation activities were by themselves insufficient to address many of the issues associated with maintaining data over time.

In the UK the result was the development by JISC (the Joint Information Systems Committee) of a Continuing Access and Digital Preservation strategy (2002)[1] and Lord and MacDonald's report into data curation for e-science (2003)[2], leading to a call in 2003 for bids to establish a Digital Curation Centre. This was established in 2004.

The DCC's brief definition needs expanding. Digital curation has different meanings for different people. Some are reported by Brophy and Frey[3] as:

- Maintaining the links between digital information and associated annotations or published materials, including citations

---

[1] http://www.jisc.ac.uk/whatwedo/programmes/preservation/continuing

[2] P. Lord and A. Macdonald, e-Science Curation Report, http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf

[3] P. Brophy and J. Frey, Digital Curation Centre Externally moderated reflective self-evaluation: Report 2006
http://www.dcc.ac.uk/docs/DCC_Evaluation_Report_Final.pdf

Digital Curation 101

- Ensuring the long-term accessibility and reusability of digital information
- Performing archiving activities on digital information such as selection, appraisal and retention
- Ensuring the authenticity, integrity and provenance of digital information are maintained over time
- Performing preservation activities on digital information such as migration or emulation
- Maintaining hardware components to enable digital information to be accessed and understood over time
- Managing digital information from its point of creation
- Managing risks to digital information
- Ensuring the destruction of digital information.

These all contain elements of digital curation, but none of them presents the whole picture. So we are still left with the question: what is digital curation?

What digital curation is *not*

We can say with some certainty what digital curation is *not*.

1) It is not *digital archiving* – 'the process of backup and ongoing maintenance as opposed to strategies for long-term digital preservation' [4]
2) It is not *digital preservation* – 'all of the actions required to maintain access to digital materials beyond the limits of media failure or technological change'[5]

Although digital archiving and digital preservation are important aspects of digital curation, they are not the whole story. Digital curation is a more inclusive concept – it is concerned with the whole range of processes applied to data *over its life cycle*, starting before that data is created by setting standards for 'good' preservable data and in systems design. It emphasises *adding value* to data sets so they can be reused. It involves a *wide range of stakeholders* cutting across disciplinary boundaries: funders, government bodies, national data centres, institutional repositories, and learned societies, as well as cultural heritage organisations such as libraries and archives. Digital curation is also concerned with *risk management*: it 'is about converting uncertainties into measurable and manageable risks' (Seamus Ross, Associate Director DCC[6]). It is also about *good data management* practices.

---

[4] as defined by the Digital Preservation Coalition http://www.dpconline.org/graphics/intro/definitions.html
[5] ibid
[6] http://www.repositoryaudit.eu/img/drambora_flyer.pdf

Digital Curation 101

Digital curation is more concerned with digital material that is *linked* to other resources in a range of formats, *large science datasets*, or *datasets that are changing* by the second*, than it is with relatively simple data such as the static documents handled by libraries and archives. Most data archiving and digital preservation practices were developed for static documents, but they do not transfer successfully to more complex data. In this new era of data-driven scholarship and research, new strategies and processes are needed to handle the wide range of data created and maintained by many different kinds of user communities.

Digital curation defined

Taking all of this into account, an expanded definition of digital curation could be:

> Digital curation is concerned with actively managing data for as long as it continues to be of scholarly, scientific, research and/or administrative interest, with the aim of supporting reproducibility of results, reuse of and adding value to that data, managing it from its point of creation until it is determined not to be useful, and ensuring its long-term accessibility and preservation, authenticity and integrity.