What is digital curation?

By Daisy Abbott, University of Glasgow

1. Introduction

There is an ever-increasing amount of data being created in digital formats, through the digitisation of existing analogue information and the creation of new 'born-digital' data from the sciences, arts, and humanities sectors. As well as generating new digital data, scientists, researchers, and scholars have begun to rely on digital content created by others. These data are at risk from technological obsolescence and from the inherent fragility of digital media. Digital curation is the management and preservation of digital data over the long-term.

All activities involved in managing data from planning its creation, best practice in digitisation and documentation, and ensuring its availability and suitability for discovery and re-use in the future are part of digital curation. Digital curation can also include managing vast data sets for daily use, for example ensuring that they can be searched and continue to be readable. Digital curation is therefore applicable to a large range of professional situations from the beginning of the information life-cycle to the end; digitisers, metadata creators, funders, policy-makers, and repository managers to name a few examples.

2. Short-term Benefits and Long-term Value

Digital curation ensures the sustainability of data in the long term, however it has immediate value for data creators as well as users. Digital curation facilitates:

- Persistent access to reliable digital data
- Improved quality of the data itself and its research context
- The use of common standards across different datasets, which in turn leads to more opportunities for cross-searching and collaboration
- Authenticity checks, improving the trustworthiness of data
- Ensuring data is valid as a formal record where appropriate, meaning it can function into the future as legal evidence
- Exploiting initial investment by ensuring that data is available for use and re-use and protecting the financial value of information
- Improved speed and range of access, data sharing and analysis opportunities, and other research benefits

Digital curation ensures long-term value by:

- Preserving data and protecting it against loss and obsolescence (particularly crucial where the data is non-reproducible or extremely valuable)
- Allowing continued access to data despite short-term funding or institutional changes
- Encouraging re-use of data
- Maximising the exploitation of digital materials cumulatively through time

- Providing information about the context and provenance of data
- The use of tools and services to migrate data, metadata, and other representation information into new formats to ensure it remains meaningful to users
- A management infrastructure for the preservation and dissemination of data in perpetuity

3. HE/FE Perspective

"*It is clear that e-learning materials, in common with other digital resources, require investment to create and exploit. These materials can be migrated, versioned, updated, reinterpreted or re-visioned to make them applicable to new teaching and learning scenarios. A better understanding is required of the value of preservation and curation methods to leverage that original investment by re-using existing resources in novel ways.*"

— JISC ITT: e-Learning Materials Preservation and Curation Studies , February 2008.

4. e-Science Perspective

"*A coherent strategy is essential in order to establish and sustain the UK as an international leader of well-curated national data assets and computational infrastructure expertly used to shape policy, support decisions and empower research. The value of data as a foundation for wellbeing and a sustainable society must be appreciated; National resources must be more wisely directed to the collection, maintenance, widening access, analysis and exploitation of these data.*"

— M. Atkinson, M. P., Britton, D., Coveney, P., De Roure, D. E., Garnett, N., Geddes, N., Gurney, R., Ingram, D., Haines, K., Hughes, L., Jeffreys, P., Lyon, L. J., Osborne, I., Perrott, R., Procter, R. N. and Trefethen, A. E. (March 2008). "Century-of-Information Research — a Strategy for Research and Innovation in the Century of Information" (CIR3).

5. Issues to be Considered

- Digital curation can be costly and requires a significant level of time investment and expertise. This can be problematic for smaller institutions, especially as the major advantages of digital curation are long-term and investment can take many years to bear fruit.
- Digital curation is an ongoing process not a one-off action. It is a chain of activities only as strong as its weakest link. It is necessary to have the appropriate financial and policy infrastructures in place to ensure that digital curation itself can be continued over the long-term.
- Best practice changes as new tools and standards are developed, and as digital media becomes increasingly complex. Digital curators must engage with technology and standards watch activities in order to keep their knowledge up to date in a quickly changing sector. In this respect, training can also form part of the activities involved in digital curation.
- The responsibilities involved in digital curation can be shared across different institutions and communities and change over the life-cycle of the data, often incorporating organisational and cultural issues as well as technical ones. There is often confusion surrounding the specific roles that various stakeholders play in the digital curation life-cycle. As such, disambiguation is urgently required.
- Digital curation offers a great number of opportunities for collaboration in data use, and the process itself can benefit from bringing together different expertise, contributors, and sharing the financial burden. However, these activities need to be carefully managed to ensure consistency of approach across different partners.
- One of the major advantages of making digital data widely available is that it can be searched, however studies have shown that much data available online is not used due to the difficulties users have in locating datasets and searching within them. How can data from diverse, varied

collections be meaningfully combined to provide faster, more accurate searching, and new research findings?
- The ownership of digital data is particularly complex as creators of both digital objects and analogue originals, databases, metadata, tools for functionality and contextual information can all have rights over the materials, as can the digital curators themselves. Managing rights is a challenging and time-consuming aspect of digital curation.
- Mechanisms for quality control, authentication, and validation of data should form part of digital curation.
- Backup of data should not be confused with digital curation. Backups are short-term recovery solutions and do not include the wide range of activities necessary to ensure the preservation and dissemination of authentic digital objects in perpetuity.
- Different disciplines use terminology in different ways which can lead to inconsistencies and/or misunderstandings between collaborators on digital curation.

6. Additional Resources

- DCC Resource Centre
- AHDS Guides to Good Practice, Information Papers, and Case Studies
- JISC Briefing Paper "Continued access to authentic digital assets"
- Digital Preservation Coalition
- ERPANET
- International Journal of Digital Curation
- SHERPA DP
- LIFE Project
- Draft DCC Digital Curation Lifecycle Model

Digital Curation Centre
Appleton Tower, 11 Crichton Street, Edinburgh, EH8 9LE | t. +44 (0)131 651 1239

The DCC is funded by Joint Information Systems Committee