

Challenging Role of Librarians in Digital Data Curation

Devaraj Devarasaiah

Stanley Madan Kumar

Abstract

Digital curation involves maintaining, preserving and adding value to digital research data throughout its life cycle. It emphasizes on long term preservation and persistent access. Care should be taken for data security, integrity and fair use. Digital data degrades more quickly than paper documents. As such, curation is not a one-time preservation process but a perpetual one. The paper discusses the role of library professionals and the challenges they are to confront in digital data preservation process.

Keywords: Curation, Data Preservation, Data Re-Use, Preservation Paradox

1. Introduction

Curate is a Latin term meaning to arrange, to look after, to take care of and to present. From the common man's point of view curation has mostly been the business of museums, art galleries and archives. Conceptually, curation has been the core business for libraries. For ages, librarians have been collecting, describing and organizing information records to ensure ease of retrieval and access to the information contained therein through library classification schemes and cataloguing codes.

Since last couple of decades, with the advent of computers and software, librarians began using them to automate and speed up the reference services. Current Awareness Services (CAS) and Selective Dissemination of Information (SDI) evolved as the computer usage allowed customization of information services for specific users and user communities (Oldham,2011).

2. Evolution Of Research Data

Gathering research data is time consuming and costlier affair. However, making the data available at the right time and in right format to the research community at large not only enhance the knowledge on the subject with increased output but also enhances the rate of return on research investment.

Until very recent, the output produced by scientists and scholars were available on as textual on paper or digital version of the same. Librarians painstakingly collected these precious items of data that could easily be stored in a couple of notebooks. As such, it was natural that libraries would curate these documents. In the present digital era majority of the knowledge output is neither text nor paper or book or journals. Instrumentation and computerization enable scholars to collect voluminous data within a short time (Heidorn, 2011)

3. Importance of Data Re-Use

Knowledge of on-going trends and user demands for digital information, including the information deluge are so dramatic that they continue to put



substantial pressure on the present information infrastructure and working practices.

In academic research, data re-use is of considerable importance as it allows –

- ❖ Subsequent research to be carried out without the time and cost overheads.
- ❖ Data from a number of different sources to be combined to give a wider data set.
- ❖ Teaching to be carried out based on real data sets.
- ❖ Greater inter-disciplinary and cross-disciplinary interaction.
- ❖ Future re-analysis of data with alternative or improved techniques (Goble et al, 2006).

Data archiving is a process where data is simply turned over to a repository at the conclusion of a study/project. In fact, data archiving should begin early in a project and incorporate a schedule for depositing products over the course of a project's life cycle and the creation and preservation of accurate metadata, ensuring the usability of research data (Jacobs & Humphrey, 2004).

4. Obstacles in Data Collection

The challenges of digital curation of scientific data lie in the need to preserve not only the data set itself but also its ability to deliver knowledge to a future user community. A true scientific research asset allows future users to re-analyse the data within new contexts. In order to carry out meaningful preservation we need to ensure that future users are equipped with the necessary information to re-use the data (Conway et. al 2011)

Library profession has foreseen and responded to the emergence of new opportunities in data curation. Gold (2010) had predicted widespread academic library and librarian involvement in the development of computer-based data curate strategies.

Without a trustworthy record, people and institutions cannot make informal decisions, verify existing information, evaluate evidence, hold others accountable, construct accurate histories or develop new knowledge, An authentic record does not preserve itself, and even best intentional records' creators often lack the resources or expertise to act as permanent custodian for non-current records (Prom, 2011).

5. Librarian As Researcher To Evaluator

There is a shift from the 'librarian-as-searcher' to 'librarian- as - evaluator'. Librarians are teaching the users how to evaluate the quality and credibility of the source more than how to locate sources. More emphasis is on providing access to information rather than providing information.

Studies have shown that most of the organizations use portable media for long term storage of digital records. Though this looks like pragmatic approach, is fraught with risks like obsolescence of medium itself (e.g. punch cards, magnetic tapes, floppy disks).

6. Loss Information Is Lost For Ever

Until a decade ago it was taken for granted that digital objects would last forever. With the phenomenal advances and breakthroughs in digital technology, there are several instances of irrecoverable digital data loss due to obsolescence of hardware and software. Information, even if stored in digital format, would become inaccessible if not

kept up to date with interoperable digital technologies. Hardware and software being used are replaced with new products and procedures on a 3 to 5 years cycle.

As digital technology is ever evolving, newer issues will always be cropping up. There can never be a single strategy adaptable to all forms of digital materials forever. As such, appropriate use of technology, to facilitate retrieval of most relevant information with minimum efforts of the users is the great challenge for the present day library professionals (Stanley, 2012).

7. Digital Curation

Data curation is to maintain, preserve and add value to research data and to construct reliable digital repository for current and future use. Increasingly, data are being recognized as first class intellectual objects that can undergo quality checks, peer review, distribution and re-use. The data can be cited and contribute to the reputation of the creator of the data for good or ill (Heidorn, 2011).

Lord & McDonald (2002) and Chand & Ramesha (2014) differentiate digital curation from archiving and preservation. According to them, curation is the activity of managing and promoting the use of data from its point of creation to ensure it is fit for contemporary usages and available for discovery and re-use. This means continuous updating and keep it fit for the purpose.

Several factors impede archives from developing digital curation programmes. The complexity of electronic records, digital preservation and digital curation literature, techniques and softwares is a clear concern to many archivists. Many institutions are collecting 'born-digital' materials on an ad hoc

basis but lack systematic management plans for the materials (Prom, 2011).

8. Return on Research Investment

Researchers think of using data for personal purposes, often limited to the duration of the project on-hand. Librarians, on the other hand, think of wider access and longer time periods. In order to achieve this, the data curator (i.e. librarian) must interact with the data generator and the user community in order to establish well structured guidelines and policies.

Life cycle of a digital object is a vital concept in digital preservation and curation, and varies greatly from traditional preservation. However, the basic principles of selecting, describing and organizing content remains the same. Digital curation will become a systematic institutional function only to the extent that those interested in preserving it fully collaborate with each other (Prom, 2011). As soon as a research project begins, data collection starts. Librarian should ensure that data are in proper format and all related metadata are collected.

Gathering research data is time consuming and costlier process. However, making the data available at the right time and in right format to the research community at large not only enhances the knowledge of the subject with increased output, also enhances the rate of return on research investment.

9. Paradox of Digital Curation

On one hand we want to preserve digital information in tact in its original form, and on the other, we want to have access to this stored data using most advanced tools. Unfortunately, continued preservation and accessibility of digital information cannot be guaranteed. Long term digital

preservation is haunted by short media life, slow read times, old media and defunct web sites, rapid technological changes making existing hardware and soft ware obsolete (Chen, 2001)

10. Conclusion

Securing long term and sustainable access to research data through preservation is a huge task that many international, national and regional organizations are now addressing (Grendley, 2009) The increasing coverage of digital preservation issues both nationally and internationally demonstrates the growing awareness of the problems and risks.

In our rush to 'go digital' little thought is being given as to how do we access these records/files after a lapse of time. Even if the files survive, the hardware and software are modified, recreated or new version developed which may not be compatible with old formats of files and there is the possibility of much digital information being lost. This emphasizes the importance of digital curation.

Individuals and organizations creating digital information whether through digitizing non-digital materials or creating 'born digital' resources will need to take steps to ensure their material is safely stored and has appropriate documentation (Stanley, 2015).

Digital curation is not a one-time preservation process whereby information is kept in a byte-silo and buried to be discovered later by future civilizations. The content must be constantly nurtured, used and refreshed.

Knowledge economy constantly redefines the skills it demands from library professionals. These skills are decided by the users' information requirements

which change constantly. Therefore, instead of squabbling on the digital/analogue boundary, librarians today have to shoulder newer responsibilities.

Technology is constantly being advanced. The parameters change constantly. We need to keep upgrading to keep up with the advances. With all these complex situations we must remember that the container changed but the content itself lives on !

References

1. Berman, F and Lavoie, B. (2010). Sustainable economics for a digital planet: Ensuring long-term access to digital information. California Blue Ribbon Task Force on Sustainable Digital Preservation and Access, La Jolla.
2. Chand, R.B and Ramesha, B. (2016). Digital curation of research data in social science research library: A successful experiment at ISEC, Bangalore (India). International Conference on Library Space-Content Management for Networked Society (LC-LISCOM) (Dharmaram Vidya Kshetra, Bangalore) (Oct 18-20).
3. Chen, S. (2001) Paradox of digital preservation. *Computer* 34(3)(March) 24-28.
4. Conway, E, Graretta, S.L and Matthews, B (2011). Curating scientific research data for the long term: A preservation analysis method in context. *International Journal of Digital Curation* 6(2) 38-52.
5. Goble, C, Corco, O, Alper, P and DeRourse, D (2006). E-science and semantic web: A symbiotic relationship. *Discovery Science* 9th International Conference DS 2006 (Barcelona, Spain) (Oct 7-10).

6. Gold,A (2010). Data curation and libraries: Short term dependents and long term prospects. <http://digitalcommon.calpoly.edu/lib-dean/27>
7. Grendley,N (2009). Digital preservation: Saving for the future. *Research Information* (Feb/March 2009).
8. Heidorn, P.B (2011). The emerging role of librarians in data curation and e-science. *Journal of Library Administration* 51: 662-672.
9. Jacobs, J.A and Humphrey,C (2002). Preserving research data. *Communication of ACM* 47 (8):27-29.
10. Lord,P. and McDonald,A (2003) E-science curation report. Data curation for e-science in UK. An audit to establish requirements for future curation and presentation. <http://www.jisc.ac.uk/media/documents/programmes/preservation/e-science-report-final.pdf>
11. Oldham,L (2011). Curation: It's not just for museums any more. *Computers in New Zealand Schools* 23(1):262-273.
12. Prom,C (2011). Making digital curation a systematic institutional function. *International Journal of Digital Curation* 6(1):139-152.
13. Stanley, M.K (2012). Digital divide: A blessing in disguise. *NextGeneration Librarianship: Dr. M. Koteswara Rao Festschrift Volume* Ed by LS Ramaiah, K Veeranjanyulu and G. Sujatha, Hyderabad, BS Pubn. Pp.63-67.
14. Stanley, MK (2013). Preserving intellectual heritage in digital libraries era. National Conference on E-resources and E-learning: Challenges and Opportunities for Libraries (CHMK Library Calicut University, Kozhikode)(Feb 1-2).
15. Stanley,M.K. (2015). Digital curation of research data: Challenges and Strategies. Integrating ICT in Agricultural Libraries in India: Policies, Issues and Challenges. National Conference of Agricultural Librarians and User Community (NCALUC 2015) (IVRI, Izatnagar) (June 17-19).
16. Stanley, M.K. and Devaraj (2015). Digital curation: Carrying usable research data through time. National Seminar on Contemporary Trends in Librarianship (Sri Krishnadevaraya University, Ananthapuramu, AP) (Dec 28-29) Pp 157-163.

About Authors

Dr. Devaraj Devarasaiah, University Librarian, University of Agricultural Sciences, Bangalore.

Email: devaraj@1805@gmail.com

Dr. Stanley Madan Kumar, University Librarian (Retd), University of Agricultural Sciences, Dharwad.

Email: sunnystanmk@gmail.com