

# The International Journal of Digital Curation

## Issue 2, Volume 2 | 2007

### Editorial

Peter Buneman  
University of Edinburgh

December 2007

Apart from a few erudite sticklers for correct English, most people now use the word “data” as a mass noun. Like “water” and “oil”, it does not have a normal plural form, and it denotes something that is to be consumed, burnt, or flushed away once it has served its purpose. If language influences the way we think, this is an unfortunate transition. Even though the Internet can deliver digital data as a fire-hose delivers water or a gusher oil, digital data is not to be thought of as a consumable or renewable substance. Many digital data have intrinsic, lasting value. They should be treated with care; they should be annotated, organised and preserved so that they are available for future use; they are definitely singular.

This is the third issue of the *International Journal of Digital Curation*. It comes at a time when there is increasing awareness of the importance of the subject. The National Science Foundation has announced the Sustainable Digital Data Preservation and Access Network (DataNet) Program; and digital curation figures prominently in the current call for proposals under the European Commission's Seventh Framework Programme. In addition, the third International Digital Curation Conference, which is about to be held in Washington, DC, has attracted an unprecedented attendance.

This issue follows the usual structure of a mixture of refereed papers and short articles. There are four refereed papers: two focus on scientific data, one on textual data and one on education. Sayeed Choudhury et al. observe that while astronomy does a good job of preserving primary data, there is no systematic approach to collecting and archiving the derived data that is published in peer-reviewed journals. This is in marked contrast to molecular biology, where placing the derived data in a public database is a prerequisite to publication, while the source data is often discarded. In the sense that molecular sequence data is relatively small, the biological problem is easier than the problem in astronomy. In fact what one wants in both subjects is that all data relevant to a publication is archived for future reference and experimentation. This paper describes an architecture for such a preservation system. What is particularly interesting about the architecture is that it involves replication. The archive does not depend on a specific organisation or resource for its survival. Carole Palmer et al. report on how a curriculum on digital

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. ISSN: 1746-8256 The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre.



curation can be integrated with a curriculum in bio-informatics. They observe that there is an increasing demand for data integration in biology and that this has led to an increasing demand for people who are trained both in biology and in the technology available for data integration. This paper discusses the key components of such a training program; it also indicates how work in digital curation that has evolved from library and information science would reinforce such a program.

Tim Brody et al. revisit the way in which format profiles should be used in what they call information repositories. They argue for a more flexible approach in providing a toolkit which each information repository can use to determine its preservation policy - for example, when to migrate to new formats. Their analysis appears to be directed at conventional (text) documents. It would be interesting to know whether it could be extended to the broader class of scientific data formats. Paul Watry, in elucidating a theory of preservation, describes a variety of current activities associated with the preservation of scientific data. This is a challenging area: even recording the enactment of a workflow used in the derivation of a scientific data set is demanding; migrating that workflow to a new software environment while guaranteeing reproducibility is a huge challenge.

The general (unrefereed) articles report on a number of activities such as conference and project reports associated with digital curation. These are extremely valuable, but unlike the refereed papers, most of the authors of these papers have strong links with the Digital Curation Centre (DCC). Although the IJDC is a publication of the DCC it is intended to be an international journal devoted to all digital curation activities. Everyone is encouraged to submit to this section.