The International Journal of Digital Curation
Issue 2, Volume 4 | 2009

Editorial

Chris Rusbridge,

Digital Curation Centre

October 2009

I am very pleased to present the second issue of Volume 4 of the *International Journal of Digital Curation*, packed with 8 research papers and 6 general articles, and this time seasoned with some of the spice of controversy. I always enjoy the refreshingly different viewpoints of the material published in IJDC, and can claim to learn a great deal from each issue.

Controversy first: since digital curation and preservation are extremely inter-disciplinary, bringing together people from widely different backgrounds, it is not surprising that terminology clashes and contested assumptions occur. In this issue, we have a letter to the editor from Andrew Wilson, taking Ronald Jantz to task for part of his paper "An Institutional Framework for Creating Authentic Digital Objects", published in IJDC 4(1), in which Jantz suggested that a digital certificate might be sufficient to assure authenticity. Wilson urges a more nuanced analysis, suggesting that Jantz ignored long-established archival work, particularly of the continuum view. We also have Jantz's response, acknowledging the different perspectives, but suggesting that there is much further work to do before we can claim a "clear, unambiguous definition of authenticity" in this digital world.

As usual, the papers and articles appear separately, in alphabetical order of first author family name. Also as usual, I would like to present them to you in a slightly different order, not from any view of priority or merit, but to make a more coherent story. All the peer-reviewed papers in this issue derive from papers presented at the 5th International Digital Curation Conference, held in December 2008. I might add that we have plans to broaden our catchment area by publishing papers developed from other relevant conferences, including some from the iPres series of conferences; papers selected from iPres 2008 however are still in the reviewing cycle (although as you will see, we have two general articles in this issue from that source).

Two papers are linked by their association with data on the environment. Baker and Yarmey develop their viewpoint with environmental data as background, but their emphasis is more on arrangements for data stewardship. In particularly they discuss the interesting notion of "distance from origin" in data repositories, from the local to the remote archive. They envisage data "moving over

time through many repositories, with a cumulative value as defined by all participating repositories and *spheres-of-context"*. Jacobs and Worley, on the other hand, write from the viewpoint of a distant archive (from the data origin), reporting on experiences in NCAR in managing its "small" Research Data Archive (only around 250 TB). These data are particularly important for re-analyses, and are heavily in demand. They remark that the "notion of data services and the feedback from the user community [does] not appear to be addressed directly in the Digital Curation Centre Curation Lifecycle Model."[1] They conclude with some interesting remarks on sustainability, and point out that "successful curation requires a stable commitment from people, facilities, and an institutional organization".

Halbert also looks at elements of sustainability, in distributed approaches that are cooperatively maintained by small cultural memory organizations. Although his paper refers particular to the Private LOCKSS Networks adopted by the MetaArchive Cooperative, Halbert stresses that PLNs are not the only option for those interested in such an approach. His paper includes a useful survey of other work in this area.

Naumann, Keitel and Lang, on the other hand, report on work developing and establishing a well-thought out preservation repository dedicated to a state archive. They look particularly at the requirements (particularly strong for such archives) for authenticity, integrity and trustworthiness, and their implications for metadata. Among many fascinating remarks, I particularly enjoyed the idea of "Sacred Content, Free Metadata". Metadata must be allowed to change (although such change must be recorded), but the content should never change.

Sefton, Barnes, Ward and Downing also address metadata, plus embedded semantics; their viewpoint is that of document author, particularly of those whose documents may be moved between the two most widely-used word processing systems, Microsoft Office and OpenOffice. Gerber and Hunter similarly address metadata and semantics, this time from the viewpoint of compound document objects, or perhaps structured assemblies of related material. These approaches however are quite different, yet both aim to make digital resources more useful for further analysis.

Finally, we have two papers loosely linked through standards, though from different points on the spectrum of the general to the particular, as it were. At the particular end, Todd describes XAM, a standard API for storing fixed content; while from the more general end, Higgins provides an overview of continuing efforts to develop standards frameworks that might be useful for those thinking of preservation and curation, with community participation.

Moving on to general articles, in this case I would like to mention first my colleagues Pryor and Donnelly, who present a white (or possibly green?) paper on developing curation skills in the community. Resulting from a workshop of the Research Data Management Forum (a joint effort between the DCC and the UK's Research Information Network), they base their work on a proposition to the workshop from Prof Sheila Corrall that "data skills should be made a core academic competency". How could this be achieved?

---

[1] Data curation lifecycle http://www.dcc.ac.uk/docs/publications/DCCLifecycle.pdf

Next, I would highlight two very interesting articles that originated from iPres 2008. These are Dappert and Farquahar who look at how explicitly modelling organisational goals can held define the preservation agenda. Woods and Brown describe how they have created a prototype virtual collection of 100 or so of the thousands of CD-ROMs published from many sources, including the US Government Printing Office.

Shah presents the second part of his interesting independently-submitted work on preserving ephemeral digital videos, and particularly the ContextMiner tool.

Finally, Knight reports from a Planets workshop on its preservation approach, while Guy, Ball and Day report from a UK web archiving workshop.

A varied and fascinating collection, indeed!