The International Journal of Digital Curation
Volume 7, Issue 1 | 2012

Editorial

Kevin Ashley,
Director,
Digital Curation Centre

March 2012

There's a great deal to read in this issue of IJDC, the first of 2012. I don't intend to add to that by describing every paper in detail in this editorial, as I have done in the past. I will let the authors speak for themselves and pick out some themes and highlights here. Following that, there is news of some changes to IJDC itself, which I know many (including the previous editor) will be pleased to hear about.

We have a good mix of papers on research and practice from this year's International Digital Curation Conference, which took place at Bristol in early December. Many are complemented by the multimedia coverage available on the event pages on the DCC website, with accompanying slides linked from the programme page[1] and videos of the presentations and question sessions available at the event video gallery.[2] The video gallery also lets you see the keynote presentations that aren't represented here in IJDC. They are well worth devoting some time to; I've got more out of re-watching them even though I was there at the time.

Bill Underwood's paper on the use of grammars to parse binary file formats won the best paper award at this year's conference and struck me as a particularly satisfying way to apply well-honed concepts in computer science to the relatively new task of validating file formats for digital preservation. I realise that the power of the techniques he describes isn't necessarily apparent to all of our readership, so I'll make an attempt to describe why this paper spoke so strongly to me. I didn't study computer science formally, but a number of its classic texts were essential to me in my early career, among them Aho & Ullmann's Principles of Compiler Design (1977). The ability to use a small number of formal statements to describe the entire structure of a computer language, and then use automated systems to understand those statements and construct a program (a compiler) which would automatically recognise valid and invalid programs in that language, seemed almost magical to me. Techniques like these allowed new compilers – and hence new programming languages – to be constructed very rapidly. Prior to this compilers were constructed by hand, an

[1] IDCC 11 Programme Page: http://www.dcc.ac.uk/events/IDCC11/

[2] IDCC 11 Video Gallery: http://www.dcc.ac.uk/events/IDCC11/video-gallery

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. The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

 doi:10.2218/ijdc.v7i1.209

expensive and error-prone process. To use the same techniques to describe digital file formats seems obvious once it is stated, and it should allow validators to be constructed far more rapidly than has been the case so far. This has the power to greatly improve the efficiency of a basic digital preservation task.

Automated tools still require some validation to ensure that they are behaving as designed. Fetherston and Gollins describe the nature of a corpus that would be necessary to carry out this validation. The case they consider is one closely related to format validation, that of format identification. It is surprising in many ways that so much has been achieved without such test corpora available to validate and compare the tools we might choose to work with. The corpus does not yet exist, but the authors have done an excellent job in describing the characteristics it should have and in drawing comparisons to similar corpora in related fields of study.

We also have a selection of papers not related to last year's conference. Groth, Gil, Cheney and Miles look at the perennial issue of provenance as it relates to the web. You may take issue with some aspects of their perspective; if so, draw this to the attention of the relevant W3C groups, whose work this paper draws upon. Standards will result from this work and we're glad to be able to bring it to the attention of the wide audience that will be necessary to ensure those standards are fit for purpose.

Knight describes a problem that can all too easily arise when we attempt to scale preservation tasks to industrial scale and in the process place undue loads on the systems of people who may hardly be aware of our existence.

Finally, to the news about IJDC itself. Although the exact details are still being finalised as the editorial is being written, I'm glad to be able to say that either at the time of publication, or shortly thereafter, IJDC articles will be citable using DOIs. Actually, I can say a little more – they are already citable using DOIs, with the minor inconvenience that those DOIs aren't visible to anyone at present. The team which supports our OJS journal platform at UKOLN have gone over back issues and assigned DOIs to all existing articles. The template for new and past articles will make those DOIs visible to readers. At the same time some formatting changes will also take place which will ensure that each article is more self-descriptive when viewed as a standalone PDF, something which we know a number of authors and readers have asked for. It's taken longer than we planned but I'm very glad that this development is finally with us and grateful to those who made it happen.

References

Aho, A., Ullman, J.D. (1977). *Principles of Compiler Design.* Reading, Mass: Addison Wesley.