

The International Journal of Digital Curation

Volume 8, Issue 1 | 2013

Editorial


Kevin Ashley,
Digital Curation Centre

Alex Ball ended his editorial in the last edition of the International Journal of Digital Curation by noting changes that we had made to the workflow and production practices for the journal. Although these are not all visible to authors or readers, they do make the journal more straightforward to produce and will help us to reduce article turnaround times. I'll now begin by describing forthcoming changes that will be visible to authors and readers of IJDC and will involve us putting into practice changes we advocate at the Digital Curation Centre that are frequently described in the pages of this journal.

One small but significant change involves us exposing information about an article's creation and history. We will make clear the date when the article was first submitted and when the final version was accepted for publication. This should help authors establish primacy where necessary and will help readers to understand the contemporariness of the research or practice described. It will also encourage us to maintain improvements in article processing time. We may also expose information about the dates of intermediate revisions. A final decision has yet to be made and we will review practice in existing journals, as well as listening to our authors and readers before deciding on the details.

The change which we hope will have the widest impact relates to data – specifically data collected by the author which underpins their paper. For future publication in IJDC we will require that the data behind articles is assigned a permanent and resolvable identifier, and that it is placed in a custodial environment that gives an appropriate degree of assurance about its longevity. The identifier must appear in the article and we will provide clear guidelines on how this is to be done. Furthermore, we will require that data is made available for open access except when there are clear reasons why this is not possible, such as when identifiable human subjects are described in the data.

On a related note, we also make clear that we support the concept of data as a primary scholarly output and therefore permit, even encourage, its citation in references on an equal footing with other resources.



Work published in this and earlier editions of IJDC and elsewhere makes clear why these changes are good for scholarship and ultimately for society. The paper in the journal is often an advert for the research but is not the whole story. To understand, to replicate and to further someone else's findings we often need to see their data, or at least its structure. In other cases we may find entirely new things to do with data that were unforeseen by its creators.

But what of the present? This first issue of IJDC Volume 8 brings us 17 papers from IDCC13, held in Amsterdam in January, as well as four independent peer-reviewed papers. Of the IDCC13 submissions, nine report on practice and eight were research papers, which were also subject to full peer-review. The practice papers may deserve particular attention, since the 15 minutes speaking time allocated to their authors at IDCC rarely allowed them to describe their work in depth.

[Lecarpentier et al.](#) describe EUDAT, one of a number of projects in the European Commission's FP7 programme creating cross-disciplinary collaborative data infrastructure. Judge for yourself whether they have successfully identified common requirements in different research domains which make the provision of common data services for collaboration realistic and useful.

Two contributors report on different aspects of data publication. [Callaghan et al.](#) report on the JISC-funded PREPARDE project (in which the DCC is a partner) describing the workflows for an earth sciences data journal. Description of practice in an emerging area, such as data journals, is invaluable to help others who are a few steps behind and facing similar concerns. [Hoogerwerf et al.](#) consider the related but more general question of the linking of traditional scholarly publications with data, citing a number of examples explored in the OpenAIRE FP7 project. It is interesting to contrast those approaches that appear to favour the paper as the primary scholarly output and those which treat all the entities on a more equal basis.

[Lagoze et al.](#) describe an approach to a common problem in many research domains: that of dealing with confidential data and moving data between the protected and unprotected domains. They describe a prototype solution (CED²AR) which manages this process and, unusually for such solutions, recognises that some metadata may also require protection. Perhaps they and EUDAT need to start a conversation.

Four articles deal with different aspects of education and training. [De Smaele et al.](#) describe their work on re-skilling university library staff to deal with research data issues in a collaborative service setting – work that has already received recognition and reuse in a number of other European countries. Three other papers deal with education and skills development at an earlier stage – the point where people are still categorised as students of one sort or another. [Kelly et al.](#) describe work to educate LIS students through placements in working scientific data centres. The other two articles both consider the need of graduate students in all disciplines learning about data as it applies to their own research. [Carlson et al.](#) describe their findings on education needs derived through interviews with students and preview the training that was developed as a result; [Scott et al.](#) describe case studies and a guide that has been used to illustrate the differences in data handling between disciplines and helps to ensure students have a general awareness of what constitutes good data handling practice.

[Doorn et al.](#) were responsible for one of the most popular break-out sessions in the conference. Read their paper on the relationship between data sharing and research integrity to understand why. Although set against the background of recent exposures of research fraud in the Netherlands and the reports which resulted, the experience will be depressingly familiar to many readers and the conclusions applicable worldwide.


Two of the research papers from IDCC13 deal with different aspects of standards in digital curation. [Yarmey and Baker](#) take a thorough look at the process by which scientific metadata standards are developed, noting that the process is rarely one-off and that the economic drivers and mechanisms are different from those that characterise the development of standards applicable in the commercial world. They advocate a process which is collaborative rather than hierarchical. [Austin et al.](#) describe work that is closer to the commercial world, relating as it does to engineering and more specifically to the preservation of information regarding materials testing. They describe innovative work which uses one set of standards for a testing process to derive another set of standards for storing the information that results from these tests. This has the potential for wider applicability and more efficient standards generation in some fields.

[Spencer](#)'s paper follows on directly from the work of [Featherstone and Gollins](#) in IJDC 7.1. They described the need for a test corpus for format identification tools and described the characteristics that such a corpus should have. Spencer describes a technique to generate what he describes as a 'skeleton' corpus using simple, semi-automated techniques and outlines further work to be done in this area.

[Bicarregui et al.](#) discuss best practice in the management and preservation of research data in what many call 'big science.' Many believe that this is a solved problem and that such large-scale collaborations can be trusted to do the right thing. As the paper shows, the truth is somewhat more complex. Even when people know what to do, there is always the possibility that they don't know an efficient way to do it. The paper describes guidance produced by the MaRDI-Gross project aimed at the project engineering staff in future big science initiatives.

[Yakel et al.](#) look at an issue facing many working on a slightly less grand scale, who often face a choice as to where to place digital materials for long-term safekeeping. Trust is fundamental and we have some existing and developing mechanisms to determine who deserves that trust, such as ISO 16363. Yakel and her colleagues examine the problem of whether the things that these standards test relate to the characteristics that a repository's designated community associate with trustworthiness.

Two further papers examine the utilisation of cloud services for preservation from two perspectives. [Von Suchodoletz, Rechert and Valizada](#) describe the potential for emulation-as-a-service using cloud instantiations of a variety of software environments. This potentially makes it much easier for a wider variety of repositories to consider emulation as a viable means of access for some digital content. Their paper is theoretical but goes into considerable detail about the architectural requirements and a business model for a service. [Rosenthal and Vargas](#) describe a practical experiment using cloud storage as an alternative to local instances for an



existing preservation system: the LOCKSS network. On present evidence, they conclude that this is not a cost-effective proposition.

[Bazzanella et al.](#) discuss a solution to what they characterise as a divergence between the notion of the ‘Cool URI’ and a persistent identifier. Whether their Entity Name System fixes the problems described and whether those problems are clear-cut divided many at IDCC13. However, all agreed that there are issues of concern here somewhere; persistent identification of some sort is critical for preservation.

Four further research papers in this issue are not from the conference. [Boutard, Guastavino and Turner](#) articulate what they see as particular problems faced by those preserving artistic works with specific technological components. One example of such work would be a musical piece that requires a particular software suite (and possibly specific hardware) to be realised in live performance. Such problems are not specific to the digital arena – artists of the 1960s onwards have produced video works which are dependent on properties of analogue television systems and cathode ray tubes, technologies which will prove increasingly difficult to preserve. The authors set out a framework which they assert goes some way to improving the situation for digital archives.

[Neugebauer and Murray](#) tackle a concern which we’ve rarely dealt with in IJDC – the nature of advocacy for open access and the extent to which institutional services, such as repositories, have a role in such advocacy.

[Yoon](#) addresses an area that is attracting increasing attention: the preservation of personal digital collections or user-generated content. Her study looks at one particular manifestation, the blog, and examines bloggers’ own attitudes to what requires preservation. Their belief that content and context are more important than behaviour or appearance is reassuring and confirms the value of approaches taken by a number of blog preservation advocates.

Finally, [Kim, Warga and Moen](#) describe an attempt to produce generalised competencies for digital curation professionals by examining the content of job advertisements in the USA. It may not surprise some that their findings show that the situation is ‘complex.’