The International Journal of Digital Curation
Issue 1, Volume 4 | 2009

Report from the Digital Curation Curriculum Symposium (DigCCurr) 2009

Alexander Ball and Michael Day

UKOLN,

University of Bath

May 2009

Summary

The second Digital Curation Curriculum Symposium was held on April 1-3, 2009, in Chapel Hill, North Carolina, with the theme "Digital Curation Practice, Promise and Prospects". The Symposium featured sessions dealing with issues from the cutting edge of digital curation research, while others showcased recent developments in digital curation tools. At the same time, the Symposium also considered how to equip the new generation of information professionals with the necessary skills to put this research and development into practice.

Introduction

The Digital Curation Curriculum Symposium, (DigCCurr, pronounced 'Didge Seeker') is an event series organised by the University of North Carolina at Chapel Hill (UNC), with the aim of providing a forum for discussing both the latest research in digital curation, and how to teach this to library and information students. It is linked to the DigCCurr Project, a collaboration between the School of Information and Library Science at UNC and the US National Archives and Records Administration (NARA) to produce a graduate-level programme in digital curation. The first in the series was held in 2007, and due to its success a second one was held on April 1-3, 2009. The theme for this second symposium was "Digital Curation Practice, Promise and Prospects".

The event was held at the William and Ida Friday Center in Chapel Hill, and attracted around 260 delegates from 11 countries. Apart from the opening keynote talk and the closing plenary, the Symposium consisted of 28 sessions, with four running concurrently in each of the seven time slots; consequently, this report only covers around half of the total programme. There was a healthy mix of paper, presentation and panel sessions on a wide variety of topics.

April 2, 2009

Opening Session

Following the welcome and introductions, the keynote talk was given by **John Wilkin** of the University of Michigan. He introduced HathiTrust, a multi-institutional repository for preserving the outputs of digitization programmes. The repository is centralized but mirrored between two sites, with backups held at a third location. In terms of scale, HathiTrust currently has 25 partners and 3 million titles online, with titles being added at a rate of up to 700,000 per month. Wilkin predicted HathiTrust will hold around 18.6 million titles by 2012, of which about 5 million will be in the public domain.

The repository has both public and private benefits. For partner institutions, it provides accessible electronic versions of print titles held by those institutions, with the full support of the publishers. The electronic access also allows partners more options for managing their print collections: for instance, keeping fewer duplicate copies on the shelves, or collaborating on print archiving. On the public side, it provides open access to public domain and permitted works, and across all the texts allows non-consumptive research, where the repository itself performs text processing on demand and returns the resulting statistics instead of text passages. At the moment, this is limited to returning the number of times a search term occurs on each page of a chosen work, but in future it is hoped to provide natural language processing compatible with SEASR data analysis. APIs are also being developed so that institutions can write their own front-end interfaces.

Health and Social Science Data Curation Collaborations

The first set of concurrent sessions included a panel presenting two different case studies of research teams curating their data in collaboration with the social science

digital data archive ICPSR.[1] **David Thomas** explained that ICPSR had set up a new, dedicated server for outside organizations to use to archive their data themselves, but using ICPSR tools. The new server had all the tools available, but without certain permissions that may have constituted a security risk. This was not as simple as it sounded, as some of the tools relied on those permissions, or on being on the main server; thus adaptations had to be made to get them running properly. These adaptations were not just behind the scenes, though, so they introduced difficulties for ICPSR staff when they came to support external users.

One of the first users of the system was the Fenway Institute, a research team based in a health centre and specializing in HIV/AIDS and LGBT (lesbian, gay, bisexual, transgender) health care issues. **Chris Grasso,** one of the team at the Institute, explained some of preparation that had to be done, including the many planning meetings, three-and-a-half days of training with ICPSR and a visit by ICPSR staff to the Institute to check the latter's IT infrastructure. When it came to using the system for real, there were difficulties with diverse file formats and missing codebooks to be overcome, while LGBT terms had to be added to the ICPSR metadata schemata. Furthermore, when making the data available to ICPSR members, extra care had to be taken to anonymize the data, for example, by re-encoding postal codes and aggregating data to remove small sample sizes. Some data were too sensitive and were restricted to on-site access only.

The Roper Center is a university-based research team specializing in opinion polls; **Cindy Teixeira** reported on the Center's experience in using the ICPSR system as part of the Data-PASS Project. In contrast to the Fenway Institute example, the aim here was for a shared archiving process, producing identical data holdings at both the Roper Center and ICPSR, with standardized metadata and a joint citation. A workflow was devised where Roper staff were responsible for acquiring and processing data, while ICPSR staff performed quality assurance, and archived and released the data. There were some issues to resolve, such as reconciling naming conventions and DDI implementations. The latter issue was overcome using XML and XSLT to create two versions from the same set of metadata.

Funders' Perspectives

Also in the first set was a panel presenting funders' perspectives, moderated by Joel Wurl of the National Endowment for the Humanities (NEH). First, **Joyce Ray** gave an overview of the recent funding activities of the Institute of Museum and Library Services (IMLS), starting by noting the importance of data lifecycle models and long-term data management in a recent report on *Harnessing the Power of Digital Data for Science and Society*, published by the Interagency Working Group on Digital Data to the National Science and Technology Council (2009). The presentation mainly focused on the activities of the IMLS in supporting education and professional development in the digital curation domain. Ray mentioned some interesting projects funded as part of the National Leadership Grants for Libraries Program, including an early grant made to Johns Hopkins University to support collaboration with astronomers and the National Virtual Observatory on data curation. A more recent

---

[1] Interuniversity Consortium for Political and Social Research. URL: http://www.icpsr.umich.edu/

grant had been awarded to Purdue University and the University of Illinois, Urbana-Champaign to explore the different disciplinary data sharing practices of researchers (Brandt et al, 2008).

In the next presentation, **Neil Grindley** gave a short overview of digital curation activities funded by the Joint Information Systems Committee (JISC) in the UK. As part of this, he commented that JISC was in the happy position of having the Digital Curation Centre, soon to move into its third phase. He noted that there was always a need in the digital curation domain to prove that what we are doing continues to be effective and valid. Finally, Grindley was pleased to announce that JISC would be funding digital curation activities through a forthcoming JISC Research Data Programme.

Kathleen Willis then gave a brief introduction to programmes funded by the National Historical Publications and Records Commission (NHPRC). Their current grant programmes were largely focused on electronic records, the digitisation of historical records, professional development, and the development of strategies and tools. As the NHPRC was a relatively small funding agency, Willis noted that the commission needed to be strategic in making grants. She concluded that ideal projects should be both collaborative and cohesive, and needed to consider both immediate gains and longer-term outcomes.

Joel Wurl gave the final presentation in the session on the funding activities of the National Endowment for the Humanities (NEH). The presentation gave information on some of NEH's current funding opportunities, including international collaboration on the *Digging into Data* challenge and the JISC/NEH Transatlantic Digitization Collaboration Grants. Grant opportunities at the NEH's Division of Preservation and Access were mainly focused on digital preservation, sound recordings or moving images, and preventative conservation.

Education for Digital Curation

Following lunch, there was a half-hour poster session, followed by the second set of four sessions. One was a paper session chaired by Maria Guercio of the University of Urbino. **Magia Krause** gave the opening presentation, discussing how the cognitive apprenticeship model, a concept developed in educational psychology, had been applied to digital curation education at the University of Michigan. This model provides a structured means of gaining practical knowledge and skills by being "encultured" in particular domains or practice-based communities, and was linked in Michigan with internships and peer mentoring. It was noted that the model required active support from course mentors.

Next, **Stefan Strathmann** of the Göttingen State and University Library provided an introduction to nestor's approach to digital curation education in Germany. The nestor initiative had been funded by the German Ministry of Education and Research (BMBF) as a national network of expertise in digital long-term preservation. Strathmann noted that academic institutions from Germany, Austria and Switzerland had been collaborating as part of nestor on developing a shared approach to education and training in long-term preservation. This included occasional seminars, more comprehensive training events (schools), the publication of a *nestor Handbook* (Neuroth et al, 2008) and the development of a cooperative curriculum. Strathmann

mentioned that the nestor project formally ends in June 2009, but that partners will then combine to form a sustainable membership-based organisation.

The next presentation, by **Jean Dryden** of the University of Maryland, was a report from a survey of archivists' attitudes towards copyright and the reuse of digitised archives. The survey focused on archives that had made content freely available on the Canadian Archives portal, and explored how some archives sought to maintain control over the use of their archival collections – typically through using technology or terms and conditions of use agreements. The survey also tried to understand how archivists utilised the various sources of copyright information available. In conclusion, Dryden suggested that there was a need for structured professional education programmes on copyright so that archivists could take a better informed view of archive use.

In the final presentation, **Margaret Henty** of the Australian National Data Service (ANDS) provided an overview of attempts to develop a national approach to research data management in Australia. The aim of ANDS was to "transform disparate collections of research data" into a "cohesive corpus of research resources." The service was focused both on developing frameworks and some tools, as well as on building capabilities by direct engagement with the community.

Digital Curation Tools and Demonstrations

The Symposium played host to a number of demonstrations of digital curation technology; these were divided among sessions in each of the two afternoon sets. In the first session, **Andreas Rauber** (Vienna University of Technology) presented two tools from the Planets Project: PLATO, a preservation planning tool for repositories and archives, and HOPPLA, a simplified preservation environment for the home or small office. **Heather Bowden** (UNC Chapel Hill) promoted the Digital Curation Exchange website, a nascent social networking site and resource hub for digital curation researchers and practitioners.[2] **Jeffrey van der Hoeven** (Koninklijke Bibliotheek) showed the Dioscuri modular emulator in action, while **Raymond J. van Diessen** (IBM) exhibited Preservation Manager, a tool for keeping track of available "pathways" (hardware and software environments) for rendering files, and the Universal Virtual Computer emulation system. Lastly, **Richard Pearce-Moses** (Arizona State Library, Archives and Public Records) demonstrated PeDALS, a Persistent Digital Archives and Library System seeking to make digital archiving more automated and scalable.

The second session saw demonstrations from: **David Pearson** (National Library of Australia) with Prometheus, a digital preservation workflow system, and Mediapedia, an online encyclopaedia of digital storage media;[3] **Chirag Shah** (UNC Chapel Hill) with ContextMiner, a web-based application for retrieving online contextual information (videos, blog posts, etc.) about a subject;[4] **Reagan Moore** (UNC Chapel Hill) with rule-based software system iRODS; **Geri Bunker Ingram** (OCLC) with digital collection management software CONTENTdm; and **Andrew McHugh**

---

[2] Digital Curation Exchange http://digitalcurationexchange.org/
[3] Mediapedia http://www.nla.gov.au/mediapedia/
[4] ContextMiner http://www.contextminer.org/

(HATII, University of Glasgow) with DRAMBORA, the Digital Repository Audit Method Based on Risk Assessment.[5]

Both sessions began with a series of 'taster' talks to introduce the tools, after which delegates were free to move around the different demonstrations and see the tools in action.

Curation of Scientific Datasets: Trends, Current Initiatives, and Solutions

Also in the last set of the day was a session featuring presentations on four different aspects of curating scientific datasets. The first was from **Michael Day** and **Alexander Ball** (UKOLN, University of Bath) who looked at a range of disciplinary and institutional issues concerning digital curation, not least the thorny issue of whether datasets should be dealt with by discipline-based data centres or institutional repositories. They also gave some initial findings from two near-completed projects. One was the DCC SCARP (Sharing, Curation And Re-use, Preservation) Project, a set of ten immersive, discipline-based case studies seeking to find both good practice and gaps in the digital curation performed within these disciplines. The other was a scoping study for a Dublin Core Application Profile (DCAP) for scientific data, along the lines of the Scholarly Works Application Profile (Allinson, Johnston & Powell, 2007).

Jane Greenberg (UNC Chapel Hill) neatly followed this with a presentation about the DCAP developed for use in the Dryad Repository. Dryad specializes in datasets that underlie papers in the field of evolutionary biology and related disciplines. Development of the DCAP followed the guidelines of the Singapore Framework, starting with functional requirements and a domain model. To enhance interoperability, the description set was drawn from existing vocabularies: Dublin Core, the Data Documentation Initiative (DDI), PREMIS, Ecological Markup Language (EML), and Darwin Core. A vocabulary server called HIVE is being developed to integrate different discipline-specific vocabularies using the Simple Knowledge Organization System (SKOS).

Heike Neuroth (Max Planck Digital Library) reported that the Alliance of German Science Organizations has identified six priority areas for improving the research infrastructure in Germany, of which one is the preservation and reuse of primary research data. A national working group is therefore writing a position paper on research data, dealing with issues of open access, disciplinary and international differences, scientific recognition for data (along the lines of impact factors for journals) and sustainable infrastructure. In connection with this, disciplinary round tables and data management workshops will be held, alongside studies of sustainable business models and other organizational and legal issues. Towards the end of 2009, the Alliance will issue a call for pathfinder projects implementing innovative discipline-based digital curation.

John Kunze (California Digital Library) closed the session with an explanation of how the California Digital Library (CDL) had been rebuilt on the principle of using micro-services. CDL uses: NOID for generating identifiers; a PairTree convention for locating an object in a directory structure based on its identifier; Reverse Directory Deltas for versioning; BagIt for packaging files; GrabIt for moving large batches of

---

[5] DRAMBORA http://www.repositoryaudit.eu/

packages; wget for moving files across the Internet; JHOVE2 for file format identification; XTF for indexing; and so on. The particular challenges for dealing with data are that there is a constant tension between standards and innovation, and between its usefulness as evidence and its instability whilst still active.

April 3, 2009

Digital Curation Vignettes: Personal, Academic, and Organizational Digital Information

The first set of sessions on the second day included one featuring presentations on two digital curation surveys and a practical curation project. **Andrea Japzon** of Drexel University presented the results of a survey of the attitudes of 26 individuals to their personal collections of physical and digital items. She found that those surveyed tended to value their digital assets at the collection level, but their physical assets at the individual item level; furthermore, they associated their digital assets more with personal achievement, and their physical assets with family. Information horizon maps drawn by the respondents showed that they valued digital items more for their everyday usefulness, while physical items were valued more in the long term, independent of their utility.

Joan E. Beaudoin, also of Drexel University, explained that academics are not depositing their work-related digital images in the institution's repository in the way they used to deposit their slides and photographs in the departmental libraries. She has surveyed four archaeologists and four art historians, and even in such a small sample found diversity of practice with regard to saving and backing up digital images. There was a general attitude that only other people's images could be of wider interest. None of those surveyed knew what an institutional repository was, and while they all knew of library or support staff who could help them create images, half thought they would get in the way building a collection of images, and only one respondent thought they could help with archiving. In short, there was an appetite for advice on good practice, and a need for repositories to be better publicized.

The last vignette came from **Maria Esteva** of the Texas Advanced Computing Center, who had rescued the digital records of a private Argentinian company that had closed after 20 years of scientific and cultural heritage work. She had approached the task like an archaeological dig. The files were transferred onto a brand new server that would act as an offline dark archive, and checked for integrity. Formats were identified using DROID and FileMerlin, and from this and interviews with staff, Esteva had constructed a timeline of hardware and software changes within the company. She had then installed appropriate viewers for the files, alongside audit control software to monitor for data access and corruption.

Esteva finished her presentation by demonstrating how text mining techniques could be used to trace changes in how closely each staff member worked with his or her colleagues.

Cooperative Approaches to Digital Preservation

Meanwhile, a panel in the same set provided an opportunity for various LOCKSS-based initiatives[6] to present the advantages of using cooperative, distributed networks for the long-term preservation of digital content. In his introductory remarks, **Tyler Walters** of the Georgia Institute of Technology repeated a comment made by Reagan Moore (UNC Chapel Hill) on the first day of the conference, to the effect that there was a need to throw away systems while being able to maintain content and metadata. The core of the session was made up of presentations of four Private LOCKSS Network (PLN) projects, providing overviews of: the MetaArchive Cooperative by **Martin Halbert** (Emory University); the Alabama Digital Preservation Network by **Aaron Trehub** (Auburn University), the Persistent Digital Archives and Library System (PeDALS) by **Richard Pearce-Moses** (Arizona State Library), and the Data Preservation Alliance for the Social Sciences (Data-PASS) by **Jonathan Crabtree** of UNC Chapel Hill. The presentations, collectively, demonstrated how all of these LOCKSS-based initiatives exist within rather complex technical and organisational contexts. For example, Pearce-Moses commented that curators were now largely working with business processes rather than 'records.'

Archives in the Wild

The next set included a session consisting of three papers dealing with novel digital curation approaches. **Frank McCown** of Harding University argued the case for using what he called the Web Infrastructure – caches of web content such as Google's cache and the Wayback Machine – as a distributed web archive, and using OAI-ORE to re-aggregate the resources. He then introduced the ReMember Framework, a prototype web service that will make it easy to repair OAI-ORE resource maps (ReMs) when resources go missing, and back up resources to WebCite if they are not already cached somewhere. The key to its success will be making repairing ReMs sufficiently fun and accessible that people will want to do it as a hobby.

Richard J. Cox of the University of Pittsburgh argued that with information being more freely available than ever before, people are becoming increasingly do-it-yourself in their attitudes. There are numerous self-help books now available on personal organization and personal information management, but there is precious little information provided on personal archiving. Cox argued this represented a missed opportunity, especially given the success of digital scrapbooking sites like Flickr, Facebook and so on. He concluded that to get digital curation in the mainstream, we need to make it more specific, lay out evidence that it works, and package it in an easily understandable way.

Finally, **W. Brad Glisson** of HATII, University of Glasgow, looked at the place of digital forensics in curation. The most crucial part of his talk centred upon the curation of deleted files. As files are not removed from a drive when deleted, merely rendered invisible to the operating system, it is technically possible to recover a large proportion of them using specialist tools. The question is whether it is ethical or desirable to do so, and following Glisson's presentation there was a lively debate on the matter, with the archiving of the US Presidential records clearly at the forefront of people's minds.

---

[6] LOCKSS (Lots of Copies Keep Stuff Safe) http://www.lockss.org/

Gaps and Persistent Challenges

A parallel panel gave an opportunity for mature reflection on long-standing digital curation issues. **Clifford Lynch** of the Coalition for Networked Information (CNI) began by making four challenging points, by way of stimulating discussion. He first commented that that we had not been very good at making the case for digital preservation, evidence being the imbalance in intellectual property law, whereby exceptions for preservation were still severely constrained. Secondly, Lynch noted that we still did not have good answers to the sustainability question, and that we needed to be more realistic in our assumptions about economic models. A third point was that we did not do 'triage' very well, the potential deluge of content meaning that best practice would probably be less important than survival practices. He asked what would happen to the historical digital assets of a corporation or institution that suddenly died. Lynch's fourth point was that we needed to give more attention to personal documentation.

Next, **Donald Sawyer** of VIE, Inc. (formerly of the National Aeronautics and Space Administration) provided examples from the 45-year history of the National Space Science Data Center of dramatic change in hardware, systems and processes. Even within a well-managed operation, Sawyer commented that we should not be surprised if some digital information gets lost or corrupted through data entry errors or problems with hardware or software, amplified by ongoing resource constraints that mean that organisations are unable to maintain adequate attention to detail over the long term.

Kevin Ashley gave the third presentation, a review of challenges in digital preservation research. He started by introducing a series of reports that were mostly published around six or seven years ago, and reviewed the current state of the art with their recommendations. Work still outstanding included: dealing with new generations of content types (e.g., virtual worlds or musical scores), the accelerated ageing of systems and software, the development of software registries, automated provenance generation, and the definition of designated communities.

The question and answer session that followed covered a lot of ground, including: libraries' poor record of engaging with new content types (Lynch); the misguided focus on authenticity from a purely computing science point of view (Howard Besser); the unmanaged distribution of "institutional" content around the web, for example on cloud storage or GMail (Lynch).

Personal Digital Archiving

Following lunch, one of the sessions featured four presentations on personal digital archiving. **Jeremy Leighton John** of the British Library gave an overview of some BL projects concerned with managing personal digital collections. The Digital Manuscripts Project was focused on analysing the personal collections of four recent scientists, with content stored on a variety of obsolete media types. The Digital Lives project was a collaboration with University College London and the University of Bristol and was looking at the lifecycle of personal information. Deposit seemed to be based on traditional end-of-life deposit mode, perhaps in this particular case, literally!

Cathy Marshall of Microsoft Research then gave an entertaining talk on how people manage their own personal (digital) collections, largely based on researcher

interviews. They revealed that there was a tendency to believe that storage was cheap and that it might be possible to keep "everything." At the same time deleting content is hard work, and is mostly done unsystematically or by accident. In practice, attitudes to content are based on their perceived value. This forms a kind of hierarchy. High-value content would be identified and maintained for the long term, medium value content "preserved" through regular use. With content perceived to be of lesser value, there could be a more ambivalent attitude, for example with draft versions of documents. Marshall proposed some general principles: that it was easier to keep than to cull, but (paradoxically) easier to lose than to maintain. These principles seem to sum up the digital preservation conundrum very well!

The third presentation was by **David Pearson** of the National Library of Australia, a reflection on several years experience of dealing with digital content at the library. Pearson commented that purely reactive collection policies would mean that libraries would either receive no content or would become swamped! The final presentation, by **Andreas Rauber** of the Vienna University of Technology, was a short overview of the assumptions and design decisions that underlay the HOPPLA (Home Office Painless Persistent Long-term Archiving) system.[7]

Scientific and Humanities Data Curation

Another session in the same set featured three papers examining practices with regard to digital data in science and the humanities. **Wendy Duff** of the University of Toronto started the session, reporting on a survey of teamwork among digital librarians and digital humanities researchers. Respondents cited as reasons for working in teams: a wider skill set, greater productivity, greater methodological variety, greater available effort, greater capacity for data study, and enjoyment. Team communication tended to be either through e-mail or face-to-face, with little use of wikis and other groupware tools. The main challenges were mostly administrative, although personality clashes and mutual recognition of skills were also cited. The survey revealed a need for teams to budget for collective sense-making, a need for better advance planning and governance, and a need for more training in teamwork, perhaps in the form of apprenticeships.

Jinfang Niu of the University of Michigan presented the results of testing a Documentation Evaluation Model for social science data. In a survey of 387 researchers, she verified that sufficiency and ease of use were perceived as indicators of documentation quality. She noted, though, that these two factors were not independent of the audience of the documentation; for example, perceived ease of use is strongly affected by the reader's absorptive capacity, itself dependent on a number of factors such as prior experience and background knowledge.

The third presentation was from **David Giaretta** of the Science and Technology Facilities Council (STFC), a UK-based agency. He outlined something of the history of data preservation at the European Space Agency (ESA); initially preservation had not been much considered, as improvements in data quality tended to make older data obsolete, but now ESA has dedicated funds for preservation. It does not normally preserve processed data, but only the raw data and the means to derive processed data on demand. ESA is participating in three digital curation initiatives. In the CASPAR

---

[7] Home Office Painless Persistent Long-term Archiving (HOPPLA)
http://www.ifs.tuwien.ac.at/dp/hoppla/

Project, it is concerned with testing preservation strategies by simulating changes in software, hardware and the community of data users. In the PARSE.Insight Project, it is contributing to surveys and case studies within space science, with the aim of producing a preservation infrastructure roadmap. Finally, it is a member of the Alliance for Permanent Access.

Skills for Significant Properties: Debating Pragmatics and Philosophy in an Area of Digital Curation

The final set of sessions included a panel looking at the balance of practical and theoretical skills required by digital curators, with particular reference to the field of significant properties. The panel was chaired by **Stephen Grace** (King's College, London), who began by arguing that significant properties are the 'soul' of digital preservation, with aspects such as characterization (file format conformance), diplomatics (judgements made in the course of analysing files), epistemology (justifying beliefs about a digital object) and conceptualization (computers as venues of interpretation and representation).

Also on the panel was **Christopher "Cal" Lee** (UNC Chapel Hill), who argued that as digital objects are in essence hierarchies of symbols, philosophy – as a study of choosing the right symbols – had much to contribute to digital curation. For example, the issue of significant properties was rather like the issue of personal identity: what properties must a digital object retain in order to remain the same thing?

Finally, **Sheila Anderson** (King's College, London) argued that as significant properties are mostly decided by people in their own contexts, one really needs to consider digital objects as part of a social-technical framework: the object level is really too fine. She also compared data models and standards to boundary objects in sociology: objects that exist in different ways in different communities, yet are sufficiently similar to provide a point of translation between them.

Among the views expressed in the ensuing debate, was one that significant properties are (by definition) the criteria for determining if an object coming out of the preservation "black box" is the same one that went into it. These criteria do not have to be universal, but they need to be clear to those providing the preservation service, so they can manage loss to the satisfaction of the users. Some pursued the argument that an object focus is insufficient. For digital objects, every act of access is a transformation or interpretation, creating a new object; therefore, one should focus on what makes two experiences of a digital object similar enough. In terms of the curriculum, it was argued, students need to made aware of the issues, and to appreciate how documenting significant properties allows them to distinguish cases of losing properties by mistake, and losing properties as part of a justifiable, reasoned methodology.

Digital Curation Tools and Strategies

One of the other parallel sessions included the presentation of four papers on digital curation tools and strategies, chaired by **Adrian Cunningham** of the National Archives of Australia.

David Giaretta of STFC gave the opening presentation, an overview of preservation workflows, strategies and infrastructures that built on concepts developed in the Reference Model for an Open Archival Information System (OAIS). In thinking

about conformance to the model, Giaretta argued that the OAIS Information Model was the key to successful preservation, retaining knowledge about content, rather than just the "bits." He mentioned that an updated version of the model would be published shortly, including revised perspectives on important things like authenticity and significant properties. He then provided an overview of preservation workflows developed by the European Union-funded CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval) Project, focusing on the various components needed to support preservation from the perspectives of data producers, curators and consumers. If nothing else, the CASPAR preservation workflow diagrams illustrated the extremely complex dependencies that need to be addressed when developing preservation strategies for particular kinds of content or user (designated communities).

Dimitris Gavrilis of the "Athena" Research Centre in Greece then gave a short overview of the MOPSEUS digital library system, developed to support the specific needs of small libraries. The premise of the system was that some smaller organisations - e.g. public or school libraries – would neither be able to afford commercial digital library systems or have the technical expertise to set up (or support) open-source software solutions themselves. The MOPSEUS digital library service is based on Fedora, but has been especially configured to be simple to set up and use.

Robin Rice of EDINA and the Data Library at the University of Edinburgh, then provided an outline of recent work supporting the curation of research data in the United Kingdom. She first introduced DISC-UK (Data Information Specialists Committee – United Kingdom), a group set up in 2004 to support data professionals working in UK Higher Education.[8] Rice explained that DISC-UK had just completed the DataShare Project, a "Repositories Enhancement" project funded by the Joint Information Systems Committee. DataShare had involved the setting up of exemplar institutional research data repositories at each of the four partner institutions (the universities of Edinburgh, Oxford, and Southampton, and the London School of Economics) and had been interested in exploring the implications that data curation services would have for other institution-based services (primarily institutional repositories). The four partners had all adopted different technical approaches to the development of their repositories and had selected multiple ways of interacting with institutional publications repositories. Three of the partners had also conducted analyses of institutional data curation requirements by means of the Data Audit Framework (DAF).[9] Lessons learned from DataShare included some insight into researcher motivations for data sharing (and management), and the potential importance of data professionals and institutional infrastructure (like repositories). Rice's final warning was for delegates not to conduct institutional audits unless they were prepared to open a can of data management worms!

Finally, **Mike Smorul** of the University of Maryland, gave a presentation on the implementation of a collection-monitoring tool called the Audit Control Environment (ACE). This uses cryptographic techniques to continually monitor and address the various threats to the integrity of digital archives. The system had been successfully tested on a number of different collections, both distributed and centralised.

---

[8] DISC-UK http://www.disc-uk.org/
[9] Data Audit Framework http://www.data-audit.eu/

Closing Plenary

In place of a closing keynote, the final plenary of the Symposium invited delegates to comment on what should happen next. One person suggested running refresher courses on digital curation for existing professionals; this may have been tongue-in-cheek, as UNC Chapel Hill will be running such a course in the summer of 2009, and is encouraging other institutions to hold the course as well. Another suggestion was that the next DigCCurr should attempt to attract more delegates from the museums sector.

Conclusions

DigCCurr 2009 featured an impressive array of speakers, in an ambitious programme covering everything from practical preservation tools to the underlying philosophy of digital curation. The curricular aspect to the discussions is something that sets this symposium series apart from other conferences of its kind, and this is no bad thing; given the old adage that the best way to understand something is to explain it to others, one can see in the development of a digital curation curriculum a sign of a discipline that is reaching maturity.

As with the previous event (Day, 2007), the full programme meant that it was not possible for one person, or even two people, to gain a comprehensive overview of the entire symposium. Nevertheless, there was a sense of things falling into place, particularly with regard to the technical infrastructure for digital curation. The papers and panels certainly gave the impression that, with the tools now available, the digital curation community is focussing on getting people to use them, and on embedding good practice within the workflows of academics, researchers, and people in general.

The programme for the symposium, along with a selection of speaker biographies, abstracts and slides from the presentations, is available from the DigCCurr 2009 website.[10] The proceedings are available for download or print-on-demand from Lulu.com (Tibbo, Hank, Lee & Clemens, 2009).

References

Allinson, J., Johnston, P., & Powell, A. (2007). A Dublin Core application profile for scholarly works. *Ariadne 50*. Retrieved May 8, 2009, from http://www.ariadne.ac.uk/issue50/allinson-et-al/

Brandt, D. S., Carlson, J., Cragin, M., Heidorn, B, Palmer, C., Shreeves, S., & Witt, M. (2008). Investigating data curation profiles across multiple research disciplines [poster]. *4th International Digital Curation Conference, Edinburgh, 1-3 December 2008*. Retrieved May 7, 2009, from http://www.dcc.ac.uk/events/dcc-2008/programme/posters/Data_Curation_Profiles.pdf

Day, M. (2007). Report from the DigCCurr 2007 International Symposium on Digital Curation, Chapel Hill, NC, April 18-20, 2007. *International Journal of Digital Curation 2* (1), 102–111. Retrieved May 8, 2009, from http://www.ijdc.net/index.php/ijdc/issue/view/2

---

[10] DigCCurr 2009 http://www.ils.unc.edu/digccurr2009/

Interagency Working Group on Digital Data to the National Science and Technology Council. (2009). *Harnessing the power of digital data for science and society.* Retrieved May 7, 2009, from http://www.nitrd.gov/About/Harnessing_Power.aspx

Neuroth, H, Liegmann, H., Oßwald, A., Scheffel, R., Jehn, M., & Strathmann, S. (Eds.). (2008). *nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*, v 1.5. Göttingen: Niedersächsische Staats- und Universitätsbibliothek Göttingen. Retrieved May 7, 2009, from http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf

Tibbo, H. R., Hank, C., Lee, C. A., & Clemens, R. (Eds.). (2009). *Proceedings of DigCCurr 2009. Digital Curation: Practice, Promise and Prospects. University of North Carolina at Chapel Hill, NC USA.* Retrieved May 8, 2009, from http://stores.lulu.com/DigCCurr2009