

ICE Forum: June 2011

1. Introduction

On 29th June 2011 JISC sponsored a meeting called the International Curation Education – ICE- Forum at University College London which brought together university and professional educators and trainers in broad areas of digital preservation. Around 70 people were in attendance. WK, AD and SM attended on behalf of the DPC and many members were present in their own right including Kevin Ashley, Marieke Guy and Joy Davidson (DCC), Neil Grindley and Simon Hodson (JISC), Richard Davis, Patricia Sleeman Ed Pinsent and Frank Steiner (ULCC), Maureen Pennock (BL), Sarah Higgins (ARA), Laura Molloy (HATII), Neil Beagrie,

These informal notes are intended to give DPC members an informal briefing. They are not intended as an official record.

2. Helen Tibbo, University of North Carolina at Chapel Hill– Educating the Curator: Curation Education in the United States

Helen introduced the day by reviewing progress in education in the US looking at previous and current work in the US. Previous projects included the DigCCurr 1 and 2 projects, ESOP1 21, Closing the Digital Curation Gap and the Certificate in Digital Curation - all of which have been led or supported by the University of North Carolina at Chapel Hill and many supported by IMLS. The University of Illinois at Urbana Champaign, University of Arizona, UCLA, University of Michigan, University of Maryland, Purdue University, University of North Texas, University of Tennessee and the Pratt Institute in NY have all also had a variety of programmes in this field. These are quite mixed in their programmes and outcomes but they point to a range of professional development for students, early years professionals and PhD studentships. There is a mix of scientific and humanities skills and they draw in museums, archives, libraries and scientific research institutes. There were representatives of almost all of these courses in the audience.

3. Michael Seadle, Humboldt University – Why do we need people who can do digital ‘curation’?

Much of what we talk about under the heading of digital ‘curation’ splits between traditional librarianship and/or digital humanities and basic computer science. There’s a balance and a split between these two; if you don’t have the IT skills you can’t build the systems: and if you don’t understand the ‘why’ then you’ll never build systems that are useful. We tend to treat the time element of curation as a computing problem and assume that reading in 2111 will be the same as in 2011 or in 1911. But we know that people read differently in different periods. So curators need to address the problem of cultural problems of reading and how it changes over time. Actually, our capacity and methods of making judgements and making sense changes too – not just reading. So there is a challenge of ‘digital cultural migration’ which will make sure that future users and reader can not only open content but understand its meaning and intent. It sounds flakey but actually what we do already: Oliver Twist for example was written as a serial in monthly instalments. But it’s now published most often as a single edited and published volume with copious notes: a format migration and content migration with some additional data to help sense making. Jokes, reportage, dialect and parodies only make sense with context. So the point is that there’s a lot more to our job than recognising and implementing format migrations. Triggers for intervention – and perhaps tools - need to extend to changes in meaning in context and cultural references. We can recognise some clues and examples based in strings and patterns. There are also some clues and some probabilistic matching possible through examination of context which can be programmed in artificial intelligence. Human intervention is already hard to do though more ‘dependable’ and it’s getting harder; machine based approaches are easier but not very good. So when it comes to where and when we need to train – we need to train them to think about how to spot times and places where cultural migrations may be needed, how search and matching algorithms find and process things. That means library and archive schools we need to understand comp science a lot more: and we need to design software that addresses human problems not just technical ones. And given the volumes and complexities we face that means larger scale and more complex tools too.

4. Comments after this session

- Re-interpretation has not been traditionally the role of the archivist – it has been the job of others to extract and research collections. Archivists ensure safe custody and integrity of collections. Maintaining information and understanding over time are not necessarily the same job. Perhaps

there's a distinction to draw around DIP/AIP/SIP. Also a distinction between future proofing (early intervention) and just in time processing.

- Provenance and context is also changing. Linked data is part of a solution because we can retain the context in new ways
- Linked data is a religion and not sustainable through time – the links need to make sense through time also.
- Libraries are not resourced or skilled to address this issue either – subject librarians are increasingly under pressure and even in research institutions there is unlikely to be scope for this work. Annotation through users is possible and works well in some contexts - but that's not scalable because of the wall of data.
- There's no such thing as a 'traditional' library or archive any more. We need to endorse and accept the idea that these roles no longer really exist. The boundaries between professions are not really sensible now and the relationships between them need to adapt.
- WK – IT'S THE QUALITY AND SIZE OF THE CROWD THAT MATTERS and the extent to which it is focussed and incentivised around sensible issues
- Appraisal plays a part here too. The scalability issue can be handled in different ways from just automation.
- Context helps meaning making and so if an archive retains a strong control of context then one can always match the skills and cultural meanings with the user. There are very many tools that can be brought to bear this way.
- Authenticity – let archivists be archivists: authenticity and integrity is essential and if we teach anything else then we're on thin 'ICE'.

5. Alan Bell, CAIS of Dundee – Are our qualifications in records management / archiving still working

We need to provide prospective employers with prospective employees with the right skills. The evidence shows that in 1996 only around 18% of the skills which employers are looking for – even in an archive – are about records management or archiving. The core skills are therefore small in comparison with generic skills about report writing, budget, finance and what not. The truth is not just about whether the qualifications are fit for purpose, but whether the teaching staff are competent to transform them. The academic staff are not necessarily able to keep their skills up let alone update their courses. Dundee has therefore devolved its teaching by bringing experts into their programme to write the materials for distance learning. That gives them access to 30 tutors who deliver teaching to standards defined by a small core teaching staff. It means real diversity and strength in the expertise to ensure supported, flexible, valid, relevant and scaleable teaching. That's a tough job but it's based on 'knowing what we don't know'.

6. Steve Hitchcock, University of Southampton – training repository manager

This is a large topic and it's hard to crush it into a short presentation. The skills for repository managers revolve around scale, costs and rights and for this presentation we will concentrate on scale. Scale is a well rehearsed problem and there's are some big numbers here. Content in academic repositories is more clearly understood and the trend is towards research data over electronic documents. Remember that the repository layer assumes a replication and preservation function too. Examining the current stats, we can see is that the repository layer is much much larger just now than the archival layer. Repository managers are well placed to help with the selection, appraisal and filtering of content before it hits the long term preservation layer. Understanding the relationship between the two is central to the development of preservation services and better repositories. Ultimately repository managers may not be specifically charged with preservation. They can outsource some of it but they cannot outsource responsibility. The tools for preservation can be put into place using freely and widely available tools.

7. Gordon McKenna, Europeana– cultural heritage collections

Europeana is the 'European Digital Library' and provides access to all of Europe's culture. It exists to inspire the creation of new content and to encourage new types of access. Large collections of digitised items and interface in 27 languages. It drives users to providers' sites and it acts as an aggregator and 'enricher' of content and it provides a multi-lingual methodology for access. Culture Grid is the UK National Library of Data and it works with Europeana.

8. Simon Hodson, Curation of research data and curation education

There is clearly a need for many people to be involved in digital preservation but when it comes to the preservation of research data there's a clear need for researchers to be involved. There's a strong case for the need to preserve research data and there are a variety of external drivers for re-use, access and aggregating data sets as well as the 'sticks' of the research council mandates and policies on research data. Research councils have not been clear on the topic of what actually constitutes research data and that's for good reasons. JISC has funded projects in this space and there have been some very good approaches to scoping the scale of the problem. The SUDAMIH project for example had an explicit and wide statement about 'the whole range of materials that researchers must assemble and analyse in order to produce research outputs', while others took different approaches. Of course this introduces issues of research data policies and their implementation. It provides the basis for selecting and potentially deleting content. 'This is all very nice but how do we persuade researchers to give a damn?' There is a real issue of awareness raising and information management skills. Researchers want to do research after all and we shouldn't prevent them from doing so! There is a continuum from data management through to preservation. The programme produced some interesting training materials in their own right – Incremental and Sudamih for example but also DCC guides. A number of projects produced specific results and feedback. JISC will continue in this work and provide support for a range of research data management training.

9. Joy Davidson, Collaboration what can we do next and most easily

We've never needed much persuasion to collaborate but it's noticeable that we sometimes don't see some sectors. We need measures and benchmarks across the board for preservation and there's a specific need for benchmarking for educational resources and courses. We're not clear what employers are looking for and whether we are building courses the right way round.

10. Kate Fernie, DigiCurve

This is an EC funded project focussing on the cultural sector. It is identifying and analysing existing training opportunities in training in digital preservation and related. Europe has something like 47 national libraries and many thousands of smaller local and public/university libraries. There are lots and lots of people involved and they have needs for training. There are well established courses in traditional skills but long term access is a new area. True professional development is a lifetime requirement especially in this highly dynamic field. There is quite a lot of opportunities here and there are new skills for new jobs – at least new job descriptions. The project started with a survey of what is currently available. Early results show responses from Europe and America but and cross domain responses. The courses are very diverse from 'getting started in digital preservation', through to more advanced courses, subject specialist courses, legislation, planning, workflows and what not. Mainly in-person and large group workshops with lectures and practicals rather than genuinely online. They are mainly delivered by specialists rather than generic staff trainers and few of the courses have any kind of certification or assessment. A companion study will follow shortly reviewing the cultural institutions. New members can join the network – and DPC is a founding member.

11. Cal Lee, University of North Carolina – what do you really care about if you care about digital preservation

'Too many miss how different architectures embed different values'.... This means giving voice to individuals, ensuring accountability, awareness about the past, integrity of processes. Digital resources exist in interacting components which each exist in layers but it's not clear which of these layers are canonical – that varies in different settings and there are different types of interaction with them. So you can engage with the physical manifestation of data on a carrier with an electron microscope, you can use it in a filesystem, or through a hex editor or as an aggregate in some way. There's ethics and values (and rights) all the way down.

12. Seamus Ross, University of Toronto – Educating and validating the abilities of emerging digital management professionals

There is so much you need to do to play in this field. It's like a decathlon. Knowledge of represented information is vital but also note that 'software is a cultural artefact' and we need to understand both. Judy Blake was quoted by Katherine Sanderson as saying that 'Curators do novel work that is required by everyone doing science'. There's been a lot of discussion whether we need subject specialists or technologists – but it's becoming clear that subject specialisms are more important than generic

technologists because, at least as far as scientific data is concerned, it is a lot harder to train a scientist. We also need to train consumers and producers of data. There's an inter-relatedness of activities however and there's a need to work across a range of boundaries. We need case studies, we need access to a diversity of tools, we need experimental laboratories, realistic quantities and diversities of digital information created in different environments, public and private partnerships for placements and we need apprenticeships. What we need to know is not always obvious – project management, risk analysis, policy definition implementation and monitoring, repository design and management, workflow definition, technology, and audit. So a dp person needs to think like a humanist, behave like an engineer, ethics and subject knowledge.

13. About this document

Version 1	Document initiated	29/06/2011	WK
Version 2	Released to members	01/07/2011	
Version 3			