Going digital, experiences at the Wellcome Library

Dave Thompson

Digital Curator, Wellcome Library

https://docs.lib.purdue.edu/iatul/2008/papers/13

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
GOING DIGITAL, EXPERIENCES AT THE WELLCOMe LIBRARY

Dave Thompson, Digital Curator, Wellcome Library
d.thompson@wellcome.ac.uk

Abstract:

The Wellcome Library acknowledges that born digital material will form part of its research collections. The question is, how do we implement this? This paper talks to the experiences of the Wellcome Library as it develops and implements its strategy to acquire and manage born digital material. It draws on our experiences of working with ‘real’ material, and looks at the pragmatic ways in which workflows and practices have been developed and the archivists successfully engaged in working with this new medium.

Key words:

Wellcome Library, digital curation, digital preservation, archives

The Wellcome Library

Through its collections and services, the Wellcome Library provides insight and information to anyone seeking to understand medicine and its role in society, past and present. More than 30,000 readers visited us last year, including historians, academics, students, health professionals and consumers, journalists, artists and members of the general public.

Part of the Wellcome Collection, a major new £30 million public venue developed by the Wellcome Trust on the Euston Road in central London, the Library has over 750,000 books and journals, an extensive range of manuscripts, archives and films, and more than 250,000 pictures. We are one of the world's major resources for the study of medical history and we also provide access to a growing collection of contemporary biomedical information resources relating to consumer health, popular science, biomedical ethics and the public understanding of science.

The Wellcome Library goes digital

The Library acknowledges that born digital material, and especially born digital archival material will form part of its future research collections. The commitment to this form of material is embodied in its Library Strategy 2006 – 2009. Increasingly, archival materials are produced in digital form only; today's email & word processed documents are tomorrow's research material. This increasingly digital environment is recognised as a major driver of change for the Library. If we don’t engage with born digital material then Library collection growth is going to be limited, and the continued provision of research material will eventually stop. However, ‘going digital’ presents many challenges, none of which are easily overcome, but digital material is inherently flexible, and can provide new forms of access & dissemination and there are an increasing number of tools can help us build new services

What kind of material are we talking about?

There is almost no limit, beyond the practicalities of management and access, to the types of material that are now being produced in digital form. From an archival perspective much of the material that has always formed archival collections is now produced digitally. Examples abound, organisational records, personal correspondence, diaries, laboratory note books, administrative records. In communicating our intentions to collect digital material to our donor/creator community the Library sees it as important that it emphasises the continuity of the acquisition of materials, it is the same kind of material that donor/creators have always provided us with. Only now we are preparing to accept it in digital form. This also opens up the
possibilities of acquiring new forms of purely digital communication such as datasets, computer programmes, blogs and wiki.

The digital problem

The issues around the long term acquisition and management of digital material are well known, are well documented and will not be not discussed here. There are many papers discussing hardware or software obsolescence, format migration etc. and they are easily found. There are many conceptual and competing frameworks for dealing with digital material, many formal standards exist or are in development. We struggle to chose which to use and to decide which is the most appropriate. The digital preservation problem is increasingly about fitting the theories into practice and finding/selecting the right or appropriate tools to serve those theories.

Our problem becomes how do our institutions identify & apply most appropriate ‘solutions’ when none is a ‘complete’ solution? Whilst it is good that formal standards exist and that more are being developed, this immature field can present overwhelming confusion to institutions just beginning to acquire their first digital material.

The Wellcome Library Digital ‘Nike’ strategy

The Library has taken a new approach to resolving the issues it faces, one that is not based around the use of a single project and one that does not treat digital material as being ‘different’. In short the Library simply began to accept born digital material from its donor/creator community. We knew we would be looking to create a hybrid solution that created both digital and physical collections, we know that material continues to be created on paper alongside digital material. The clear message that we are sending to our users and donor/creators is that digital material isn’t ‘instead of’, it’s ‘as well as’. This assures all our user community – and staff too – that the core business of the Library isn’t changing.

How does this strategy work? In effect we are learning by doing, acquiring ‘real’ digital material from our donor/creator community and working out how to deal with it as we go. In practice this means that we work closely with certain donor/creators to discuss with them how they might transfer digital materials to us, what material we are prepared and able to accept, and where the balance lies between digital and physical materials.

We’re also not working in isolation, we have engaged in projects and ad hoc collaboration with other institutions to look at aspects of working with digital materials. We have collaborated with the Bodleian Library Paradigm Project, and with the National Library of Wales. Collaborative activities like these have proved to be a valuable way to build capability and gain experience in a field with which we have little experience as yet. Thus far our approach has worked, though it is not entirely without risk, we have not yet met any virus infected material, and we have not yet ‘lost’ or ‘destroyed’ any material through our inexperience. Only time will tell if this strategy continues to be an appropriate one.

The framework for our work

If we don’t have a single digital project running we do have a broad framework that supports our work. This is our Digital Curation in Action project, begun in 2005. It isn’t a formal project as such, more a series of long term phased activities examining how the Library will bring born digital material into its collections and from whom we will acquire material. It is also identifying the issues around the long term storage and management of digital material and beginning to develop workflows that identify key tasks and assign responsibilities. Under this framework there has been some practical testing of tools available for managing born digital material, we have a developmental instance of the Fedora digital object repository, and are working with tools such as DROID/PRONOM and the National Library of New Zealand metadata extraction tool. By these means we have begun to develop the formal business rules for acquiring, managing & providing access to digital material. Again, the use of a ‘rolling’ framework has proved to be a successful way to approach the many issues to be faced.

Archivists do rule the world
One of our key principles and core pillars of our work is that sound archival practice is very relevant when applied to the management of digital material. The archival process already contains many of the necessary processes, eg the use of formal donor loan agreements, early communication with donor/creators, the maintenance of an ongoing relationship, etc. The process of ‘description’ is also more appropriate than ‘cataloguing’ in the provision of essential metadata describing digital materials. Hence the support of the archival staff is essential, going digital is changing their work patterns. In the Library we enjoy considerable support from our archivists for our digital agenda and we are learning that archival workflows can be modified to accommodate digital material without undermining professional principles.

Using archival practice also allows Library to use the existing skills of staff to build a sustainable foundation for working with digital material. There are few public training courses, little time to require staff to read additional journals or other literature and only limited resources for staff development. Using what we already know is making economic as well as professional sense.

The archivists are supported by one new full time employee, a Digital Curator, whose role is to examine and adapt archival practice and support professional archival practice by resolving technical issues. This role is partly IT person and partly archivist/librarian. The digital curator is also the driver for the Library’s going digital strategy which frees the archivists to concentrate on the work of collection development. To date this has proved another robust and flexible approach.

So how will we work with digital material

We propose to use methods already adopted by others, a kind of best practice. At this stage the plan is kept deliberately simple and forms the basis for future learning.

We will always retain an unmodified ‘original’ bitstream permanently, regardless of its format, regardless of issues of obsolescence etc. All material will be removed from portable media – CD-Rom, disk, tape etc, and stored on network drives. This removes the need for complex and expensive media migration exercises, and we do not yet have such volume of material that this will cause storage problems.

For current material, eg current MS Word documents, we will produce a ‘preservation master’ manifestation from the ‘original’. The preservation master will be the focus of all preservation efforts and activities. It will be migrated as necessary and will have maximum metadata describing it. Any and all dissemination manifestations will be derived from our preservation master.

For legacy or obsolete material, eg WordStar or early forms of MS Word documents, the proposal is to normalise these to an ‘open’ format, eg OpenOffice, which is both an open standard file format and a current one. This normalised manifestation will be our ‘preservation master’, again this will be actively managed and dissemination manifestations derived from it. For both originals and preservation masters we will create descriptive & technical metadata as the basis for resource discovery & future life cycle management. This metadata will remain ‘fixed’ for original manifestations but will reflect the active preservation management that our preservation master manifestations receive. Preservation masters will be subject to a Technology Watch process that will monitor for material ‘at risk’, eg of impending obsolescence, and prompt for some form of preservation intervention.

How are we supporting this process?

The Library is implementing donor/creator education and awareness activities, promoting that we are now collecting born digital material and providing information about how we work. For instance we have a Digital Curation section on our Library’s website. This includes a number of FAQs that talk to our work and provide basic answer to questions that we are commonly asked.
Staff development/training is an on-going activity and working with digital materials is beginning to be included in our professional development. Digital issues, processes and activities are presented at regular staff meetings and details of relevant public courses circulated to key staff. The Library also has membership of both the Digital Curation Centre and Digital Preservation Coalition. Both these organisations provide support in the form of papers, journals and workshops for organisations ‘going digital’. Membership also provides crucial access to peers in other organisations who are working on the same issues and who are prepared to share their experiences. At a more basic level there is good internal communication and sharing of ideas through regular meetings, and this has helped us be prepared to change and to be flexible as we learn or as we face particular difficulties.

A practical example of a workflow

A simplified practical example of a workflow we use for new digital acquisitions can be expressed like this,

• Create activity log for each transfer
• Acknowledge receipt of material by letter, enclosing FAQ outlining what our processes are
• Anti-virus check – month in quarantine - anti-virus check - confirm with donor/creator material as ‘clean’, or otherwise
• Create manifest listing all material in transfer & send to donor/creator, transfer donation to network storage & verify
• Create (Manually) TechMD, eg DROID, checksums & file on network with activity log & copies of correspondence etc
• Create working ‘copy’ of material for archivists for arrangement & description, either on CD or network fileshare
• Send transfer media, CD-Rom, disk, tape etc, to our off site storage facility for 18 months, just in case we require it again, and then dispose

This is our current practice. Whilst this workflow works for now it has not been tested on large volumes of material, on obsolete or ‘difficult’ material. Again, only time will tell if this workflow needs to be expanded, modified or completely replaced, but we are prepared to be flexible and change as experience dictates.

Technical issues

Going digital is not easy, it’s not cheap and it does require some changes in work practice. Currently in the Library we don’t have a full suite of end to end management tools that will help us acquire, ingest, manage and make available the material we are acquiring. Hence much of our current work is done manually and so is both time consuming and costly. For instance we have no integrated way to automatically create technical metadata, it is currently created manually using Droid/PRONOM. This will change as we begin to automate processes with the purchase in 2008 of a digital object repository.

Appraisal of large volumes of digital material is especially complex and time consuming. The necessity to view digital material with a computer and appropriate software means that up to now we have created bespoke processes and systems for appraisal as the material demands. In the long term this isn’t a sustainable option. Learning how to work with digital material without breaking it takes time and practice, something we are getting better at. This is a clear benefit of working with real material – we are learning real lessons.

Problems

Issues of working with digital material are not the only problems we face. We have no mandate to collect any material, we are not a legal deposit library, our donor/creator community has no obligation to transfer anything to us. This means that to collect material we have to work hard with our donor/creator community to secure material for our collections. This is less of an issue for physical material, but we need to intervene more quickly with digital material to avoid issues of obsolescence. To collect useful contextual metadata means we need to emphasise the
urgency of digital preservation and work directly with those who created the material. Our typical archival ‘lag’, the time from creation to deposit, is between 10 and 15 years, too long for digital material, and could mean that every transfer to the Library is of obsolete material.

We are also beginning to see evidence also of changing roles within organisations. Especially so between Archivist/Records Manager and IT Managers in relation to digital material. Archives managers seem unfamiliar with providing or transferring digital material to us, it appears to make them nervous with issues around security, privacy and just the mechanics of preserving digital material. On the other hand IT managers, who may be responsible for maintaining, backing up and providing access to material for an organisation may not understand point of long term preservation, or the reasons why material is transferred to an archive or library. Often organisations have poor record keeping practices, making them ask how ours can be better, especially in the long term?

Acquisition issues

What we have found is that acquiring digital material is no more or less easy than acquiring physical material. We work closely with our donor/creator community and there is certainly a willingness to transfer digital material to us. On the other hand the mechanics of extracting material from systems, identifying it etc seems to be proving daunting. In the case of both physical and digital material, acquisition can be a time consuming process with no certainty of success, and it is certainly the case that we don’t yet have regularly established transfer schedules for digital material in the way we do for physical.

We have also noticed that our archival ‘lag’ is between 10 and 15 years, that is the time between creation of the material and its transfer to us. This is too long for digital material. Such a period increases the risks of both obsolescence and of accidental, or indeed deliberate loss. We are also finding it harder to acquire material from those in their in mid-career who may be less certain that material will have future significance, or who may feel more protective of their own work. We have no simple answers to these issues but it is interesting to note that these ‘social’ issues are as important as the technical issues.

What lessons have we learned?

We have acquired born digital material on the same principles as the other material in our collections, we collect it, manage it, make it available for the long term. Material must fit with our collection policy. Not treating digital material as ‘different’ has helped us place a new medium into a workable context.

On the other hand working with digital material is more difficult than we thought, digital preservation & life cycle management of this material isn’t a push button activity. It is more time consuming than we thought, the accession/ingestion process is time consuming and laborious especially when compared to that for physical material. We clearly need a new set of tools to help store, manage and make this material viable for the future and available to our users. To achieve this automation of processes will be essential, but can be difficult to build whole, end to end processes when the tools are not available. Collaboration and co-operation are successful strategies to help find ways around these ‘gaps’, examining what others are doing can save time and effort.

Interesting side effects

There have been some interesting side effects to our going digital. Experience with digital material is changing roles within the Library. It is both highlighting and strengthening the role of the archivists by stressing that their professional experience and practice provides much of the methodology for working with digital material. In turn this has lessened fears that additional staff resources, or expensive re-training will be necessary.

The focus on digital material is also raising expectations of what our library will hold in the future. There is genuine interest in plans to acquire material that has no physical manifestation
such as blogs, wiki and websites, as well as in the idea that digital material can be manipulated to provide or create new knowledge or products.

The different requirements of digital material are challenging ways we think about engaging with physical material, reassessing roles and responsibilities within the Library. This is changing the ways we want to record and document our interactions with physical material and its creators. The process of early intervention, so desirable for digital material is changing the ways we think of acquiring physical material, maybe early intervention for some physical material is also desirable.

Conclusion

‘Going digital’ presents many challenges, and not all have immediate solutions. What we have found is that by engaging with digital material we are learning valuable lessons. More so than if we had taken a more academic, or research based approach. In this way sound archival practice and a ‘just do it’ approach is proving to be a very flexible and relevant way to work. We learn by doing.

Experience with digital material is changing roles within the Library, and is changing professional practice. What we are doing is supported by our archivists and is seen as a positive opportunity and an extension of professional development. The consequence is that the different requirements of working with digital material are challenging the ways we think about engaging with physical material and so our practice evolves. Not everything we do is wholly ‘internal’; we have used projects and ad hoc collaborations with other institutions as a valuable way to build capability and gain experience. Not that we’ve ‘finished’, we still have much work to do before we can say with confidence that we can ‘properly’ manage born digital material long into the future.

References