

The International Journal of Digital Curation

Issue 1, Volume 3 | 2008

What to Preserve?: Significant Properties of Digital Objects

Report on the JISC/BL/DPC Workshop of April 7, 2008, British Library Conference Centre

Helen Hockx-Yu,
Planets Project Manager,
The British Library

Gareth Knight,
Centre for e-Research,
Kings College London

April 2008

Summary

This article reports on the JISC/BL/DPC workshop on significant properties, which took place on April 7, 2008 at the British Library Conference Centre, London. The intention of the workshop was to bring together the relevant projects and report on progress to date. It was also hoped that the workshop will lead to collective recommendations for future areas of research and development.

Introduction

“What to preserve”? has been a key conceptual problem facing those involved with digital preservation in the last decade. The concept of “significant properties” or “essential characteristics”, as a mechanism for determining the characteristics that must be preserved for digital objects to remain accessible and meaningful overtime, has been the focus of a number of notable digital preservation research projects in recent years. They include a range of projects funded by the Joint Information Systems Committee (JISC) and the European Union, in which work has been undertaken to investigate the factors affecting decisions on significant properties, to establish generic models for determining them, to develop tools and services for describing and extracting them, or simply to understand complex digital object types, using the concept of significant properties as a starting point.

Bringing together these projects and sharing their results and outcomes with the UK digital preservation and information management community formed the impetus of this workshop. Digital preservation is of strategic importance to all three organisers. JISC has been running an ongoing digital preservation programme and invested significantly in cutting-edge digital preservation research. The British Library has established a dedicated digital preservation team to address digital obsolescence issues and co-funded the Planets project¹ with the European Union to develop practical preservation tools and services for large-scale digital collections. The Digital Preservation Coalition (DPC) has been the cross-sectoral focal point in the UK for digital longevity and actively advocates joint actions to tackle the many challenges related to digital preservation. The significant properties workshop (Digital Preservation Coalition [DPC], 2008) was the second joint event by the same organisers. An earlier workshop on archiving of electronic journals took place in March 2007 and led to the funding of the e-Journal Registry Scoping Study, which investigates the feasibility and scope of a possible registry of archived e-journals².

The workshop was chaired by Chris Rusbridge, Director of the Digital Curation Centre, who did a meticulous job choreographing a packed programme. More than 140 people attended the workshop, from a wide range of sectors and professions. It was rare yet encouraging to see a workshop about digital preservation which was *not* predominantly attended by librarians and archivists. The diversity of the participants’ backgrounds signals the increasing awareness of digital preservation across sectors and professions. Interest in the topic, and in digital preservation in general, was cited as key motivation to attend the workshop by the majority of those who have provided feedback.

Notes from the Organisers

Neil Grindley welcomed the participants on behalf of JISC. He briefly introduced the organisers and explained the motivations to hold the workshop. JISC has the remit to support the innovative use of technology in UK teaching and research but they are looking to work with people and organisations outside the higher education sector and the UK so that its work will benefit the wider community. Seven of the 11 speakers were presenting work funded by JISC. Neil explained that the decision on what it is

¹ Planets <http://www.planets-project.eu/>

² JISC e-Journal Registry Scoping Study
http://www.jisc.ac.uk/whatwedo/programmes/programme_preservation/2008ejournalregstudy.aspx

exactly about digital objects that is worth preserving at a basic level might represent a choice between keeping and disposing of information, which in itself is not always entirely a straightforward choice. The purpose of the day was to focus on the most important characteristics of the digital objects which we wish to keep. Work to determine the significant properties of digital objects has been ongoing for some time. For JISC, the key funder of digital preservation research in the UK, it is important to fund effort that builds on earlier work, which will enable information managers and preservation specialists to gain detailed understanding of how the component parts of a wide variety of digital objects may need to be considered if they are to survive the ravages of time. Neil also asked the participants to help think about the next steps. Should more studies be conducted to understand significant properties of more digital object types, such as geospatial information, scientific and medical imaging and databases, or should investment go towards building tools that exploit our current understanding of significant properties?

Keynote: Significant Properties of Digital Objects

The keynote was given by **Dr. Andrew Wilson**, Director of Information Policy at the National Archives of Australia. Andrew has been credited, along with Adrian Brown of The National Archives with encouraging the renewed interest in significant properties in the JISC community. Andrew's keynote provided a summary of research into significant properties of digital objects during the previous ten years, highlighting the contribution made by the JISC-funded CEDARS and CAMiLEON projects in the late 1990s and early 2000s to defining the purpose of significant properties for digital preservation. Andrew highlighted the inherent difficulty of maintaining the significant properties of digital resources, indicating that a strategy of benign neglect is not an effective method for maintaining access to significant properties of digital resources. Instead, intervention is necessary at various stages of the digital lifecycle to ensure that the significant properties remain accessible and usable. However, he noted that the process of maintaining access to digital resources through migration or emulation must be carefully monitored, to ensure that the significant properties remain present and unchanged in the message that is intended to be communicated. Andrew concluded his keynote by asking delegates to consider the feasibility of identifying and cataloguing significant properties for different types of digital object in a production environment – a theme that was widely discussed throughout the day.

InSPECT: Investigating the Significant Properties of Electronic Content over Time

The opening presentation was given by **Stephen Grace**, Preservation Manager for the KCL Centre for e-Research (CeRch), who outlined the work being performed by CeRch and The National Archives in the JISC-funded InSPECT project. The project team are developing a framework and methodology that will assist institutions when evaluating the properties of a digital object that should be maintained. The project provides a framework for the identification and classification of significant properties based upon a four-stage process:

1. Define the intellectual components of a record that must be maintained.
2. Identify the technical properties of each component that are required to recreate it.
3. Classify the function performed by each property and assess its relative

value.

4. Measure each property through the use of a pre-defined assessment method.

Stephen indicated that the project methodology is informed by work in diplomatics, a type of forensic palaeography that seeks to identify the information that is necessary to establish the integrity and provenance of a digital record. There is a recognition that the properties of a digital resource are not necessarily of equal importance to the recreation of the information content. A core set of technical properties are required for each type of digital object (e.g. text in a word-processed document, the number of pixels in an image). However, other properties may be altered or removed with little impact upon the intellectual content that is being recreated.

The latter half of the presentation provided an outline of the measurement schematic that had been developed by the project team to catalogue and quantitatively measure the properties of a digital object. The draft schematic outlines a set of information that is considered useful for describing each property: the property title and definition; an indicator of the measurement method; and the facility to record the degree of variation that is allowed when recreating the property. Information on the significant properties of a digital object may be stored by a maintenance agency, such as The National Archives' PRONOM software or the Global Digital Format Registry (GDFR), or packaged with the digital object itself to assist a digital archive with an allotted task. As a testbed, Grace indicated the project will apply the measurement schematic to the description of properties attributed to four types of object: raster images, digital audio, presentational text documents and e-mails. The project will also seek to collaborate with similar projects to extend and revise the schematic to consider the application of significant properties for other types of digital objects. Finally, the project will integrate its outputs into the PRONOM database, to enable other institutions to identify the significant properties of objects for which they are responsible.

Significant Properties of Vector Images

David Duce presented the findings of the Significant Properties of Vector Images study, which was performed by Systems Simulations Limited and Oxford Brookes University. The project was the first of the recent JISC-funded four-month studies to publish its research into the significant properties of a specific type of digital object. The amount of visual data created as vector images has increased in recent years, through development and adoption of creation tools. Vector images may be created to visualise information as two- and three-dimensional still and moving images, as well as interactive resources. To limit the scope of the study, David indicated that the project team focused on the analysis of 2D vector images, with a secondary concern that the method may be extended to analyse other vector types at a later date.

David began by providing an overview of the NAA Performance model and its application to vector graphics. He expanded upon the issue raised by the InSPECT project that the definition of certain properties considered to be significant is often based on the assessment criteria of the assignee, suggesting that a key challenge of rendering a vector image to a new environment was to decide which components of the visualisation were considered to be important to recreate the intended meaning. A common issue encountered when considering visualisation was the potential difference

between the creator's intent and the viewer's interpretation. In some circumstances, a list of the significant aspects of a visualisation may differ if the process is performed by the creator, other artists or critics with an interest in the work, or archivists whose task it is to maintain the resource.

To formulate the significant properties of vector graphics, the project adopted the methodology outlined, ISO/IEC Computer Graphics Reference Model (ISO/IEC 11072:1992) – a conceptual model that specifies five inter-related layers of interaction between the software application and the user. He indicated that the “virtual scene”, which contains the core information (e.g. the co-ordinates of a shape or the underlying data values) may be considered essential for recreating the information content. Some elements of the stylistic representation of the data may also be useful. However, it may be necessary to assess these factors on a case-by-case basis. For example, colour may be used to communicate information on the interpretation of the visual image or for differentiation only. In circumstances in which the colour is a significant property of an image component, it should also be considered if the recreation of the colour requires that the colour shade is a precise match or if it is necessary only to indicate that it is a specific colour.

To conclude his presentation, David made several recommendations for digital archives that wish to preserve vector images and for future development work. He recommended that archives should minimise their reliance on potentially inaccurate assessments of each component in a vector image by storing the application data on which the visual image is based, in addition to the illustrative vector image. If the data is stored in a vector image, work should be done to convert it to one of several candidate formats, such as WebCGM, SVG and PDF/A that may be suitable for long-term access. Finally, he indicated that further work may be necessary to create tools able to extract the significant properties of vector images.

Significant Properties of Moving Images

Mike Stapleton, Technical Director of System Simulation Limited, presented the findings of the significant properties study into moving images. He began by providing an overview of the moving image market, highlighting the influence of technology on changing work practices. The presentation indicated that the significant properties of a moving image resource may be understood at different levels of granularity: a frame contains information that is essential for communication of its information content which, when merged into a sequence, will produce additional properties that must be considered. The properties associated with a frame share many similarities with earlier studies on the significant properties of raster and vector images – the width, height, bit-depth, aspect ratio, colour model and colour space are essential for the correct recreation of the visual appearance of a still image. Further properties emerge when a collection of still images are combined, often in conjunction with audio or text components: the inter-relationship of each frame must be defined to establish the sequential order of the footage; and the speed at which each frame should be displayed should be indicated by specifying the frame rate.

Mike went on to make several recommendations for digital archives that wish to curate moving image data. He indicated that good preservation practices required consideration of the production process. To maintain a high-quality moving image, a digital archive must obtain the digital master that has been produced by the creator and

should not be reliant on lower-quality derivatives that have been produced for distribution purposes. To enable existing hardware to store and playback moving image data, many file formats use compression algorithms to remove ‘superfluous’ information. However, the use of compression may introduce “smearing” and other emergent artefacts that were not present in the original. He argued that the relative importance of the information to the recreation should be a subjective choice that is made on a case-by-case basis and that appropriate quality metrics are necessary to establish if the compression information and other artefacts introduced into the frames are tolerable or otherwise, according to an objective or subjective measure.

Significant Properties of Software

Brian Matthews of the Science and Technology Facilities Council (STFC) presented the findings of the study on the significant properties of software packages. It is increasingly common for research projects funded by the JISC and other research councils to produce software tools that are used to perform specific functionality. However, software has received scant attention as a target for preservation by previous research studies. Software is a key component of the NAA Performance Model, on which each project has based its work. However, it is commonly considered as a method of rendering information content and is therefore interchangeable with similar software applications. Through the consideration of software as a target for preservation, the Performance Model may be applied to other preservation methods, such as emulation.

Brian began the presentation by providing some justification for the preservation of software. He suggested that software represented a considerable investment by research institutions. It provides evidence of the research outputs and work practices of an institution; may be used as a method of maintaining access to legacy data that has previously been created and which continues to be useful; and it allows continued use of software code that has been produced to perform a specialized function. The latter may be considered particularly important for research institutions that have produced in-house software intended for specialist purposes by a limited set of researchers. If access to the software and code base cannot be maintained, the institution may be required to reinvest time, knowledge and funding into the redevelopment of software for current systems.

A theme that was present throughout the presentation (and the associated study) was the need to maintain a flexible understanding of digital curation and preservation, to apply it to software packages. The lifecycle of software may be broadly comparable to other types of research data - software code is written in a development environment using a particular programming language and intended to fulfil a specific purpose. It is compiled at various stages in the production process, which binds it to a specific delivery environment. However, the strategy that is required to maintain access to the software and the associated significant properties that must be recorded will differ from other types of data. Source code allows greater flexibility, allowing an institution to rebuild it for a different software environment or re-implement it using current programming techniques (if sufficient documentation exists to understand it). For compiled software code, information on the environment in which it must be executed should be maintained. For example, the hardware architecture, operating system and dependencies are useful for recreating the software code through a virtual machine, emulation or using physical hardware. To conclude his presentation, Brian reiterated

that the preservation of software was at an early stage of development and that further work was necessary. However, he indicated that the most effective method of preserving software was to use good practices when developing software tools - a concept that many resource creators will find familiar.

SPeLOs: Significant Properties of E-learning Objects

The final presentation of the morning session was provided by **Richard Davis** of the University of London Computer Centre (ULCC) on the topic of significant properties for e-learning objects. Learning objects represent a particular type of digital object that must be considered as to its purpose and the environment in which it is used, in addition to its technical composition. A learning object may contain many different types of information, encoded in different file formats and designed for use through a software interface (e.g. Blackboard, Qmark, QTI, etc.) that is used for the purpose of learning and teaching by an appropriate authority, such as a teacher or publisher.

The project methodology is built upon the concept that a Learning Object (LO) must be understood through consideration of its intended purpose and use by the Designated Community. A notable argument made throughout the presentation was that the definition of properties as significant must be extended to consider other methods of use, in addition to requirements of access. To be considered an effective learning object, it must also be possible to repurpose and reuse the information content contained within the LO. This may require the recording of qualitative domain-specific information, such as an indicator of the purpose of the learning object (e.g. an instructional, collaborative, or assessment exercise) and the function that it performs in the course. The argument, which was also raised to varying degrees by previous presenters, proved to be an interesting topic that provoked some debate during the discussion sessions.

Richard went on to identify 26 types of properties that may be attributed to a learning object that must be maintained. The list included contextual information on the creation and assessment process, as well as details of the delivery mechanism. He indicated that the properties may be mapped into the InSPECT framework and that the SpeLOs project team intended to work with the InSPECT project to refine it to cater for the requirements of e-learning objects.

The morning presentations highlighted significant properties research-funded by JISC, which are mostly exploratory in nature. Much of this work is laying valuable theoretical foundations for more in-depth investigation. The afternoon presentations were intended to provide the wider context and some practical aspects of applying and implementing the concept of significant properties.

Content Characterisation in Planets

Adrian Brown of the National Archives (TNA) opened the afternoon session with a presentation on the content characterisation work within the Planets project. TNA is a partner in Planets, a four-year project coordinated by the British Library with 16 partners across Europe, bringing together expertise from national libraries and archives, leading research universities and technology companies. Content characterisation service is a key building block within the Planets architecture and

serves the purpose of understanding properties of digital objects which are significant to their long-term accessibility. Planets is developing solutions to automate the process of identifying, validating and extracting essential characteristics of digital objects. Adrian introduced the main areas of Planets' work, including a methodology for describing significant properties, a suite of tools and services for automatically characterising a range of specific object types in accordance with this methodology, and a supporting registry of characterisation information.

He then compared the OAIS model, in which the logical *information object* is derived from the physical *data object* via a process of interpretation using *Representation Information*, and the National Archives of Australia's Performance Model, wherein the *performance* (information object) is produced through the interpretation of a *source* (data object) by a *process* (representation information), and pointed out how closely the two models correspond to each other. In the context of the models, Adrian explained the difference between significant properties and representation properties (information). While the former are about the intellectual intent and apply to the abstract information object and properties of the intellectual intent, the latter are specific technical manifestations of the information object and apply to the data object, e.g. format, encoding schemes, algorithms. Successful preservation, he argued, requires both types of information. Adrian pointed to a white paper recently produced by Planets which explores the concept of Representation Information Registries, surveys the current state of the art and discusses the role of registries in Planets (Brown, [2008](#)).

Adrian then introduced in detail the PRONOM-based Planets Characterisation Registry which models both significant properties and representation information to support the complete preservation process. He presented a number of use cases and explained how the Registry serves as a repository of representation information, how it also acts as a tool registry and allows appropriate characterisation tools to be identified and executed. The Registry in addition has capabilities to describe significant properties, thus enabling validation of preservation actions by comparing the characteristics of digital objects (in a typical before-and-after scenario). The Planets characterisation framework uses third-party tools such as DROID and JHOVE but also includes the eXtensible Characterisation Languages which have been developed to support automatic validation of document conversions and evaluation of migration quality. The *eXtensible characterisation definition language* (XCDL) provides an abstract way of expressing object properties in a format-neutral way. The *eXtensible characterisation extraction language* (XCEL) defines how information for specific object formats can be extracted and described in XCDL. Planets has implemented an extractor component which extracts content from files as well as a comparator which is capable of identifying degrees of equality between XCDL documents (Becker, Rauber, Heydegger, Schnasse & Thaller, [2008](#)).

Adrian concluded the presentation by providing a look into the future, how additional functionality such as a risk assessment service will be added to support preservation planning and how the Registry will be fully integrated with other Planets components to provide pro-active preservation management capabilities.

Preserving Digital Assets: Issues and Considerations

Roger Lloyd presented the issues and considerations related to preserving digital assets at Barclays Wealth, the wealth management division of Barclays Bank PLC with over 6,900 staff in 20 different countries. To support such a large and complex business operation, data must be accessible, transferable and retrievable. Clients' data, as well as a vast amount of business records, are regularly transferred between different offices worldwide. This requires compliance with legislation and corporate rules to ensure protection of personal data. Data security is therefore also a key consideration.

When deciding what data to preserve and how, many factors need to be taken into account. Roger went through a list of business drivers and explained that the key issue is to protect the Barclays brand as a weak brand means weak business. Costs are equally vital and relate to many aspects, such as maintenance, sustainability, legacy media, access, marketing, risk, compliance and audit. The digital assets management solutions that have been put into place cover infrastructure, data handling and software. A common platform provides maximum data integrity but also allows for independent provenances of local offices. A company-wide classification scheme has been defined and implemented to improve data management and access. Data cleaning is performed to enhance the quality of data input, and regular reviews take place to re-evaluate the data retention schedule and to counter data obsolescence. A communication programme in addition has been set up to engage with senior management and responsible staff to obtain buy-in for proper data handling procedures.

Roger introduced the many software components and applications required to preserve and secure data. A key requirement is software performance, which is seen as critical to ensure digital objects remain accessible and meaningful.

SCARP Project: Digital Working and Disciplinary Factors in the Definition of Context and Significance

Colin Nelson of the Digital Curation Centre (DCC) presented the SCARP (Sharing Curation and Re-use Preservation) project, funded by JISC to understand disciplinary practices and research cultures in the areas of data deposit, sharing and re-use, curation and preservation, as a setting for development in digital curation. SCARP is conducting a set of case studies as examples of digital curation practices and opportunities in specific domains of knowledge. An immersive approach has been taken to work alongside researchers and practitioners to understand their curation needs. A final synthesis report will pull together the differences and common factors in explaining the variety of digital curation practices across disciplines. Findings of the studies will also be integrated with the work of the Digital Curation Centre.

In order to reify "digital objects", SCARP not only looks at digital information, but also at how researcher and practitioners interact with digital information, the disciplinary-specific way of digital working. Many factors have been examined, explained Colin, to comprehend the academic as well as the industrial context, including instrumentation, regulation compliance requirements, laboratory setting, field research, products and market. Colin then presented SCARP's effort to decipher the disciplinary landscape for the Architecture profession. After offering detailed

analysis of both the academic and industrial structure, Colin looked at the tools which support digital working, including commercial software and information models which ensure that digital information are in forms that permit use and re-use.

He then focused on lightweight representation in Product Lifecycle Management (PLM), a new-generation product representation aimed at supporting users at different stages of the product lifecycle in rapidly browsing, retrieving and manipulating product information. In recent years, lightweight representations have been successfully developed into commercial systems, and their applications cover many different industries. They however have limitations and cannot satisfy the information need of different users at the different stages throughout the product lifecycle. Researchers at University of Bath and UKOLN have proposed a new framework of lightweight representations for the product lifecycle which supports different lightweight representations for different users and partners, while at the same time maintaining the essential information. The framework integrates a markup method and the Representation Information Registry/Repository constructed by the DCC (Ding, Ball, Matthews, McMahon, & Patel, [2007](#)).

Colin concluded that the disciplinary context and structure of the industrial sector are crucial to understanding digital working. The significant characteristics of digital objects are defined by the use at each stage of the lifecycle for which the object was created. Decisions on the level of significance need to include the perspectives of the different designated users.

Digital Object Semantics

Stephen Rankin of the Science and Technology Facilities Council presented the semantics of digital objects on behalf of **David Giaretta**. Stephen opened the presentation by looking at a set of key OAIS concepts and saying that the ultimate goal for digital preservation is to ensure that information to be preserved is independently understandable to (and usable by) the designated community. *Representation Information*, which maps a *Data Object* into more meaningful concepts, is the key. An example of representation information for data stored in a FITS (Flexible Image Transport System) file may include the FITS standard which defines the format and a data dictionary which defines the meaning of keywords in the file which are not part of the standard. Significant properties of an information object often need to have specific representation information to denote how they are encoded.

Stephen then pointed out that the concept of representation information extends beyond the format of a data object. Knowledge of format alone is not sufficient to interpret a data object, especially complex scientific datasets. Stephen concluded that significant properties only give hints about some representation information but more is needed to interpret and understand digital objects. The amount of representation information is determined by the knowledge base of the designated user community, which needs to be clearly defined. So the key question about significant properties is: for whom are they significant?

Preserving Attachments from an e-Mail Collection: The Good, the Bad, the Ugly and the Thought-Provoking

Cal Lee of the School of Information & Library Science, University of North Carolina, was one of the researchers who worked on the CAMiLEON Project, which developed a model for expressing significant properties and used this to assess emulation as a digital preservation strategy (Hedstrom & Lee, 2002). Cal gave a presentation on the SigProp project, a partnership between the State Archives in North Carolina and Kansas, focusing on significant properties of office formats widely used in state government. He started with examining both the concepts of significant properties and representation information. Significant properties affect digital objects' quality, usability, rendering, and behaviour and often have a range of possible rather than discrete values. Representation information allows significant properties to be enacted (reproduced) in a given technical environment or set of environments. He also looked at four types of properties (supported, observed, measured and intended) and examined their implications within a preservation strategy.

Cal presented the good, the bad, the ugly and the thought-provoking aspects of preserving attachments from email messages in the SigProp project. The object of this challenging endeavour is the 41 Outlook personal folder files (.pst) of James B. Hunt, Jr, Governor of North Carolina between 1993 and 2001, which were transferred to the State Archives in 2002. The collection contains 61,973 messages and 13,746 attachments. Outlook Attachment Sniffer (plug-in to Outlook) was used to extract attachments from PST files which were then converted to Open Office formats. Many problems were encountered in this process. Some attachments had viruses and some had file names which were too long to be handled by the software. In addition, a daunting number of potential properties are supported by file format specifications such as the Microsoft Office Binary File formats. Cal explained the problems with the Representation Information for Word, which not only has many undefined fields but is rarely sufficient to reproduce an exact rendering of the text's character or paragraph properties. Much of the specification is about where to find representation information in a file rather than how software should handle, interpret or render it. Microsoft Word documents are bundled with many properties, such as conditional values, application state at time of save, properties for printing and a large variety of hidden data. The increasing realisation of hidden data has spawned massive industries for e-discovery and tools to remove data. The challenge is to decide what properties to preserve.

Cal observed that industry trends suggest that (at least some) hidden text is best not preserved. Microsoft itself has also brought changes to Office to abandon some hidden data in the recent releases. Cal ended the presentation by sharing the challenges, and suggesting ways of determining some of the observed and measured properties for preservation.

Where Do We Go From Here?

The last session on the workshop programme was a lively discussion chaired by Chris Rusbridge and contributed to by representatives of the organisers as well as the audience. It started with a summary by Andrew Wilson, who pointed out the complexity of determining significant properties for digital objects, and that it is an impossible task to carry out manually given the large volume of digital information with which we deal. Minimising human effort and automating as much as possible is

the way forward to address this complex component of digital preservation. Wilson then looked at the differences between the Archival Information Package (AIP) and the Dissemination Information Package (DIP) and argued that AIPs need to retain as many properties of a digital object as possible so that different DIPs can be generated, in any format, with or without certain properties.

Wilson's suggestion was supported by a number of participants. Because there is no way to foretell what properties the future users might consider significant, keeping as many flexible options as possible seems to make sense. Some however raised questions about the costs involved with this approach, and argued that the strategy should be keeping as many flexible options as one can afford.

A number of participants commented on the fact that it is neither desirable nor possible to keep all the properties of a digital object in its exact original form. Analogies were drawn with appraisal and selection, a process during which some documents and objects were chosen to be discarded, and with conventional publication, where the publishers choose a set of significant properties from the author and convey them to the readers correctly. Digital preservation will not produce the perfect copy. When moving from one technical platform to another, there will be changes. Even with emulation, the approach commonly regarded as being capable of not altering the original data in any way, differences in the emulated copy were observed. It is important to accept this and understand what characterises the intellectual creation, the work, which can be encoded technically in different ways.

An observation has been made about the fact that current research seems to focus only on fixed digital objects, rather than objects under state of change such as databases. In addition, little was mentioned about preservation of associations between objects. Experimental work in database preservation at the University of Edinburgh and within Planets was reported. It was also pointed out that activities within Web archiving have recognised, and are dealing with, the issue of preserving links.

Participants contributed actively to discuss how to take things forward. The key areas where future work was suggested included:

- Relationship and differences between representation information and significant properties
- Analysis of the needs of the designated community
- The level of granularity at which decisions are made about significant properties, e.g. instance level, format level, format category level. How do we manage this, taking into account economic as well as technological viability?
- Extended sets of properties of object types which have been investigated based on specific community needs, e.g. medical x-ray images as an extension of the vector images study
- Automatic tools that identify, extract significant properties and help assess or validate the outcome of preservation actions against original requirements
- Map results of current projects to find the common ground
- Conduct practical research into archived material over a long period of time to understand how decisions about significant properties have been made in practice, explicitly or implicitly.

Conclusion

There is no doubt that the significant properties workshop generated a high level of interest. This is not only reflected by the large number of participants and their diverse backgrounds, but also by the discussion which continued after the workshop via various blog entries. The workshop successfully demonstrated the breadth and depth of significant properties research to date, and pointed to a number of future directions to move things forward. It is important to realise that some of the work presented is ongoing, such as the InSPECT and Planets projects. Full account should be taken of the results emerging from these projects.

The relationship and difference between significant properties and representation information were touched upon by a number of speakers but remained somewhat unclear. It was obvious that even the speakers were at some variance in their understanding of these terms. This is a topic that clearly warrants further discussion and clarification. Both types of information are crucial to successful preservation. Clear understanding about how they relate and map to each other might provide us with the common ground between the technical and curatorial aspects of digital preservation.

References

- Becker, C., Rauber, A., Heydegger V., Schnasse, J., & Thaller, M. (2008). A generic XML language for characterising objects to support digital preservation. In *Proceedings of the ACM Symposium on Applied Computing (SAC'08), Track 'Document Engineering'*. Fortaleza, Brazil, March 16-20, 2008. Retrieved 17 April 2008, from http://www.ifs.tuwien.ac.at/~becker/pubs/becker_sac08.pdf.
- Brown, A. (2008) *White paper: Representation information registries*. Retrieved 24 April 2008, from http://www.planets-project.eu/docs/reports/Planets_PC3-D7_RepInformationRegistries.pdf
- Digital Preservation Coalition (2008). Presentations and photographs for JISC/BL/DPC workshop on significant properties of digital objects. Retrieved 28 April 2008 from <http://www.dpconline.org/graphics/events/080407workshop.html>
- Ding, L., Ball, A., Matthews, J., McMahon, C., & Patel, M. (2007). *Product representation in lightweight formats for product lifecycle management (PLM)*. Retrieved 29 April 2008 from http://www.ukoln.ac.uk/projects/grand-challenge/papers/DET2007_Ding_Paper.pdf
- Hedstrom, M., & Lee, C.A. (2002) *Significant properties of digital objects: definitions, applications, implications*. Retrieved 28 April 2008 from http://www.ils.unc.edu/callee/sigprops_dlm2002.pdf