THE IS&T ARCHIVING 2008 CONFERENCE

June 24 – 27, 2008

Bern, Switzerland

This was the first time that the IS&T Archiving conference was held outside of N. America. The venue for this meeting was the University of Bern in Switzerland and around 175 delegates attended the event, excellent for a conference new to Europe.

The quality of the papers submitted was also very high and there was a good mix of oral and interactive papers. Probably due to the European venue the conference attracted presenters and delegates from the Far East as well as the strong US and European groups.

The conference followed a tried and tested IS&T format that also translated well into a European setting. A preliminary day was devoted to Short Courses, providing introductions and in depth studies of topics of interest to the target audience for this conference – see Section 5. The conference papers themselves were presented in 3 forms.

1. Focal papers. These are longer presentations given orally to the full conference. These longer presentations are allocated to invited papers or proffered papers that are ranked by the reviewers as being on topics of special interest to the conference.

2. Standard length oral papers. These form the majority of the oral papers presented to the full conference.

3. Interactive papers. These papers are presented as a poster in a dedicated session of the conference with a short oral preview during the relevant conference session.

Finally, this year's conference also featured a strong resurgence of interest in microfilm solutions for data storage.

1   SHORT COURSE DAY 24TH JUNE

We had a full short course programme at this conference with around 100 places taken up spread over 12 very different courses. With around 175 delegates registered for the conference this seemed to be a fairly good tutorial turnout.

The courses were organised into 5 tracks as listed below.

1.1   Case studies

This was the most attended of the 5 tracks.

The track started with a 2 hour course given by Simon Tanner of King's College London, entitled "Making digital preservation affordable: values and business models". Simon discussed the strategic perspectives towards being able to effectively finance digital preservation. The audience and other stakeholders define the economic factors by which digital information is valued, used and ultimately retained. In looking to finance digital preservation there are a number of different issues to consider including business planning, risk management, possible revenue streams and a clear cost benefit relationship. This tutorial

offered a means of developing a cost and benefit justification for digital preservation to help secure the financial underpinning needed to make institutional digital preservation a realistic proposition. Simon is well known in this community and was Program Chair for this conference so his tutorial was well attended with a broad cross section of delegates.

The second 2 hour course was given by Richard Wright from the BBC, entitled "How to Save Audiovisual Archive Content by Digitisation – and then How to Save the Digits (in archives great and small)".This tutorial reviewed audiovisual preservation / digitisation projects across Europe, beginning with major broadcast archives but also including the requirements and general possibilities for smaller collections. A particular issue is the many audiovisual collections which are part of institutions which have little or no audiovisual technical expertise. However, IT and Internet technology has come to the rescue and audiovisual archives are at the start of a revolution: content hitherto inaccessible can now be put on YouTube. Digital storage costs are 99% cheaper than 15 years ago. There is a groundswell of support for the public domain leading to market pull for the availability of older content. Finally, the future roadmap for this material was presented.

The 3rd 2 hour course was jointly given by 3 people, again from the UK. Stephen Grace (Centre for e-Research), Neil Grindley (Joint Information Systems Committee), Grant Young (The Technical Advisory Service for Images) presented "Significant Properties and their Role in Digital Preservation". To enable digital images and other objects to remain both accessible and meaningful over time, it is critical to understand what aspects of them need to be preserved. Recent work in this area has looked at this issue using the concept of 'significant properties', and has looked at a variety of digital object types, including raster images, vector images, moving images, structured text, audio and software. To take an example, three significant properties of a vector graphic might include, line width, opacity and colour, the primitives and associated attributes of which may be described differently by different graphics systems such as PDF/A (see Section 1.4). An organisation with curatorial responsibility for digital objects cannot assert or demonstrate the continued authenticity of those objects over time, or across transformation processes, unless it can identify, measure, and declare the specific properties on which that authenticity depends. Nor can it undertake the preservation actions required to maintain access to those objects, unless it can characterise their current technical representations with sufficient detail. Confidence at the object level is also informed by the trust placed in the organisation curating the files. This tutorial looked at the nature and extent of recent work in the area of 'significant properties' and gave a more detailed view of the ongoing work being carried out by the InSPECT Project (Investigating the Significant Properties of Electronic Content Over Time). It then considered the wider potential impact of this work on creators and users of digital images.

The final case study offered was from Dr. Krystyna W. Ohnesorge and Dr. Hartwig Thomas of the Swiss Federal Archives (SFA). This 2 hour presentation was entitled "SIARD - a file format for archival of relational databases". Often central information about today's administrative activities is stored in databases. Thus it is an important task to archive database content. The SFA - and other archives - are obliged to archive this type of administrative "documents" in a

way that ensures accessibility of database content for at least fifty years. Unfortunately no standardized format exists for archiving database content over several decades. Although the concept of relational databases in the 1970s was based on the assumption that data have a longer life expectancy than software or hardware, only the query language SQL has been weakly standardized. Oracle table spaces, MS Access MDB-files and most other database file formats are proprietary and therefore not suited for archival use. The SFA have developed a storage format for relational databases and the software SIARD (software-independent archival of relational databases) for handling database content archived in this format. The SIARD format is an open format based on the standards SQL:1999 and XML. Therefore it is an appropriate format for archival of database content. A first prototype of the SIARD software was used by the SFA for productive archival of relational databases of government agencies. The development of a second version of SIARD, the "usability release", was started in 2007 and will be productive in the second quarter of 2008. The SIARD format is an open and published format of the SFA. The SIARD format contains all the base tables of the database as well as the description of the structure of the database, its meta data, in XML files which are contained in a single ZIP file per database. The database metadata describing schemas, tables, columns and other database content are compatible with the SQL:1999 standard. The structure of the SIARD XML files is published by the SFA as XML schema definitions and an explanatory description. Conformity to the SIARD format can thus be automatically verified. The SIARD software permits examining database data as well as loading archived data into a commercial database system.

1.2 <u>Technology</u>

Sabine Süsstrunk from the Ecole Polytechnique Fédérale de Lausanne (EPFL) gave a 4 hour course entitled "Color Image Workflows and Architecture for Archiving Applications". The attendees seemed to be a good mix of beginners and those who are already users. This course covered workflow from image capture to visualization to archiving and discussed the appropriate image parameters for each step. It was emphasised that images optimized for archiving, viewing, and printing usually do not contain the same digital values, The reasons for this were explained and the necessary adjustments to the image state (colour encoding, resolution, compression, processing, and rendering) were explained. Sabine emphasised the issue of gamut colours, using gold leaf as an example of one that is out of display and printer gamuts.

The second course was a 2 hour event given by Hans I. Bjelkhagen of the Centre for Modern Optics, North Wales on the subject of "Colour display holography". Although holography has been around since the mid-1960s the interest in using holography to display objects has been rather limited mainly because of the monochrome image, often green or orange only. Recent development in improved techniques and recording materials, most important being the possibility to record 3D images in full colour has caused an increase interest in display holograms. In theory colour holographic imaging technology can be applied to many potential applications and markets of interest to this community such as the display of expensive or unique art pieces, museum artefacts, advertising, etc. This tutorial made a number of claims about current colour holography technology.

- The virtual colour image behind a holographic plate represents the most realistic-looking image of an object that can be recorded today.

- The extensive field of view adds to the illusion of beholding a real object rather than an image of it.

- By choosing the optimum recording laser wavelengths within the spectrum, good colour rendering can be achieved. This was also the subject of a paper in the main body of the conference – see Section 5.

- Colour holography may become an important reproduction technique for 2-D objects such as oil paintings. Holographic reproductions provide extremely realistic-looking images, showing the texture details such as brush strokes and the painter's signature.

- The holograms will not fade or change colour even if they are continuously on display.

Alan Hodgson also presented a 2 hour version of a course on "Image Science for the Archiving Community". This describes Image Science as a broad topic covering the parameters that can be used to describe an image. This course gives a concise overview of what Image Science can offer the Archiving community and is equally applicable to images in the digital or analogue domain. It is therefore pertinent to any hard copy technology such as paint, traditional photo or new recording media such as inkjet as well as digital images. It covers issues such as image content, noise, tonal range and digitization as a brief overview. It covered the image science considerations of digitization by whatever means, display and any subsequent printing and included case studies from printed and digitized images. The delegates came from a broad spectrum of the Archiving community with no previous experience but plenty of interest in the topic. The section on multi-spectral imaging was particularly of interest to this group.

1.3 **Hard Copy Issues**

The first course on this track was given by Franziska Frey, Rochester Institute of Technology and Martin Jürgens, a European conservator and was a 4 hour lesson on "Contemporary Photography: Digital Prints". The course focussed on the materials, identification, and stability of digital prints used by contemporary artists. It provided attendees with the knowledge and tools to handle the issues surrounding the acquisition and preservation of prints made from digital files, as well as an understanding for the trends in imaging technology and artists' use of modern photographic printing techniques. From sitting in on part of the course it was obvious that there is still a good deal of interest in hard copy image permanence.

The second course in this track was due to be given by Stephanie Ogeneski of the National Anthropological Archives, Smithsonian Institution in the US entitled "Digitizing Historical Negative Collections". Unfortunately this course had to be cancelled as the presenter became unavailable. I attended this course last year in Washington where it was well subscribed and provided a thorough introduction to those working with historical negative collections, giving a fundamental understanding of the relationship between analogue and digital images. Participants involved in digitization projects of negative collections

explored best approaches when embarking on a digitization project with these materials to ensure accuracy in reproduction. It gave an overview of the historical material and how that material is conceived, captured and translated from analogue to digital, including the issues of digital guidelines and tone reproduction.

1.4 <u>Preservation Formats</u>

The subject of this track of the short course programme was computer file formats. The best formats to choose for long term preservation of digital files has been the subject of much debate over previous Archiving conferences. The 2 courses in this track were chosen to reflect the 2 formats that are emerging as the chosen ones by this community – PDF/A and JPEG2000. As a result this was a well attended track.

The first 4 hour course was given by Leonard Rosenthol of Adobe Systems entitled "A Look Inside the Portable Document Format (PDF)". It provided a peek inside the various aspects of PDF including such areas as fonts, graphics, colour, interactive elements and security. Special emphasis was given to the use of PDF for "long-term archival storage of digital documents," as spelled out in ISO 19005 (PDF/A-1).

The second 4 hour course was given by Robert Buckley of Xerox entitled "JPEG 2000 and other formats for image preservation". It began with an introduction to commonly used digital image formats and image compression methods, laying the groundwork for discussing and comparing the formats for image preservation, with an emphasis on TIFF and JPEG 2000. TIFF is typically used to store uncompressed images. JPEG 2000 is a still image compression standard based on wavelet technology.   JPEG 2000 is attractive for image preservation and access because it can handle a wide range of applications, including gigabyte and high-dynamic range images, spectral imaging, digital cinema and on-line image collections. This course explained the key parts of the JPEG2000 standard, covering its features, demonstrating its capabilities, and discussing who's using it and why.

1.5 <u>Cameras and Scanners</u>

Unlike the other tracks, Cameras and Scanners consisted of 2x4 hour presentations from the same presenters, Peter Burns of Carestream Health and Don Williams, a consultant.

The first half of the day dealt with the topic "Evaluating Digital Scanner & Camera Imaging Performance". As many of today's standards for characterizing imaging performance are based on image science principles the session began by introducing this perspective and then describe its application to scanner and digital camera performance in an archiving environment. There are standards and accompanying tools to help the user control tone reproduction and evaluate manufacturers' claims of resolution, dynamic range, and noise and these were described. They then went on to identify several common image artefacts associated with digital image capture and through examples described how performance parameters can be monitored by summary measures acquired automatically as part of a quality assurance process.

The second half consisted of a workshop entitled "Scanner & Camera Imaging Performance Workshop". This workshop, new for 2008, aimed at those interested in the practical application of imaging performance evaluation and control for collections. Following a brief review of current practice and standard methods they addressed several common problems faced by those providing imaging services, or seeking to improve image content. In each of the cases addressed they discussed the selection and development of test plans, performance measurements and simple analysis. Attendees had the opportunity to perform the evaluations using analysis software provided, illustrating the uses and limitations for the methods described.

2 FILE STORAGE

It is important to consider the issue of computer file formats as this is a critical factor in the cost models used to decide on storage solutions. As such it is a major component in the financial viability of microfilm solutions for various applications.

Uncompressed TIFF is a widely accepted standard for the storage of master images in digitisation projects. However, it looks from the programme that JPEG2000 has finally come of age in this community. This format has a great advantage over uncompressed TIFF of a much reduced file size. There is also a tendency for PDF/A (the Archiving "flavour" of PDF) to be used, particularly as this is now fixed as an ISO standard.

The issue of file formats was covered in great detail in one section of the short course programme – see Section 1.4.

3 MICROFILM TECHNOLOGY

There appears to be a genuine resurgence of interest in microfilm technology within this community. The realisation of the resource and cost implications of data migration has prompted an increase in interest and subsequent awareness of high density storage media with a human readable optical format and a high inherent stability in dark storage. Microfilm technology appears to fit this bill.

Microfilm as a long term storage medium has a number of attributes.

1. The data is human readable. As long as there is a human eyeball and a magnifying lens the data can be accessed. This concept is becoming known as "Images on film" in this community. An alternative concept, "Bits on film" allows this technology to store digital images in what are essentially 2D bar codes.

2. There is fear within this community that data will not survive the migration cycles that will be necessary for the long term preservation of digital images. The long perceived lifetime of microfilm diminishes this issue.

3. The data is held in a compact form on a medium that has a track record of long term stability.

4. The recorded image is in read only form so is immune to virus attack or manipulation.

The issue of colour is a major one in this community. Much of the content has a colour component and as a result a lot of interest is generated by the potential of the colour microfilm products. However, this is not the only solution. The use of monochrome films to store colour separation images is not new – it has been used for many years in the motion picture industry.

Colour separation also has potential extension into multispectral imaging. There is substantial interest in multispectral imaging within this community as a way of accurately characterising the spectral reflectivity of an object and hence it's perceived colour under a variety of illumination spectra. One indication of the usefulness of this technique came from a Finnish university which showed how previously unperceived detail can be revealed using such systems.

3.1 Microfilm scanning

In some cases there will be a requirement to scan microfilm to create a digital file. It has been shown that microfilm is a good storage medium for scanning. A further paper at this conference looked at the success rate of OCR of different contrast monochrome microfilm recordings of newsprint.

3.2 Evaluation issues

This conference had a paper that illustrated the type of image physics likely to be important in this market, looking at modulation transfer and point spread functions plus granularity. Point spread function in particular will be important when considering the digital bit density capability of microfilm.

One other aspect that seems likely to re-emerge in this process is the potential trade off between resolution and sharpness. The classic demonstration of this is based on television images but the same image physics was shown be shown to apply to digital images.

4 COST MODELS

Migration is necessary as computer hardware and software systems become obsolete. For example, how easy is it to read a 5¼" floppy disk containing DisplayWrite and WordStar documents? As a result various tools are being developed for migration strategies.

There are various cost models being developed and some were presented at this meeting. It should be noted that these cost models are critically dependent on the migration cycle time which could be as short as 5-10 years.

5 HOLOGRAPHY

In addition to the short course on colour display holography there was a paper on this topic from Hans Bjelkhagen and Peter Crosby. The paper was presented as a poster at the conference.

One problem with colour display holography is the accuracy of the colour rendition. This paper seeks to address this, presenting work on colour accuracy from 3-7 laser wavelength exposure. Four wavelengths seem to give a good balance between complexity and colour accuracy.