

IIC 2012 WEBSITE AND STUDENT PAPERS MEETING

Chelsea College of Art, 4–5 April 2012

REPORT

The aim of the meeting was to investigate the number and type of student conservation theses (at MA and PhD level) available for consultation, in what form, how they were held and how they could be accessed. This literature is not well-known and IIC could perform a useful service to the conservation community by providing information on or links to theses and dissertations through its website. Representatives from a range of international conservation training institutions and thesis repositories took part in a two-day meeting and the present report consists of abbreviated minutes of the proceedings taken during the sessions. (Note: A transcript could not be made; it was hoped that a recording of the proceedings would be available, but unfortunately it was of insufficiently good quality.)

Participants registered

David Garcia, Dean of Camberwell, Chelsea and Wimbledon Colleges of Art, UK (welcome)

Tatiana Ausema, Coremans Fellow, Preservation Studies Doctoral Program, Department of Art Conservation, University of Delaware, USA (on line)

Karen te Brake-Baldock, INCCA coordinator, Cultural Heritage Agency of the Netherlands (RCE), Amersfoort/Amsterdam, The Netherlands

Jocelyn Cuming, MA Conservation Pathway Leader, University of the Arts, London, UK

Sarah Gould, EThOS, British Library, London, UK (on line)

Maria Gruber, University of Applied Arts, Vienna, Austria

Anna Hoffmann, University of the Arts, London, UK

Dr Panagiota Manti, Department of Archaeology and Conservation, School of History, Archaeology and Religion (SHARE), Cardiff University, Wales, UK

Dr Jennifer Mass, Senior Scientist and Director, Scientific Research and Analysis Laboratory, Winterthur Museum, Garden and Library, Delaware, USA (on line)

Stephanie Meece, University of the Arts Research Online, London, UK

Gustavo Grandal Montero, Librarian, Conservation collection, University of the Arts, London (UAL), UK

Dr Caitlin O'Grady, Department of Art Conservation, University of Delaware, USA (on line)

Katalin Orosz, Hungarian National Museum, Budapest, Hungary

Renata Peters, Lecturer in Conservation, Institute of Archaeology, University College London, UK (on line)

Edward Pinsent, Digital Archivist and Project Manager, Digital Archives and Repositories, University of London Computer Centre (ULCC), UK

Heather Ravenberg, Postgraduate Student, University of the Arts, London, UK

Mark Sandy, MA Conservation Pathway Leader, University of the Arts, London, UK

Meagen Smith, Postgraduate Student, University of the Arts, London, UK

Dr Athanasios Velios, Research Fellow, Ligatus Research Centre, University of the Arts, London, UK

Dr Angela Weyer, Hornemann Institut, Hochschule für angewandte Wissenschaft und Kunst (HAWK), Hildesheim, Germany

Alberto Campagnolo, Postgraduate Student, University of the Arts, London, UK (moderating on-line participants)
Jo Kirby Atkinson, IIC
Graham Voce, IIC (Chair)

[Together with additional attendees over the two days, in person and on line.]

IIC would like to express its gratitude to the Getty Foundation for most generously funding the event, to the University of the Arts London for hosting the event at its Chelsea College of Arts site and to all those taking part for contributing to an interesting and worthwhile event which provides the foundation for IIC to take the project forward.

Wednesday 4 April 2012

Introduction and welcome by **David Garcia**, Dean of Camberwell, Chelsea and Wimbledon Colleges of Art

He spoke on:

- The need for archiving and retrieval of dissertations: 220 PhD students from UAL – will the dissertations be accessed or read?
- The format: Marvin Minsky, on mag[netic] tapes is inaccessible; Galileo, on paper, is readable. Hybridity is necessary. *News in Conservation* (note – we had given him some) in printed form is important, while remembering that the semantic web is important – it flattens hierarchies.
- Insights from people unknown – the importance of the semantic web.
- The importance of taxonomies and of sharing.
- Agreements and opting in / out.
- The challenge and opportunities of the semantic web.
- An archive is transforming into a database.
- The opportunities of shared taxonomies.



Graham Voce, IIC, introduced the meeting by explaining why it is taking place and what IIC hopes to do by providing a portal for access to conservation theses, a form of grey literature that is not as well known as it might be and could be better used.

The meeting had been preceded by an on-line survey which training institutions, museums and other institutions involved in education had been invited to complete (a summary of the results was made available to the participants at the meeting on the second day, Thursday 5 April 2012). Essentially this showed that theses, dissertations and reports available in training institutions ranged from unbound typescripts on a shelf to bound theses in book form, theses on floppy discs or CDs converted to Pdfs, theses on microfiches and theses on line, with abstracts, fully searchable through metadata and every combination of these. The Hornemann Institut, for example (see below), holds over 1400 theses [2012 figure] in electronic form on its website, searchable worldwide through metadata provided by the students. On the other hand, many institutions hold much of their material on paper, ranging in number from a few

tens of reports (for example, the Hamilton Kerr Institute, University of Cambridge, UK, and the Department of Archaeology, Durham University, UK) to two or three thousand (Escuela Nacional de Conservación, Restauración y Museografía, México). Some (for example the University of Applied Sciences, Cologne) hold several hundred reports in each format. The provision of search terms – metadata – was very variable. Theses might be catalogued through a conventional library system, so were searchable within that system; some were simply listed so searching was very limited. In some cases, the system could be searched through the worldwide web; more frequently searching was possible within the institution itself, but only to a limited extent outside, or not at all. Copyright was usually held by the student, or by the institution, or jointly. [Note: Summary of the position in April 2012]

The institutions supplying information included:

Art Conservation Department, University of Delaware

Centre for Cultural Materials Conservation, University of Melbourne

Centre for Textile Conservation and Technical art History, Institute of Art History, University of Glasgow

Conservation Centre of the Institute of Fine Arts, New York University

Department of Archaeology, Durham University

Department of Archaeology and Conservation, Cardiff University

Escuela Nacional de Conservación, Restauración y Museografía, México

Hamilton Kerr Institute, University of Cambridge

Hornemann Institut, HAWK, Hildesheim

Hungarian University of Fine Arts, Budapest

Institute of Archaeology, University of London

Instituto Superior de Conservación y Restauración Yachay Wasi, Lima

Scuola universitaria professionale della Svizzera italiana (SUPSI), Canobbio

Shelburne Museum, Vermont

Tate, London

University of Applied Sciences, Cologne

Winterthur/ University of Delaware Program in Art Conservation, Preservation Studies Program (WUDPAC/PSP), University of Delaware

Gustavo Grandal Montero, Librarian, Conservation collection, University of the Arts, London (UAL), gave a description of the UAL situation:

PhD theses in print are catalogued by Library of Congress headings (an internationally recognised standard). There is an online institutional repository and *recent* PhD theses are probably available on this in full text; probably not 100% at present.

A backlog of MA dissertations has been received, they are cataloguing now about 100, going back to 1990. Should these be made available through the repository? Is this research level material? The repository is very flexible in terms of what it can accommodate (including films, etc.).

It is the decision of the library and the researcher as to how accessible an item is, partly because of copyright. Images within theses can lead to copyright problems, particularly if the theses are then made available digitally.

EThOS (UK, Electronic Theses online System) – British Library – involved in the discussion (see later presentation by Sarah Gould, EThOS).

Graham Voce: What is the function of the repository: to make available or to store items?

Gustavo Grandal Montero: Legal restraints are such that you have to work within them, but the aim is to make the theses available.

Searching: by title, otherwise basic Library of Congress subject headings are used.

Panagiota Manti: At Cardiff people upload their own data.

Gustavo Grandal Montero: Users can use their own terminology and the thesaurus the database supplies, but the end user has to describe the thesis content. I (Gustavo) am not a conservation subject specialist.

Stephanie Meece, University of the Arts Research Online, UAL clarifies:

1. All PhD theses submitted since 2010 are included in UAL Research Online, and about 90% are downloadable in full; the remaining 10% are restricted access (mostly because they contain material copyrighted to other authors) but can be requested from UAL staff individually. Older theses are being digitised by the British Library for us, under the Ethos project, as they are requested by BL users (see **Sarah Gould's** presentation).
2. UAL Research Online does not accept Master's level theses at present, and our advisory board does not intend to change this in future.
3. The function of the repository is both to curate a collection of the University's research output, and to make that output more easily accessible, where copyright restrictions permit.
4. In practice, PhD theses and metadata are uploaded by UALRO staff. There is a standard subject list, which includes 'Conservation'. General discussion on the minimum standard for uploaded material, quality and quality control. The variability between different countries in what was acceptable in terms of quality of MA-level theses in general was recognised to be a problem.

Tatiana Ausema (on-line), in answer to a question on many theses there would be in the US from all the programmes: Very few (quite surprising to those in the room). The Delaware MA level thesis is about 25 pages, a description of the conservation treatment of a single item.

Angela Weyer: German MA courses and theses take five years.

Panagiota Manti: The Cardiff (and presumably Welsh Universities) network is rather like those in England; repositories are self-archiving and harvested, searchable. The Cardiff one includes MA theses with distinctions (so obtaining a mark of at least 70%) – probably about 10% of all the MA theses in a year – and PhD theses.

[General discussion on how easy it is to harvest the contents of repositories.]

Four other systems were then described.

Angela Weyer, Hornemann Institute, Hildesheim, Hochschule für angewandte Wissenschaft und Kunst (HAWK), Hildesheim, demonstrated the **Hornemann Institute** site:

The Hornemann Institute is responsible for E-publication; e-learning courses; conferences.

The site holds 1,430 academic theses; conference papers abstracts and texts; articles; projects; posters and videos – thousands of new documentation records.

It is possible to do: a full text search; sorting; to see the holdings of other academic institutions they hold on their site.

They provide a platform for restorers to publish fast, while maintaining copyright. The restorers can change or withdraw their texts, or anything else (after the professors have made corrections in the case of the student texts). The newsletter of the institute lists the new theses

published and advertises for them. Technical support is provided and there is a back-up system. Each student has an account and has control of the data.

The sort of details given for a thesis: abstract (this is compulsory, in the original language and in translation); table of contents; details; contact details; download details.

(An example was shown of one from the Hochschule für Künste, Bern; a Diplomarbeit; also various others.)

Because the author (the student) holds the copyright he can publish as he wants.

The Hornemann site acts as the data centre for the Technical University of Hannover. The site gives a DOI – digital object identifier – to each item uploaded onto the system; the address is thus fixed forever. This is done with the library, in cataloguing. But then, of course, the content is fixed and the author (the student) cannot change anything.

Problems:

- to motivate authors over the years and to make it interesting for people again and again – for new students, staff, etc.;
- Technical problems – technical development;
- User problems: No editing – and this is a problem, leading to quality questions.

NB Some authors do not know about copyrights: even though this is taught at the Hornemann Institute, it is not necessarily taught everywhere else. The Institute does not know in practice that the authors have all the permissions, although there has not been a problem so far.

Universities compete for the best students and publish for their own reputation. Some professors in conservation and restoration are very close to the students and put a lot of work into the thesis and as a result the students are not sure if they have the rights over their own theses.

[Discussion]

Karen te Brake-Baldock, INCCA coordinator, Cultural Heritage Agency of the Netherlands RCE), Amersfoort/Amsterdam, described **INCCA** (PowerPoint):

INCCA is an activity of the Netherlands, hosted at the RCE (Rijksdienst voor het Cultureel Erfgoed) Amsterdam, and is one of shared activities and artists. It is a volunteer network, with national groups; the North American one is organised into a Foundation*, but the others are just groups. The network has a website, <http://www.incca.org/>, that is open and contains a lot of information. It has various projects and links to other organisations.

[* In March 2015 INCCA-North America changed its name to Voices in Contemporary Art (VoCA), <http://www.voca.network/>.]

Resources include links to and full texts of PhD theses, if the author is willing to share and give open access to all.

There is also a database, similar to the Hornemann Institute one, used and filled by INCCA members which allows members to access each other's unpublished information, such as artist interviews. Here one can search the database, create one's own records for the database, etc. It is also possible to look for fellow members, similar to LinkedIn. The metadata are based on Dublin Core: fairly general fields such as artist, title, artwork type – and here title means title of the *document*, not the artwork. Another heading, creator: this is the person creating the document., not the artist; contributor: this could be anyone, including the artist; keywords; etc. Publisher: the documents mentioned are generally not 'published', but this term is used as it is useful for libraries.

The keywords are not controlled and are not cleaned up as they are user-generated.

Resource, describing language: this is English so the keywords are in English (although the document itself need not be).

The INCCA database is passworded so is pretty secure, but the record can be edited.

There are guidelines on making a document for the resource, including getting permissions. There is a policy on re-use, using Creative Commons licences.

The metadata fields are described – there is an (i) (= information) label beside them – and only a few are compulsory.

The record, or a draft, etc., can be saved.

It seems, though, that many people are now happy to post information on the external database; they are not so worried about keeping their information private now.

Membership is based on a willingness to share and is supposed to be based on the member supplying five records a year, but this is not happening. The real strength now is the platform itself.

It is important to remember that there may be damaging things in the report (damaging to the museum concerned).

Copyright of images: again, it is assumed that authors have worked this sort of thing out.

Artists should be told that the report is on-line – ‘published’.

The reports are not refereed as a paper would be; it is assumed that if a museum conservator is responsible or involved the quality control is built in. If the report is by a private conservator then an endorsement is needed.

There are about 700 members; they also use LinkedIn to spread news and here there are about 1,000, but there is some overlap.

If conditions of upload are to be changed, the member has to give notice of about 3 months or so.

Katalin Orosz, Hungarian National Museum, Budapest, described the training of conservation students in **Hungary** and the theses produced (PowerPoint):

Courses. Hungarian University of Fine Arts:

- a) Fine arts objects conservation: painting, divided into easel and mural; wooden sculpture; stone sculpture. The MA course takes five years, three years basic training, then a real object to work on. The students have to restore one object (or two – this depends on the object) and write a diploma work consisting of detailed documentation of the object and a thesis about the investigation, the treatment, new materials used and so on.
- b) Applied arts objects conservation. There was an independent institute for this founded in 1974; this was in the museum field and the museum developed the course. This lasted four years. In 1991 the course changed to a five-year MA in cooperation with the University of Fine Arts, then the institute became connected to the University of Fine Arts and the National Museums. However, this course is part time: one week per month of lectures and practice; formerly the aim was to give training and the students worked the rest of the time. The course covers wooden objects and furniture; metal; siliceous materials; paper and leather; textiles and leather. Nowadays the students are very young, from secondary school, and they do not have jobs, but the training is still part time and it cannot be changed for financial reasons.

The first diploma – five years – is free of charge.

For both branches of the training there is a maximum of 12 students per year. In the applied arts branch there are five specialisations and the university starts three per year. 10 students do not have to pay and two do. The fees have recently gone up to £1,300 a semester (2 semesters a year) which is very high for them.

The first training was brought in by the late Agnes Timár-Balázs [the internationally well-known textile conservator], who in 1996 wanted to disseminate knowledge of the training by means of an exhibition; the students treated original objects (from other collections or not) and there were a lot. In 1999 this exhibition moved to the National Museum and is now

annual. In 2002 a small catalogue was produced (containing for each object before and after treatment photos, the names of the student and the supervisor, object details, etc.) and the fine arts conservation students were also incorporated. Now the catalogue has English abstracts, before and after photos, etc. The students have to give a small paper and the show is open to the public; conservation colleagues come and the objects are exhibited. The photos and texts are controlled, posters made and so on. The explanations are suitable for the general public. A DVD presentation with the text is also displayed on a monitor throughout the exhibition. They receive financial support from the National Museum and the Hungarian National Cultural Foundation so they receive money to do all this. Last year [2011], however, they could only get support if they made a travelling exhibition, so they did this and this exhibition went throughout the country.

In order to be suitable as a diploma work the object must be quite complex so there is a diploma committee to assess the objects. If the object is accepted, the National Museum makes a contract with the owner of the object for the conservation work to be done.

Conservation is free of charge for the institute, but any investigation carried out has to be paid for, so they can teach the student how to cooperate with scientists and investigate the object: they do not have the facilities to do this themselves. The exhibition work and design is done by the museum; the installation is partly by them, partly by the students and teachers. The Head of Applied Conservation training edits the catalogue.

Distribution: The catalogue is not distributed electronically because of the need to obtain copyright permissions for photos of the objects; this is permitted for the printed catalogue and the posters, but not electronically.

In the last six years, students have had to give their texts and illustrations on CD and this is catalogued in the library. It depends on the owner of the object and the copyright problem as to what happens next. The catalogues are searchable, but all in Hungarian. 600 theses in 40 years in the two branches (Fine Arts and Applied).

[5 April 2012]. Having asked her Head of Department what was available on-line, Katalin adds that the library catalogues are available on line, in Hungarian but are searchable in English to some extent (she does not know how this works) The Institute library, the University library and the National Museum library are all connected and the National Museum e-publications are in the system. The catalogues are on line and documents can be archived this way, although the earlier director of the National Museum did not want to do this. However the abstracts of the best diploma works are put into the university cataloguing system, so it is technically possible.

Panagiota Manti, Department of Archaeology and Conservation, SHARE, Cardiff, described the situation for conservation theses there (PowerPoint):

The courses are: BSc Conservation of objects in museums and archaeology; BSc conservation courses

MSc Professional conservation

MSc Conservation practice

MSc Care of collections

MPhil Conservation

PhD Conservation

At present they have details about student work as an Excel file and hard copies in the conservation department laboratories, also a library (very small). Even so, it is informative and they might want to share it as an archive. At the moment they use YouTube but later they might not want to do this.

She posed the question: If we share theses, do we also share conservation work and archives? Do we share project notebooks? Valuable information? Again, there is the question of quality control.

Online theses: there are only a few at present and only PhD theses, not Masters: see ORCA – Online Research at Cardiff.

Research publications from Cardiff University: After a thesis is accepted and passed, for the student to be awarded the degree the thesis has to be uploaded into ORCA. The student can put a bar on access – if, for example, the student wants to publish it.

[There was immediate discussion of this point: for example:

Stephanie Meece from the University of the Arts repository: Making the student put the thesis on the website in order to get the degree is not legal.

This led to further discussion.]

Panagiota Manti: The aim [of uploading the theses] is to make already published research work available and to come up in Google Scholar when someone puts an appropriate search term in.

The search terms include text, year, creator, abstract, date, keywords, subjects, type of material, etc. – with tick boxes.

They use some free software from the University of Southampton – EPrints 3.

ORCA is self-archiving and an open access repository: it is not intended to be a substitute for peer-reviewed journals. At present PhD theses only are listed on ORCA, not MA theses.

PhD theses – access via ORCA is mandatory. It is mandatory also for research contracts from AHRC [the Arts and Humanities Research Council, a UK Research Council providing government money supporting research and post-graduate study in the arts and humanities] and the Science and Heritage Programme [funded by AHRC and EPSRC, the UK Engineering and Physical Sciences Research Council]: it is a requirement to deposit copy of article. (The deposit of appropriate metadata was also mentioned at this point.)

Access to the theses has been developed with the Welsh Repository Network, WRN; e-thesis harvesting with the National Library of Wales and the British Library EThOS project (see below for this). There are at present about 50,000 theses; all PhD and Masters theses in Wales have to be excellent and the Masters theses are those taught Masters dissertations which have a Welsh interest or have gained a distinction.

There are then various considerations with access to theses, including the type of material: is the thesis in the form of hard copy? digital? is it downloadable or not?

Dr Manti then took this further to consider how an e-thesis collection for IIC might work and this led on to a brief general discussion of the points raised. These may be summarised as:

- The need for examples for practice for establishment of rules
- The possible advantages of such a collection
- Issues of moderation of the theses
- Uploading of the work
- Concerns: Quality control – how can this be monitored and maintained by IIC or on its behalf?
- Would this include an archive of past projects? This was a matter of general concern as other people present had had to face similar problems; should this be via alumni self-archiving for example? Do the students have a log-in code to do this – even once they've left? (What about a self-archiving repository? Does IIC need this?)

Graham Voce: IIC's role in this would be to provide the expertise in searching; the conservation terms to use to access the databases, be they ORCA, Hildesheim or wherever.



Thursday 5 April 2012

Various extra people present (including Anna Hoffmann, MA student, not on the typed list)

Edward Pinsent, Digital Archivist and Project Manager, Digital Archives and Repositories, **University of London Computer Centre, ULCC** (site visible on screen):

Edward Pinsent started his career as an archivist, working with paper records. He joined ULCC in 2004 to work on the National Digital Archive of Datasets.

The purpose: preserving datasets so they can be used on line; preserving digital objects and rendering it possible to give people access.

Not everything held is digital. Access is through an on-line point.

Athanasios Velios, Ligatus Research Centre, UAL: The repository is international.

Stephanie Meece: The UAL digital repository software is EPrints modified.

Edward Pinsent: This is a good one to choose.

The idea of the digital repository started in the USA. Universities see repositories as a good method to allow managed access to and dissemination of material on line through a website. Repositories emerged in an academic environment so the dissemination tended to match the work flow of the university.

'Publishing' is sharing the output of research.

Embargoed access can be managed, also intellectual property rights, copyright, etc. The system can handle different ways of coding, such as PDF, Word, Excel spreadsheets, etc. It can manage a wide range of formats: it is strong on text and images, but it can also manage sound, movies, etc. To a degree these can be replayed over the front end.

Panagiota Manti: How is preservation of the file format once digitised handled?

Edward Pinsent: EPrints is developing add-ons, plug-ins and so on to make this easier. (He will come back to this: see below.)

a) Metadata:

- These are chiefly descriptive information, the more the better and the repository people and the system offer strong support. EPrints, Dublin Core: author; title; data, etc. Or finer detail is possible if required. Some existing metadata schema are already installed on EPrints.
- How does the system work? There are registered users, such as students, professors and so on. They are responsible for upload of the documents, self-archiving as they do so. They *must* fill in some metadata; it varies as to how much. Thus the job of cataloguing the item is in effect passed on to the user [Note: It would be interesting to know how the terms are chosen]. Other metadata schemes [other than that installed on EPrints] can be plugged in: someone who wanted a very specific resource, e.g. Library of Congress, suggested keywords. These were then easily uploaded into the repository system, as EPrints is very flexible and can be optimised and customised.
- Other open source software is available: DSpace, Fedora, Zentity ... partly because repositories have come up through the open access initiative, because open access is good for sharing. So the repository software is free, but there are overheads in developing and modifying it and you need a programmer to do these things.
- The other thing is storage. Usually what happens is that the institution has IT support and server space, so one could say the software is open source with caveats.

b) Access:

- The system is very user friendly. There is a searchable front end allowing detailed resources of searching. One can publish content on line, or not, and share it with other people who have a repository. EPrints is compatible with the OAI (Open Archives Initiative) protocol for metadata.
- Harvesting is a way of allowing more access in greater detail. This allows the user to speak to other repositories.
- Interoperability – so that one can hook up with other repositories with other information: probably this has been done somewhere, but he doesn't have any examples. In principle it's a good thing.
- Metadata are exported as xml so they are structured. Google queries in a dull way; xml has tags telling you what the field is so the repository can interrogate it and get better results.

Athanasios Velios: Is this standard?

Edward Pinsent: Yes, it's a specification.

Athanasios Velios: Is it a schema?

Edward Pinsent: This is a consultation question! [No further discussion of these technical points took place in this session].

- Access is structured into the workflow of how you upload things: you can declare to users online what you want to happen. An example is the creative commons licence (this too comes with caveats as it is not legally binding: it's an honour system). [Other examples given.]
- The other thing to do would be not to share: you embargo something for a certain number of years, and this too is built into the system.

c) Preservation:

- One would like digital content to be accessible for a long time, but what if the file format encoding becomes obsolete? You can intervene and migrate the content to another format that may last longer. e.g. Microsoft software: a) it's proprietary; b) it is said to become obsolete very quickly (is this true? They don't know). So it is better to migrate the file – for example, a Word document – onto something open source such as PDF; not only is this very strong, but also the format specification for PDF is published so it could be fixed if it goes wrong. Also tif files are better than jpg, and so on. But there is a trade-off: does everything in the original document survive the migration? Text is usually OK, but there are often embedded functions that may not transfer. Therefore investigation of core components of file types should be done and a declaration of components that need to be preserved should be made: e.g. AV material, websites – here lots of formats, scripts, and functions.
- EPrints has tried to align itself with digital preservation standards:
 - Steve Hitchcock [University of Southampton WAIS (Web and Internet Science) Research Group, Electronics and Computer Science (ECS), one of the developers of EPrints] is trying to turn EPrints repository into a preservation repository by adding a preservation stage to the workflow, but he (Edward Pinsent) does not know how well this is going [as at April 2012].
 - UAL Research Online was a partner in Hitchcock's preservation project ("KeepIt": see <http://preservation.eprints.org/papers/>). UALRO has installed the Eprints plugin that was developed as part of this project, which provides a list of file formats in danger of obsolescence, and allows the collection manager to migrate to up-to-date formats (while preserving an archived copy of the original).
 - EPrints Bazaar: a range of plug-ins that sort out file format identification and migration.

Athanasios Velios: Book conservation is a parallel example. The original bindings have not been preserved, only the content, and migration does the same thing: it takes what we think is essential – the text – and throws away the outside.

Edward Pinsent: You don't throw away the original. This is very important.

Everyone agrees that one then 'uses' the surrogate master.

Athanasios Velios: Why not use emulation?

Stephanie Meece: Because users will not have access to machines that can read 'pdf version 1.0'; in 20 years' time, Windows will not be able to read these ancient documents.

Edward Pinsent: Emulation is the no. 2 method of digital preservation. But it requires more technical expertise.

Panagiota Manti: Is uploading of the theses mandatory?

Edward Pinsent: Oddly enough, ULCC does not manage the University of London repository.

A repository can be a small-scale thing: simply for one school.

Meagen Smith, UAL Postgraduate student: How labour-intensive is it?

Edward Pinsent: It depends on content and how much traffic you have; the compatibility of your bandwidth and that of the hosting company.

Meagen Smith: The location of the server does make a difference. Also with copyright.

The possibility of cloud storage was raised.

Edward Pinsent, commenting on cloud storage: You buy into a bandwidth and storage, it looks cheap, but then you find you need to pay more ...

Athanasios Velios: IIC does not have its own server, in fact.

Stephanie Meece: You do need a repository manager. In fact, you need two people, one to handle copyright and management, the other IT, at least part-time.

Edward Pinsent: You need a repository support group and joint information management systems.

The take up of repositories has not quite bedded down as expected. One reason is that there are people who will not share. But there does have to be some management; you need this in any IT system. Institutional change needs to happen when you buy into a system. You must not ignore the organisational side.

[Discussion]

Katalin Orosz then added some comments to her presentation of yesterday (Wednesday 4 April, 2012), inserted above, q.v. One point made is that the catalogues are in Hungarian, but are somehow searchable in English (see above), thus something – some key words – must be in English. With reference to this,

Graham Voce: The question of language: translation of abstracts into English is extra work. Does one need to find funding to do this?

Katalin Orosz: Not all students speak English and we need to teach them. Most speak more or less and we ask them for an English abstract. The work lies in the translation of earlier ones. She (Katalin Orosz) does not know if it is manageable to teach them English in fact.

Graham Voce: A question for the archivists: is international connectability possible?

Edward Pinsent: This depends on the person managing the repository. It can be done.

Stephanie Meece: We are accepting standards that make it technically possible, for example Dublin Core and several others which are EU accepted. But the technical infrastructure is not very robust. *OpenDOAR* [a directory of academic open access repositories] is another one. You can search for documents this way, but people are relying on Google. Google Scholar should rank repositories high in the results but this is a blunt instrument.

Graham Voce: There is a problem with the word 'conservation'. How does one rule out trees and the rest? Would a thesaurus help?

Stephanie Meece: Yes; if it was agreed, this would help. But it would only work forwards. But you could ask managers to add these keywords retrospectively. This is a very challenging problem.

Meagen Smith: But it is a universal experience among conservators.

Edward Pinsent: There is scope for developing metadata schemes. There is one for photos (he cannot remember the name). A specialist interest can be represented.

Athanasios Velios: We do have experience in doing this for book bindings and it is very difficult. It is not a rapid progress.

[General discussion of structuring a thesaurus and general agreement that this is a large task.]

Angela Weyer: What about regional terms? These are not in English.

Graham Voce: There is a huge European effort of standardisation of conservation terms. And regional terms will be missed.

Jo Kirby Atkinson, IIC: A practical example is Doris Oltrogge's *Datenbank Mittelalterlicher und Frühneuzeitlicher kunsttechnologischer Rezepte in handschriftlicher Überlieferung*, held at the Fachhochschule Cologne, where there is a thesaurus of selected terms used for searching, but the unselected synonyms that one cannot use are not, of course, listed (although, behind the scenes, Doris Oltrogge will have them). And, of course, there are many, many similar examples.

[Athanasios Velios gave more information on the use of a thesaurus, particularly from the IIC point of view. General discussion.]

Athanasios Velios (finishing the discussion): The point of the thesaurus is not for the person doing the blunt Google search, but once you get past this you need the thesaurus

- for a more specific search;
- b) to find the selected term and to rule out other terms.

But one may have to end up on a link to something.

Stephanie Meece, University of the Arts repository (showing the website):

In constructing the UAL system we designed EPrints to properly showcase research in arts, design and media, our specialism at UAL, which is more visual. Theses have been added in. The students submit theses in hard copy and in digital form. There are three parts to the permission form that accompanies thesis submission:

- 1) The first part refers to permission to make the printed copy accessible in the library, and goes to the library;
- 2) The second part authorises the thesis to go into the British Library EThOS project (and gradually UAL will be digitising all the theses);
- 3) The third part authorises the thesis to be included in UAL Research Online: the student agrees to all or any of these three parts or not. (In effect, however, authorising part two automatically makes part 3 happen, and vice versa (see below). We will be rewriting our permission forms to remove the redundancy. The student also agrees to

take on responsibility for any infringement of third-party copyright, because the university will not assume that risk. – NB. In practice the university helps with copyright, as this is a complex and rapidly-changing area, and most research students are not sufficiently informed to take this responsibility. The thesis is written for very particular requirements for the university: it is *not* a published book. So if the student needs a portrait, it goes in, but it may be taken out of the on line version. Sensitive data and other copyright content are removed. Or the student can request permission to reproduce, but then generally a fee is paid. There is a take-down policy: if the copyright holder requests that the image be removed, it is removed. So in practice they need to decide the level of risk, and how to mitigate it.

Examples of theses were shown to the audience on the website: one that has to be requested as its subject matter could be seen as pornographic; another with images that have been removed.

A typical thesis consists of a PDF of 130 pages: downloading it can therefore take a long time if the connection is slow (this is an example of a snag for the user).

The system generates a cover page for the downloaded PDF. There is a link to the policy of usage; creative commons attribution: non-commercial; no derivatives.

They actually do this for the students because access may be difficult for the student who has graduated and is no longer able to log in.

If the student wants to publish – fine: he/she has the copyright.

If work has to be taken down it is taken down.

They have fewer than 100 theses in the system at the moment as they do not have digitisation facilities [situation in 2012].

If the thesis is on EThOS it is downloaded by UALRO staff and added to UALRO. Ethos also now automatically harvests UALRO for our theses, in case we have something they do not.

To find a thesis: in Google, put in author's name and a few words; you can get right into the thesis. And you still get a cover page.

Using EPrints it is possible to get a plug-in to do it; you have to tell it the fields you want to include.

OpenDOAR – Directory of Open Access Repositories – lists repositories and it should be possible to search content. This is world-wide.

Tools for querying:

- DART-Europe E-theses portal; this is a partnership of European research libraries aiming to provide access to European electronic PhD theses; it only searches theses that exist in electronic form (on their system, presumably) and you have to register with them.
- EThOS (UK) Electronic Theses online System – this deals with everything that might be needed, including copyright.
- Repositories Support Project (tel. 0845 2576860) – this covers how to set the repository up, training, etc. and is a more general system than EThOS.

Edward Pinsent: Are there technical protection measures that can be used to cover copyright?

Stephanie Meece: This is a very difficult matter because one wants to share the research and a photo is a research item: if you watermark it you change it.

Edward Pinsent and Stephanie Meece agree that one can watermark things, but that watermarks can be broken.

Edward Pinsent: Is the backlog to be digitised?

Stephanie Meece: Unfortunately this is not possible at present (shortage of staff, the logistics and large scale of the task and so on). We are relying on the British Library's Ethos project for digitisation of our print-only theses; when one of our older theses is requested via the BL's catalogue, we receive a message, send the hard copy to them, and they digitise it and upload it to the Ethos website. Then we add it to UALRO. This means that our more valuable older theses should rise to the surface over time.

N.B.. All UAL students now submit a digital version of the thesis, so digitisation will be less and less a concern. Repositories at other universities are now moving to submission through the repository itself, so that all theses are captured and described by their authors and UALRO will look into this as well.

Angela Weyer: How about the url? This stays? (This was agreed). And the DOI system of citation: DOI numbers can be broken which is worrying.

Here there is a general discussion of DOI numbers as a system; as Angela Weyer says they are generated by the library and others think they are generated by the publisher, although it is true that there must be some overarching authority that allocates them. (The DOI service and registration is supervised and provided by a federation of registration agencies – <http://www.doi.org/>). It is generally agreed that it is distinctly worrying if this number system can break down.

Stephanie Meece comments in 2013: Since this meeting, we have investigated the possibility of creating DOI numbers for our online theses. It is still not clear if this will work for repositories, but I think it will become possible in the next few years.

Maria Gruber, University of Applied Arts, Vienna: Are download statistics possible?

Stephanie Meece: Yes, this can be done. For example we had 400 people in China looking at stage design. We can analyse keywords used, source of the referral to us, length of time spent on our pages, which other pages were visited, whether the visit is the first time or a repeat, etc.

Maria Gruber: We have a diploma image database, but there are several search places and this is confusing for a student.

Everybody agrees that this is a universal problem.

Graham Voce: What do students actually want? Are repositories part of this?

Heather Ravenberg, Postgraduate Student, University of the Arts, London: A lot of issues apply to students using the repository. Targeted searches are more useful than 'Conservation of an 18th-century icon' which is very general. It is also useful to know if something has been 'published' elsewhere: from the point of view of a person writing an MA dissertation, one asks, 'Is this something I can use?' It gave ideas of what had been written, but were they useable?

Anna Hoffmann, also a Postgraduate Student: It is useful as it not only gives content, but also how to compile and write a thesis. (She studied in Cologne before, where the teaching and structure of a dissertation is very different.) In Camberwell, though, this is not taught.

Stephanie Meece: Lecturers don't know the answer to this!

Graham Voce: So the repository, what it contains and how you use it: should teachers be teaching this?

Heather Ravenberg: At the BA and MA level there is some instruction, but not a lot.

Meagen Smith: If the programme being followed is less lecture-intensive then this becomes more important. They (theses) are valuable for grey areas of how you do research. If IIC or another group is helping me to get to these really useful resources, like that in Hungary, then it's really useful.

Graham Voce: This is the main aim: how IIC can enable people to find out about this sort of material as far as we can. There are so many varieties of interns, technologies, priorities of what people think is important; we cannot offer everything, but we can offer students the possibility of starting the exploration and making their own links.

Heather Ravenberg: The repositories are unknown to the students! They don't just need to know they're out there, but also how to access them.

Angela Weyer: The database of the Hornemann Institut is logging downloads or receiving requests for PhD theses from all over Europe because there is *so much* grey literature.

Karen te Brake-Baldock: Working with the educators' subgroup of ICOM-CC, and there is a subgroup of INCCA there: would this be a useful link for IIC?

Graham Voce and Jo Kirby Atkinson: Yes, it certainly would.

Maria Gruber: How about ENCoRE (European Network for Conservation Restoration Education, <http://www.encore-edu.org/>) and also INCCA?

Graham Voce: ENCoRE is already part of the umbrella, but we also need to consider Australia, for example, and people in far flung areas where maybe internet access is the main form of access and the paper stage is being by-passed. *Studies in Conservation* is now getting a lot of submissions from China.

Athanasios Velios: We can link or we can harvest, but what we need is more specific metadata on the IIC website. Would the students be willing to help with this?

Meagen Smith: Yes, under certain circumstances. And generationally they are willing to do this: they are not as protective of data as more conventional researchers.

Karen te Brake-Baldock: IIC is not building a repository?

Jo Kirby Atkinson: No, a link – or links: more of a portal.

Graham Voce: We are putting people together. IIC's job is allowing people to connect, to see how it's done elsewhere.

Karen te Brake-Baldock: It could be that there is the possibility of an EU project under the cultural programme (Note: rather than the framework programmes) and these can bring in the Getty and so forth.

[There followed a discussion of the template of a modified repository from EPrints, which could be shared with others wanting to set up a repository.]

Panagiota Manti: How do you add more metadata if the repositories are linked?

Athanasios Velios: You keep a reference point on your local database – the IIC in this case. Populate it with more specific terms related to conservation. Maybe you could replicate the graphical records, but you would need to keep the records synchronised. If you harvest once in five years (Note: this question was asked earlier and it seemed to be what people did) this is reasonable. The difficult part is populating the metadata, which students can do.

Meagen Smith: This will develop as time goes on; it grows organically from the group of experts who will come together and want it more focussed.

Athanasios Velios: The starting point is the AAT Getty Thesaurus. But there is still discussion as to whether we extend it or replicate it. Compliance is good. I would be happier if any term proposed was assessed by AAT first.

Graham Voce: Heather, any more thoughts?

Heather Ravenberg: There needs to be quality control. Not everything should be included. Presumably this must be at university level.

[Various people suggest the use of the internet and social media software to insert comments.]

Heather Ravenberg: The problem is the bad thesis.

Athanasios Velios: It is necessary to make users aware of the conclusions of the work alongside the primary sources. And this is a very big problem. The evidence is not in the abstract.

Karen te Brake-Baldock: You can have an idea of quality from the abstract?

Athanasios Velios: Not if people do not speak English very well.

Graham Voce: Availability and the rest are not IIC's problem in every respect.

Athanasios Velios: Except that IIC is really a culture and an environment.

Karen te Brake-Baldock: Which is why I suggest working with the ICOM-CC Education Group as they are addressing this on the dissemination side, but one can also discuss quality with them.

Graham Voce: The question is, how much grey literature should be accessible?

Jo Kirby Atkinson: In practice IIC cannot moderate this literature; this would be an enormous task.

Meagen Smith: But it can give an idea of how much moderation has been carried out.

Edward Pinsent: It is possible to send out a wrong signal to people.

Some marking system like Amazon or TripAdvisor may be a good idea.

Sarah Gould, EThOS, British Library (attending on line): EThOS – E Theses online Service is a web resource holding records for virtually all UK PhD theses: about 300,000 records. There are 57,000 full text open access copies of theses; also about 20,000 have a link to a repository at the awarding institution where there is a full text. This system grew out of the British Library document supply service and the information that was held in the catalogue. The theses were held on microfiche and lent out to other libraries. EThOS is an open website which anyone can use. It is limited to UK theses, PhD or equivalent level. The system is three years old. Originally it was purely a British Library service, but in 2010 it became clear that the business model was not working well and a partnership developed with higher education institutions of the UK. So it is a shared service now. It has not changed all that much. One of the underlying principles is that of open access: information is free to use and to download.

Function: EThOS harvests metadata from repositories around the UK, full texts where institutions want it done and permit it. The EThOS–institution relationship depends on the institution so some harvest metadata *and* the full text. There may also be digitised older material coming through, so institutions with paper theses send the title page and EThOS creates a record for this. If the copy is not held, the user can link to the institution holding the copy, or the user can order a digital copy of a paper copy thesis. Sometimes the end user has to pay to have this done. Or the institution may not want to do the digitisation and the end user has to travel to the institution and read the paper copy.

Some examples

- Courtauld Institute: 360 records of theses, 1 has full text pdf available;
- Royal College of Art: 203 thesis records, 46 with full text available;
- UAL: 101 thesis records, 43 with full text available.

Records vary in quality and comprehensiveness: they give author, title, etc.; abstracts are increasingly available and link to the institution. The subject is indicated by Library of

Congress subject headings and there is a retrospective programme adding Dewey information (Note: both are standard library classification schemes). A lot of records have free text keywords added.

EThOS is a very fast-moving environment: three years ago there was a holding, with digitisation leading to a bucket of content. Now it has interoperability with other systems and repositories.

Is it OAI PMH (Open Archives Initiative Protocol for Metadata Harvesting) compliant? No, not yet [situation in 2012], so they are sending metadata out. EThOS records are also on the British Library catalogue (q.v.) and they use this to produce EThOS.

For end users the service is good – about half a million downloads of full text content. All UK universities have some sort of record in EThOS.

Challenges include the duplication of records between EThOS old and EThOS new, and between EThOS new and other repositories. They are having to sort this out.

EThOS has to take account of embargoed theses and replicate this requirement. The take-down policy is clear: EThOS verifies the request from the author with the awarding institution and then takes the thesis down.

Re-harvesting of updated records: this is a two-way process and soon they will be able to harvest out and allow institutions to do this too.

Permissions from author are key. There are probably ten take-down requests a month. EThOS does not check the copyright; they assume the universities have already done this.

Athanasios Velios: Who could harvest?

Sarah Gould: A few years ago the British Library would have had the only repository, but now the institutions have their own so they want to pull back their theses onto their own systems, and to apply their own metadata. Also OAI PMH allows repositories to be interoperable, so metadata are harvestable through open access/ open archive harvesting.

Graham Voce: Can this be done internationally?

Sarah Gould: We are just starting to share metadata with international services and this drives traffic to EThOS and on into the institutions. Anything we or they can do to share metadata outwards will help the institutions.

The end users are people. They may be in the UK or anywhere; they may or may not be from higher education or commercial users.

Graham Voce: We have been discussing on focussing down searching and the use of our thesaurus.

Sarah Gould: The search facility on EThOS is overall: you can search by field, but you cannot sort or filter. It is very basic.

EThOS had been focussing on digitisation, then without any advertising suddenly they found themselves with 200,000 records and digitisation brought in by the universities went through the roof. Demand rocketed and they used one year's digitisation money allocation in 6 weeks! The knock-on effect was that the basic underlying search service was not developed, but with an institutional membership service this can be developed to improve the metadata and harvest more.

Graham Voce: What are your longer term aims?

Sarah Gould: Business as usual: a service that the British Library will offer as it does now, while improving the way we talk to the other institutions. Content is rolled out to other scholars.

Of the international systems: One, NDLTD – Network Digital Library of Theses Digitisation – is American-hosted. This is not aimed at searching by end users; it is just a library catalogue. EThOS records would then be downloadable by NDLTD.

A European one is DART-Europe.

So a lot of records will crop up in several places at once.

[Discussion]

Sarah Gould: The search functionality is not that great, but we can provide IIC with a list of records.

Graham Voce raises points IIC wants to take forward.

- Quality control: IIC would have to say that they were not responsible for the quality or opinions expressed in the resources, i.e. a disclaimer. IIC just says here it is, not what it is.
- Hosting: IIC will not host the texts.
- Languages: These are a problem.

In fact you will be matching xml fields.

Athanasios Velios: When you build the thesaurus you translate it or ask someone to do it. You can use a wiki for translation, but once you start translating you have to go on with it; it is also a huge responsibility.

Edward Pinsent: When using the thesaurus, start with English and see how many words get hits.

Karen te Brake-Baldock: In the Netherlands there is a Dutch version of the AAT headings.

Gustavo Grandal Montero: This is not a good use of resources.

Karen te Brake-Baldock: So if one was harvesting from a repository in a different language one would need search words in English added?

Gustavo Grandal Montero: From the document itself, and fields would be added by the cataloguer.

Jo Kirby Atkinson: The official languages of IIC are English and French, and French is still important for the southern European restoration community. (There was general agreement on this)

Graham Voce, continuing his list of points:

- Copyright

There was more discussion on this, leading on to

- Access

Everybody thinks the service provided by IIC should be open access – this after a discussion between Athanasios Velios and Jo Kirby Atkinson as to who can access the IIC service and whether it should be members only.

Edward Pinsent: If IIC is providing a searching service (in effect) this could be described as a premium service and so for members only (in effect, chargeable).

Panagiota Manti: But if the repositories and source material themselves are open source and available to anyone is this legal?

Gustavo Grandal Montero and everybody: You cannot take other people's open source material and then in effect charge for it.

Various comparisons were made, for example JISC (a UK body advising on the use of digital technologies in education and research, <http://www.jisc.ac.uk/>).

Athanasios Velios: So IIC has

- 1) a charging model, or not;
- 2) Thesaurus models – AAT:

One may have geographical places, or cultural objects, e.g. the Parthenon – as a term we know.

One has a context, one has a thesis and the work described usually refers to an object, a piece of information; where it is done and who did it. But in the thesis one might also have the results of the analysis of the painting (if the object had been painted) and all sorts of other things. Alongside this, we need the vocabulary to fill in the information for each bit. So there are three extra fields: we also know other people have worked on other objects or some on similar sites, using similar technologies. This is not too difficult for students to fill in.

Maria Gruber: I would add topics like consolidation, also materials as individual fields.

Karen te Brake-Baldock: Who is doing this: the student on his/her own thesis?

Another point for IIC to bear in mind: partner organisations [several partner organisations were suggested both in the meeting and by people responding to the questionnaire and these organisations are to be contacted]: the Education Working Group of ICOM-CC could be helpful in taking this project forwards and there is also a subgroup of INCCA. This is where an EU project would come in, otherwise it's a lot of work (and in any case IIC is too small an organisation at the administrative level to apply for EU money on its own because of the small number of employees). IIC should try searching under Culture Programme, Culture 2013, etc.

General summing up: The meeting has covered knowledge transfer; outreach. IIC needs to consider a project initiation document; what is achievable in a certain amount of time. Here the scope and a realistic aim are important.

IIC's website can never host the theses, but this was never the intention.

[General discussion, thanks and goodbyes.]

