

Information Access to Cultural Heritage

ECDL 2008, Aarhus Denmark, 18 September 2008

Martha Larson
University of Amsterdam, Netherlands.
m.a.larson@uva.nl

Kate Fernie
Kate Fernie Consulting, UK.
kfernietiscali.co.uk

Johan Oomen
Netherlands Institute of Sound and Vision,
Netherlands.
JOomen@beeldengeluid.nl

Juan Miguel Cigarran Recuero
UNED, Spain.
juanci@lsi.uned.es

Abstract

The workshop on Information Access to Cultural Heritage (IACH 2008) was held during the ECDL conference in Aarhus on the 18th September 2008. The workshop provided a venue to bring together academics carrying out research in the area of information access and practitioners working in the cultural heritage field. The aim was to encourage the exchange of ideas concerning the creation, curation, storage, retrieval and use of cultural heritage information. The day began with a keynote paper by Chris Batt, who looked to the future, challenging the audience to think of uses of technology to make sense of the landscape for users, shifting focus of information access to cultural heritage from *search* to *find*. The keynote was followed by a session on the European Digital Library with invited papers by Stefan Gradmann on the infrastructure of Europeana and by Rob Davies on work to develop the content of the Library in the EuropeanaLocal project. The proceedings continued with a series of papers on services being developed to provide access to cultural heritage collections and a wide range of media – images, text, music and speech. A lively series of poster boosters followed and the day concluded with the poster session and a panel discussion on future directions. Given the attendance at the workshop and the good level of discussion throughout the day, we conclude that Information Access to Cultural Heritage is a growing area of research and promising area for exchange and collaboration between researchers and practitioners.

1 Introduction

The workshop on Information Access to Cultural Heritage (IACH) was held during the ECDL conference in Aarhus on the 18th September 2008. The workshop provided a venue to bring together academics carrying out research in the area of information access and practitioners working in the cultural heritage field. The aim was to encourage the exchange of ideas concerning the creation, curation, storage, retrieval and use of cultural heritage information.

This report gives a summary of the workshop, starting with an overview and then continuing with a description of each of the workshop sessions and of the panel session that concluded the workshop.

Further information, as well as the proceedings [7] of the workshop, are available on the workshop website , Accessed 07-10-08.

Europeana Infrastructure, S. Gradmann (Humboldt-Universität zu Berlin)

EuropeanaLocal — Its Role in Improving Access to Europe's Cultural Heritage through the Europeana service of the 'European Digital Library,' R. Davies (MDR Partners)

Enabling Semantic Access to Cultural Heritage: A Case Study of Tate Online. P. Clough, J. Marlow, N. Ireson (University of Sheffield)

Enigma Helvetia: Promoting an Exhibition through Multiple Channels, N. Di Blas (Politecnico di Milano), E. Rubegni (University of Lugano), P. Paolini (Politecnico di Milano)

Having Triplets — Holding Cultural Data as RDF, K. Byrne (University of Edinburgh)

Access to the Music of Man Archive: The MARIMBRO System, A. Schneidt (Logica), K. Klenke (Hanover University of Music and Drama) T. Mandl (University of Hildesheim), J. Medinvil, R. Vogels (Hanover University of Music and Drama), C. Womser-Hacker (University of Hildesheim)

Towards Affordable Disclosure of Spoken Word Archives, R. Ordelman, W. Heeren, M. Huijbregts, D. Hiemstra, F. de Jong (University of Twente)

Table 1: Oral presentations at the IACH 2008 workshop

The IACH 2008 workshop was sponsored by the EU IST-FP6 Project MultiMatch: Multilingual/Multimedia Access To Cultural Heritage. The workshop program consisted of oral and poster presentations of papers selected from the submissions received in response to the workshop call for papers. Submissions were reviewed by 15 experts from academia and cultural heritage institutions who were invited to serve on the programme committee. Chris Batt was invited to give a keynote speech, and delivered a talk entitled 'Building Bridges to Knowledge.' Presentations on Europeana [2] and EuropeanaLocal [3] were also invited. The oral presentations of the workshop are listed in Table 1 and the poster presentations in Table 2. The rest of this workshop report is devoted to giving an overview of the topics addressed in the sessions of the workshop and the panel discussion at the end of the workshop.

2 Workshop presentations

2.1 Keynote: What do you want the future to be like?

Chris Batt (former chief executive of the UK's Museums, Libraries and Archives Council) began his keynote by asking the audience what they wanted the future of information access to be like and continued by focussing on the need to improve services for users. He described the current environment as 'the land of the informed bewildered' where users search into an unknown world and spoke about the importance of making sense of the landscape and enabling users to go beyond the shop window

presented by search engines into the deep web. Batt challenged the audience to ‘think *find* and not search’ and to envision applications of technology that support users in their discovery of cultural heritage resources instead of focussing on sophisticated search techniques.

An Intelligent Service Oriented approach for Improving Information Access in Cultural Heritage, A. Baruzzo, P. Casoto, P. Challapalli, A. Dattolo (University of Udine)

Providing linked-up access to Cultural Heritage Data, T. Blanke, M. Hedges (King's College London)

Multimedia Retrieval in MultiMatch: The Impact of Speech Transcript Errors on Search Behaviour, J. Carmichael, P. Clough (University of Sheffield), G. Jones, E. Newman (Dublin City University)

A System Architecture for Conveying Historical Knowledge to Museum Visitors, D. Dannells (University of Gothenburg)

Feasibility of a European Registry of Standards for Digital Heritage, M. de Niet (DEN: Digitaal Erfgoed Nederland)

Marking Up Cultural Materials for Time and Geography, F. Gey, R. Shaw, R. Larson, M. Buckland, B. Pateman, D. Melia (University of California Berkeley)

Searching for Linguistic Phenomena in Literary Digital Libraries, F. Sánchez-Martinez, M. L. Forcada, R. C. Carrasco (Universitat d'Alacant)

Table 2 Poster presentations at IACH 2008

2.2 Digital cultural heritage: European initiatives

The papers in the ‘Digital Cultural Heritage: European initiatives’ session focussed on current work to build the European Digital Library, a flagship initiative of the European Commission’s i2010 agenda. Stefan Gradmann began by describing the political background and timeline for building Europeana, the first prototype of Europe’s Digital Library, Museum and Archive. Gradmann went on to describe the technical infrastructure for Europeana, which builds on work by TEL, Minerva, MICHAEL, EDLNet, DELOS, BRICKS, QVIZ and other EU-funded projects. The current project will create a single-access point to 6 million digital objects from libraries, museums and archives. Gradmann described the Europeana first prototype as a catalogue-based portal that will hold index files and object surrogates, pointing users to the digital objects held by cultural institutions via simple qualified Dublin Core metadata. Gradmann presented an overview of plans for future Europeana prototypes with additional functionality, the semantic resources needed and the potential for interactive services. Rob Davies built on Gradmann’s presentation, describing the EuropeanaLocal project, which aims to provide support to local cultural institutions in making their digital content available to Europeana. Davies reported that a preliminary survey of the content available from the institutions involved in the EuropeanaLocal project had identified 20 million items, which will be made available to Europeana via a network of regional aggregators. Davies noted that the project aims to demonstrate the importance of content available from local and regional institutions for the European Digital Library. He concluded by remarking on the

importance of geography in finding content interesting to a locality or region, calling for additional geographic search functionality in future Europeana prototypes.

2.3 Cultural heritage exhibitions and collections

After the break, Paul Clough reported on research on semantic work to facilitate finding, integrating and connecting cultural heritage information based on a case study of Tate Online. The research project included a user evaluation study to inform the design of a prototype interface by identifying important information facets for the Tate's audiences. An interesting aspect of the project was work undertaken to extend the Getty's Union List of Artist's Names (ULAN) by integrating it with details of additional artists identified through the Tate's collections. Clough reported on difficulties in dealing with contradictory data, for example differences in the expression of an artist's date or place of birth in the different datasets. The Tate Collection prototype is now available online.

Paolo Paolini went on to talk about the concept of 'Instant Multimedia' and the rapid development of websites to accompany museum exhibitions. Paolini described the case study of the 'Enigma Helvetica' exhibition, which was about contrasting views of Swiss culture. The accompanying website offered users a chance to explore this 'strange and complex exhibition' through a series of audio-visual topics. Paolini reported that the methodology allowed for the development and delivery of multimedia presentations within 30 days and was now being used by curators and school groups to develop new resources. The research also looked at the frequently asked question of whether such websites stimulate people to visit museum exhibitions in person, finding that people were visiting the Enigma Helvetica website both before and after seeing the exhibition.

The final paper in this session was 'Having Triplets' by Kate Byrne. Byrne described her research project into the conversion of the Royal Commission on the Ancient and Historical Monuments of Scotland's (RCAHMS) database to RDF format. She described the benefits of conversion to RDF both in terms of enabling access to the hidden web, by revealing the content of databases to web crawlers, and in terms of the potential to connect data from different sources. Byrne's work on conversion of the RCAHMS database included experimentation with automated tools. Having estimated that this would produce an astonishing 235 million triples, Byrne concluded that there was a need for simplification. With schema knowledge and by avoiding literals, Byrne concluded that it was possible to reduce the number of triples by tenfold.

2.4 Access to Cultural Heritage Multimedia

After lunch the focus of the workshop was on providing access to multimedia formats. Christa Womser-Hacker began by talking about the MARIMBRO System, which provides access to the Music of Man archive at the University of Hildesheim and Hanover University of Music and Drama. The collection, which is currently in the process of being digitised, includes a wide range of materials relating to the ethnology of music – ranging from instruments to recordings of performances, videos and books. Specialists have developed various systems for classifying this sort of material, for example based on the instruments, the playing method, place and so on. The novelty behind the MARIMBRO system is that it offers multi-faceted ontologies and allows for switching between the facets. Womser-Hacker reported that integration of the different ontologies was challenging and required the involvement of experts in the field of music ethnology. Future developments for the system include looking at collaboration with users via social tagging.

The final paper was given by Roeland Ordelman, who described research into techniques aiming to support the affordable disclosure of spoken word archives. Ordelman reported that there had been a shift

in recent years from the use of laboratory data, such as benchmark evaluations of speech, to models based on real world data. He described the surprises and difficulties of working with live interviews, for example hesitations, accents and background noise. The archive collections worked with included radio interviews with the Queen of the Netherlands made during the Second World War and interviews with Dutch survivors of the Second World War camp at Buchenwald. Ordelman reported that the speech of some individuals was particularly difficult to understand, increasing the difficulty of training speech recognition models and developing systems to automatically transcribe their speech. Current research is focussing on understanding the errors and tuning speech recognition systems to improve results. In the future, plug-and-play systems may become available but for the present Ordelman reminded the audience that it is currently cheaper to employ people to transcribe small speech archives.

2.5 Poster Session

Following poster boasters, workshop participants had the opportunity to speak to the presenters during a lively poster session. The posters covered a wide range of topics relating to information access in the cultural heritage (cf. Table 2) including, for example, research on marking up cultural materials to support retrieval based on time and geography presented by Fred Gey of the University of California, Berkeley. The EU-funded MultiMatch project presented some of its research into the impact of speech recognition errors on search behaviour. Marco de Neit of Digital Heritage Nederlands presented a poster that called for the establishment of a European standards registry for the digital heritage.

3 Panel Discussion

The workshop concluded with a panel session, which had four discussants: Marco de Niet (DEN: Digital Heritage Netherlands), Joyce Ray (IMLS: Institute of Museum and Library Services, USA), Richard Wright (BBC Archives, UK) and Maarten de Rijke (University of Amsterdam, Netherlands).

The session opened with the panel being asked the question: ‘Which technologies have been important for information access to the cultural heritage and which will continue to be important?’ The panellists approached this question in different ways, with Marco de Niet beginning by highlighting on the coordinated actions, which, in recent years, have increased momentum in bringing cultural heritage collections and institutions on line. While de Niet and Joyce Ray focussed on user issues, Maarten de Rijke and Richard Wright spoke about the technologies and the importance of making catalogues available. De Rijke suggested that, given the massive volume of data we are dealing with, we need to give up on top down structuring and exploit bottom-up and statistical approaches. He also mentioned the importance of open tools, citing as an example the semantic annotation services made available by Open Calais. Chris Batt contributed to the discussion from the floor, highlighting the emergence of technologies that are enabling users to combine resources, for example combining images from Flickr to create a three-dimensional picture or model of a building such as Notre Dame [6]. De Niet noted the importance of user-based criteria to drive developments from the beginning.

The next question asked the panel to consider whether there are actually technologies that hinder information access to cultural heritage, picking up on earlier discussion on copyright and IPR. Joyce Ray suggested it would help if technologies were more widely available that make it possible to discover if something is in the public domain; Richard Wright mentioned ways forward: exploiting the large of amount of factual material in the public domain and making maximum use of permissible digests and excerpts for copyrighted material.

Paul Clough asked the panel if users are ‘experiencing information overload.’ A discussion followed which focussed on content and trust, with Richard Wright beginning by noting that the ‘low volume, low

interest market is the saviour of many archives.’ Joyce Ray mentioned the risk that ‘content is here today and gone tomorrow’ and championed libraries and archives as being vital for maintaining collections in the long term. Marco de Niet continued by noting the importance of user-generated content, which is often of high quality, alongside the content from cultural institutions. Kate Byrne contributed to the discussion from the floor noting the importance of provenance in terms of assessing the trust-worthiness and authoritativeness of material. De Niet noted that authority is the driving force of Europeana.

Chris Batt described quality as an opportunity, as something to aspire to. He warned of the danger of applying different standards on-line and in the real world – suggesting that institutions sometimes require their on-line data to be more perfect than their real-world collections. Following on from the earlier discussion about user generated content, which included systems that enable users to contribute answers and judge each other's answers mentioned by Maarten de Rijke, the question of preservation of this additional content was raised from the floor.

The panel discussion concluded with a call for the audience and cultural heritage institutions to face head-on upcoming challenges, e.g., ‘go out there and prepare for public access’ and to support accessibility and interoperability by making use of open source software and open standards.

4 Conclusion

The IACH 2008 workshop proved to be a very successful venue in which to bring together practitioners and academic researchers to discuss new research and new developments in information access for the cultural heritage. The workshop was well attended with ca. 40 participants from Europe, Canada, the USA, India, Sri Lanka, New Zealand and Australia. Another notable feature of the workshop was the breadth of the research topics being reported on, including specialized technologies under development, which in the future are likely to co-exist in information access systems for the cultural heritage. We concluded that organising such a workshop to bring together people with different specialities and experiences offers a valuable opportunity to exchange knowledge and build a foundation for future research and development in the area of information access to cultural heritage.

5 References

- [1] Clough, P., Marlow, J., Ireson N., Tate Collection, University of Sheffield prototype, <http://www.dcs.shef.ac.uk/~nsi/tate-collection.html>, Accessed 07-10-08.
- [2] Europeana, <http://www.europeana.eu>, Accessed 07-10-08.
- [3] EuropeanaLocal, <http://www.europeanalocal.eu>, Accessed 07-10-08.
- [4] Flickr, <http://www.flickr.com>, Accessed 07-10-08.
- [5] IACH 2008 Website, <http://ilps.science.uva.nl/IACH2008>, Accessed 07-10-08.
- [6] Knight, W., Holiday snapshots used to model the world in 3D. New Scientist Tech. <http://technology.newscientist.com/article/dn12846-holiday-snapshots-used-to-model-the-world-in-3d.html>, Accessed 07-10-08.
- [7] Larson, M., K. Fernie, J. Oomen and J. Cigarran (eds.), *Proceedings of the ECDL 2008 Workshop on Information Access to Cultural Heritage*, Aarhus, Denmark, September 18, 2008, ISBN 978-90-813489-1-1.
- [8] MultiMatch: Multilingual/Multimedia Access To Cultural Heritage, <http://www.multimatch.eu>, Accessed 07-10-08.
- [9] OpenCalais. <http://www.opencalais.com>, Accessed 07-10-08.