



# CULTURAL HERITAGE *on line*

***Empowering users: an active role  
for user communities***

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009



# Cultural Heritage on line. Empowering users: an active role for user communities

Florence, 15<sup>th</sup> - 16<sup>th</sup> December 2009

Sotto

l'alto patronato del Presidente della Repubblica

Under the Patronage of



Promoters of the Conference



Library of Congress



MINISTERO  
PER I BENI E  
LE ATTIVITÀ  
CULTURALI

MiBAC  
Ministero per i Beni  
e le Attività Culturali



Fondazione Rinascimento Digitale

Proceedings are edited by Chiara Cirinnà and Maurizio Lunghi

A special thanks to all the staff working for the organisation of the conference: Emanuele Bellini, Alessio Bertolani, Benedetta Caporioni, Maria Caterina Cunsolo, Verdiana Fontana, Gabrielle Giraudeau, Irene Lisi, Damiana Luzzi, Matteo Paternò, Marco Rufino

Special thanks to the hospitality of Teatro della Pergola:



Special thanks to the Supporters of the Conference



Many thanks to:

Firenze University Press, Scuola di Dottorato di Ricerca in " Telematica e Società dell 'Informazione" - Università degli Studi di Firenze, Università degli Studi di Roma "Tor Vergata", Fondazione Mediateca Regionale Toscana, Fondazione Giangiacomo Feltrinelli, Sistema Bibliotecario di Ateneo - Università degli Studi di Firenze, Società Dante Alighiana, Società Internazionale per lo Studio del Medioevo Latino, Fondazione Ezio Franceschini, Memoria Ecclesiae, Federation of Telecommunications Engineers of the European Union – FITCE, Associazione Comunicazioni e Tecnologie dell'Informazione – AICT, Master in Management, Marketing e Multimedialità per i Beni e le Attività Culturali - Politecnico di Torino, Fabbrica Europa, Master in Catalogazione Informatica del Patrimonio Culturale - Università Cattolica del Sacro Cuore, sede di Milano, Master in Conservazione, Gestione e Valorizzazione del Patrimonio Industriale - Università degli Studi di Padova, Master in Diagnostica Avanzata per i Beni Culturali - Università degli Studi di Bologna, Master in Indicizzazione di Documenti Cartacei, Multimediali ed Elettronici in Ambiente Digitale - Università degli Studi di Roma Tor Vergata, Master in Restauro Digitale Audio/Video - Università degli Studi di Roma La Sapienza, Master in Studi sul Libro Antico e per la Formazione di Figure di Bibliotecario Manager - Università degli Studi di Siena, Master Internazionale in Information Technology - Scuola Superiore Sant'Anna, Segreteria didattica Facoltà di Conservazione dei Beni Culturali - Università Alma Mater di Bologna, Scuola Normale Superiore, PERCRO - Scuola Superiore Sant'Anna



## **Table of contents**

Michele GREMIGNI - Welcome	7
Maurizio LUNGHI - Preface	8
<i>Tuesday 15th December, Plenary Session</i>	
<i>Invited lectures</i>	10
Laura CAMPBELL	11
Collaboration and interaction on the Web at the Library of Congress	
Luciana DURANTI	16
The Long-term preservation of digital heritage - The Case of Universities Institutional Repositories	
Daniel TERUGGI	20
Ethics of Preservation	
Andrea GRANELLI	24
Learning processes on the Net: more information or noise?	
Helen TIBBO	27
User-Based evaluation in the Web 2.0 world	
Daniel J. COHEN	28
Cooperative Web tools and user-generated content for cultural heritage: advantages and limits	
John UNSWORTH	33
Computational methods in humanities research	
Ingrid PARENT	38
Internet-Driven Convergence Between Libraries, Archives and Museums: an opportunity, an inevitability or both?	
Manuela SPEISER	41
Digital Libraries and Digital Preservation: EU-Research Perspectives	
<i>Wednesday 16th December, Plenary Session</i>	43
<i>Invited lectures</i>	
Jill COUSINS	44
Europeana: of the user, for the user	
Rossella CAFFO	45
Digital Libraries programs in Italy	
Stefano VITALI	51
The SAN Portal: a common gateway to Italian archival resources on the Web	
Gianbruno RAVENNI	56
Cultural Heritage On-line. Empowering users: an active role for user communities	
<i>Parallel sessions I</i>	
<i>Digital library applications &amp; interactive Web</i>	57
Anna Maria TAMMARO	58
Introduction to the session	
Brian KELLY, Charles OPPENHEIM	60
Empowering users and their institutions: a risks and opportunities framework for exploiting the potential of the Social Web	



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

Smiljana ANTONIJEVIC, Laura J. GURAK Trust in online interaction: an analysis of the socio-psychological features of online communities and user engagement	65
Max KAISER, Jeanna NIKOLOV-RAMÍREZ GAVIRIA, Veronika PRÄNDL-ZIKA EuropeanaConnect - Enhancing user access to European digital heritage	69
Fred STIELOW Perspectives from an online university community - Community Building in the Web Revolution	73
Silvia GSTREIN, Günter MÜHLBERGER User-driven content selection for digitisation - the eBooks on Demand Network	78
Aly CONTEH, Asaf TZADOK User collaboration in mass digitisation of textual materials	83
Serge NOIRET The European History Primary Sources (EHPS) portal at the European University Institute, Florence	87
Zinaida MANŽUCH Digitisation and communication of memory: from theory to practice	92
Andrea BOZZI Pinakes Text. A tool to compare, interoperate, distribute and navigate among digital texts	96
Frank AMBROSIO, William GARR, Edward MALONEY, and Theresa SCHLAFLY MyDante and Ellipsis: defining the user's role in a virtual reading community	98
Wendy M. DUFF, Jennifer CARTER, Costis DALLAS, Lynne HOWARTH, Seamus ROSS, Rebecka SHEFFIELD, and Cassandra TILSON The museum environment in transition: the impact of technology on museum work	103
Tomi KAUPPINEN, Tuukka RUOTSALO, Frédéric WEIS, Sylvain ROCHE, Marco BERNI, Eetu MÄKELÄ, Nima DOKOOHAKI, and Eero HYVÖNEN SmartMuseum knowledge exchange platform for cross-European cultural content integration and mobile publication	107
<i>Parallel sessions II</i>	
<i>Sustainable policies for digital culture preservation</i>	112
Maria GUERCIO Introduction to the session: conceptual framework and chain of custody for sustaining the digital preservation	113
Miquel TERMENS, Mireia RIBERA, and Alice KEEFER Does "long-term preservation" equate to "accessibility forever"?	117
Sven SCHLARB, Andrew N. JACKSON, Max KAISER, and Andrew LINDLEY The Planets Testbed: a collaborative environment for experimentation in digital preservation	121
Jan HUTAŘ, Andrea FOJTU, Marek MELICHAR, and Bohdana STOKLASOVÁ Czech National Digital Library and long-term preservation issues	125
Friederike KLEINFERCHER and Kristina KOLLER Cultural Heritage: from the library shelves to network residents	130
Felix ENGEL, Claus-Peter KLAS, Holger BROCKS, Alfred KRANSTEDT, Gerald JÄSCHKE, Matthias HEMMJE Towards supporting context-oriented information retrieval in a scientific-archive based information lifecycle	135
Maurizio LANCIA, Brunella SEBASTIANI, Roberto PUCCINELLI, Marco SPASIANO, Massimiliano SACCONI, Luciana TRUFELLI, Emanuele BELLINI, Chiara CIRINNÀ, Maurizio LUNGI Towards a European global resolver service of persistent identifiers	140



<b>Jeremy W. HUNSINGER</b>	<b>146</b>
Where did the user's go? A case study of the problems of event driven memory bank	
<b>Thomas RISSE, Julien MASANÈS, András A. BENCZÚR, Marc SPANIOL</b>	<b>150</b>
Turning pure Web page storages into living Web archives	
<b>Sam COPPENS, Erik MANNENS, Tom EVENS, Laurence HAUTTEKEETE, and Rik VAN DE WALLE</b>	<b>156</b>
Digital Long-Term Preservation using a Layered Semantic Metadata Schema of PREMIS 2.0	
<b>Daniel TERUGGI</b>	<b>159</b>
PrestoPRIME: Keeping digital content alive	
<b>Raffaele CIAVARELLA, Alain BONARDI, Guillaume BOUTARD</b>	<b>162</b>
Virtualization of real time audio processes: towards a musical notation of contemporary music	
<b>Dennis MOSER</b>	<b>166</b>
Second looks at Second Life: Considerations on the Conservation of Digital Ecologies	
<b>Bernard SMITH</b>	<b>169</b>
Conclusions and report from the parallel sections	
 <i>Papers accepted</i>	 <b>178</b>
<b>Maristella AGOSTI, Nicola FERRO, and Gianmaria SILVELLO</b>	<b>179</b>
Enabling Cross-Language Access to Archival Metadata	
<b>Paolo Budroni</b>	<b>184</b>
Rethinking how to shape a new matrix for the protection and retention of cultural heritage	
<b>Cèsar CARRERAS, Federica MANCINI, and Group ÒLIBA</b>	<b>188</b>
Improving cultural web production in small-size institutions: strategies to promote heritage in a productive basis	
<b>Susanna CHOULIA-KAPELONI, Jenny ALBANI, Polyxeni BOUYIA, Michelle KONDOU, Charikleia LANARA, and Sofia TSILIDOU</b>	<b>193</b>
E-wandering through the glories and vicissitudes of the Roman Agora and Hadrian's Library at Athens	
<b>Alida ISOLANI, Dianella LOMBARDINI, Claudia LO RITO, Daniele MAROTTA, and Cinzia TOZZINI</b>	<b>198</b>
Empowering users without weakening digital resources: is this possible?	
<b>Thomas KIRCHHOFF, Werner SCHWEIBENZ, Jörn SIEGLERSCHMIDT</b>	<b>201</b>
BAM - A German portal for cultural heritage as a single point of access for users	
<b>Lauri LEHT</b>	<b>205</b>
Involving users in the content enrichment process of digitized archival material	
<b>Franco LIBERATI, Maria Teresa TANASI</b>	<b>211</b>
Optical Supports for Digital Preservation - Problems and Prospects	
<b>Luciana GUNETTI, Eleonora LUPO, Francesca PIREDDA</b>	<b>216</b>
Designing digital formats for cultural production and exploitation: from accessibility to use value	
<b>Christoph MÜLLER, Anna WEYMANN, Bertram NICKOLAY, and Rodrigo Luna OROZCO DE ALENCAR</b>	<b>220</b>
Digitisation of Library Material: Caught Between User Demands and Preservation?	
<b>Laura PECCHIOLI, Fawzi MOHAMED, and Marcello CARROZZINO</b>	<b>225</b>
ISEE: retrieve information in Cultural Heritage navigating in 3D environment	
<b>Satellite Events</b>	<b>229</b>



*Empowering users: an active role  
for user communities*

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## Welcome

The Ente Cassa di Risparmio of Florence represented by myself today is delighted to attend this second edition of the international conference: Cultural Heritage On-line. Empowering users: an active role for user communities, well organized by the Fondazione Rinascimento Digitale and promoted jointly with two of the most important institutions in the survey of the Italian and foreign cultural heritage: the Ministry for Cultural Heritage and Activities, and the Library of Congress.

The conference has involved the participation of several national and international supporters and I would like to thank all of them for the excellent work of collaboration and the extraordinary attendance registered.

The Fondazione Rinascimento Digitale is a Foundation promoted by the Ente Cassa di Risparmio of Florence, that carries out and will carry out more and more a constructive support action to optimize and direct projects concerning the field of new technology for the cultural heritage. We are very glad to see that the Foundation, as years go, has won a high regard and a strong credibility from the most important representatives in the cultural field at the national and international levels.

In conclusion, I would like to thank all the people attending the conference for the great enthusiasm and interest they have showed, and finally I wish to extend my sincere thanks to the Teatro della Pergola for its fine hospitality and helpfulness.

***Michele GREMIGNI***

**President**

**Ente Cassa di Risparmio di Firenze**



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## Preface

The second edition of the conference, Cultural Heritage online – Empowering users: an active role for user communities, has been successfully organised in Florence on 15-16 December 2009 by the Fondazione Rinascimento Digitale – FRD in cooperation with the Italian Ministry for Cultural Heritage and Activities (General Direction for Archives and General Direction for Libraries, Cultural Institutes and Authors' Rights), and with the Library of Congress that mobilized all the network of the National Digital Information Infrastructure and Preservation Program - NDIIP partners.

We warmly thank our promoters, the local and regional authorities, and a fantastic group of thirty-five supporters who really were the engine and petrol in spreading information about the event within their user communities, involving cultural heritage institutions and professional associations, technology providers and research centres. A special thanks also to the Director of the Teatro della Pergola for his marvellous hospitality and technical support, and finally to the Palazzo Borghese that was the incredible venue for the gala dinner.

The majority of the success has been possible thanks to all the speakers and chairpersons, as well as to the FRD staff, very limited in number but determined and motivated.

The FRD is a structural foundation promoted by the Ente Cassa di Risparmio di Firenze to support the best adoption of ICT and international standards with a special attention to Internet technologies and digital preservation. The Foundation operates in all of the sectors connected with the production, preservation and diffusion of digital memories by means of research projects and test-beds, management and preservation of digital memories, implementation of innovative applications, and dissemination of this knowledge, with tutorials or on-line training courses.

The FRD invests its annual budget on research projects or prototypes in synergy with the main cultural actors, promotes awareness and training about digital libraries and digital preservation with special attention to young generations.

The main interests and activities of the FRD are in two fields: digital libraries for the cultural and humanities sector, and all the issues related to digital preservation. In the first group, the Foundation has promoted a project to set up a user community about humanities & computing, and has developed an open-source software "Pinakes" to manage digital objects in a very advanced way. Concerning digital preservation, the Foundation is very active and participates in international arenas, for example with the Digital Preservation Europe – DPE coordination action, and with the "Digital Stacks" project, coordinated by the Central National Library of Florence - BNCf, about reliable digital repositories paving the way to the first test-bed for the national legal deposit of digital contents on behalf of the Ministry of Culture in Italy. The Foundation has also promoted an innovative architecture for National Bibliographic Number - NBN persistent identifiers for digital contents, and thanks to the BNCf and the Consiglio Nazionale delle Ricerche – CNR, they have developed jNBN an open-source software already requested by other countries for evaluation. Finally, the Foundation participates with an expert in the PREMIS Editorial Committee managing the development of the standard under the coordination of the Library of Congress. By these three projects, the Foundation attempts to face the challenge of digital preservation with a global and complete strategy.

This year, the conference topics were again related to digital libraries, digital preservation and how Internet is changing scenarios and paradigms, but focussing on the user needs and point of view; so instead of investigating the technology offer, we favoured the works about organisational issues, new Internet schemas and roles, user requirements and constraints, and cultural and economic limits too, that can be seen as obstacles in the way of completely adopting ICT in the cultural heritage sector.

In particular, the first day of the conference proposed eight invited lectures that investigated user needs and expectations, analysing how to better involve users and the cultural heritage community in creating and sharing digital resources. Other topics evaluated were the current trends and use of interactive Web 2.0 tools by cultural institutions, benefits and future opportunities but also limits and constraints for users to create new contents, or possible risks trusting too much any info available on Internet. Some basic concepts underpinning dematerialisation of traditional archives have been presented. The need to choose what digital contents we want to select for long-term preservation, also involves some ethical issues. One important message was that the challenge of the future information society hinges on the use through cooperation among all the sectors, and in particular the current development for Archives, Libraries and Museums - ALM sounds promising.



*Empowering users: an active role  
for user communities*

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

The plenary session of the second day conference was started with the presentation of national and international scenarios, followed by two thematic sessions with scientific speeches selected through a Call for Papers that observed the advancement of the research on the user-institution relationship towards the development of cooperative Web 2.0 tools and on sustainable digital preservation policies.

The scientific program was enriched by two important training opportunities the day before and the day following the conference: the Tutorial “Long Term Preservation of digital assets: basics, concepts and practices” and the Tutorial “Dublin Core - Building blocks for interoperability”.

The conference has seen a great interest not only among the specialists in the cultural heritage field, but there was also a large participation of Italian and foreign students, several European Projects, representatives of local administrations, research centres and the private and corporate sector. There were about 400 people at the conference: one-fourth of those present was formed by students; as for nationality, there were people from all countries: in particular three-quarters of participants were Italians, followed by Americans, Germans, Dutchmen, Estonians, and a varied representation of almost all of the European countries.

In conclusion, many activities have been carried out since the first edition of the Cultural Heritage on line conference in 2006, the Fondazione Rinascimento Digitale has developed contacts and cooperation with cultural institutions worldwide, and we feel confident to be able to work successfully with you all before the next conference in 2012. We'll keep in contact with you before that.

Thanks again to everybody!

***Maurizio LUNGI***

**Scientific Director  
Fondazione Rinascimento Digitale**



*Empowering users: an active role  
for user communities*

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

**Tuesday 15th December, Plenary Session**  
*Invited lectures*



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## **Laura CAMPBELL**

### **Collaboration and interaction on the Web at the Library of Congress**

As we think back to Thomas Jefferson's initial contribution to the Library of Congress, we are reminded that technological progress did not begin with the digital era. But it has been through the continued refinements of digital technologies over the past 20 years that the Library has been most able to make its treasures widely accessible and to share its resources with a national and international audience.

The Library of Congress's embrace of digital technologies is a natural extension of its commitment to innovative access and use of the Library's unique collections. Just as the initial investment in the Library has paid off in myriad ways, so have the ongoing strategic investments in information technology proven essential as a way to bring scholarship to the information leaders of the next century.

As we close the first decade of the 21st Century, the Library is cognizant of its need to continually refine its approaches to engage with an interactive audience fluent with the Web and computing technologies. This refinement is nothing new, as the Library has consistently been a pioneer in identifying these technologies and utilizing them to share its wealth of resources.

#### **1990s Strategy: Increasing Accessibility**

The Library of Congress is engaged in a constant process of technological refinement in order to continually maximize the value and accessibility of its collections, but the full embrace of collaborative and interactive digital technologies essentially began in the 1990s.

In 1990, the Library began a pilot called American Memory to digitize some of the Library of Congress's unparalleled collections of historical documents, moving images, sound recordings, and print and photographic media; items that make up the "nation's memory."

By 1994, as the Internet transformed the presentation and communication of human knowledge, the Library established the National Digital Library Program, a systematic effort to digitize the historical treasures in the Library and make them readily available on the Web to Congress, scholars, educators, students, the general public, and the global Internet community.

The NDLP has been an unqualified success, celebrating its 10 millionth digitized item in American Memory (<http://memory.loc.gov/ammem/index.html>) in 2004, and continuing to expand online historical content as an integral component of the Library's commitment to harnessing new technology to fulfill its mission "to sustain and preserve a universal collection of knowledge and creativity for future generations."

The iconic materials that comprise American Memory remain highly popular today. Strong interest in the American Memory materials resulted in development of The Learning Page (<http://memory.loc.gov/learn/>). The Learning Page promotes use of primary sources in the classroom and acts as a teacher's "front door" to Library holdings that enrich teaching experience.

Additionally, the Library worked to make the legislative workings and daily bills of the U.S. Congress more accessible through the establishment of the Thomas website (<http://thomas.loc.gov/>) in January of 1995.

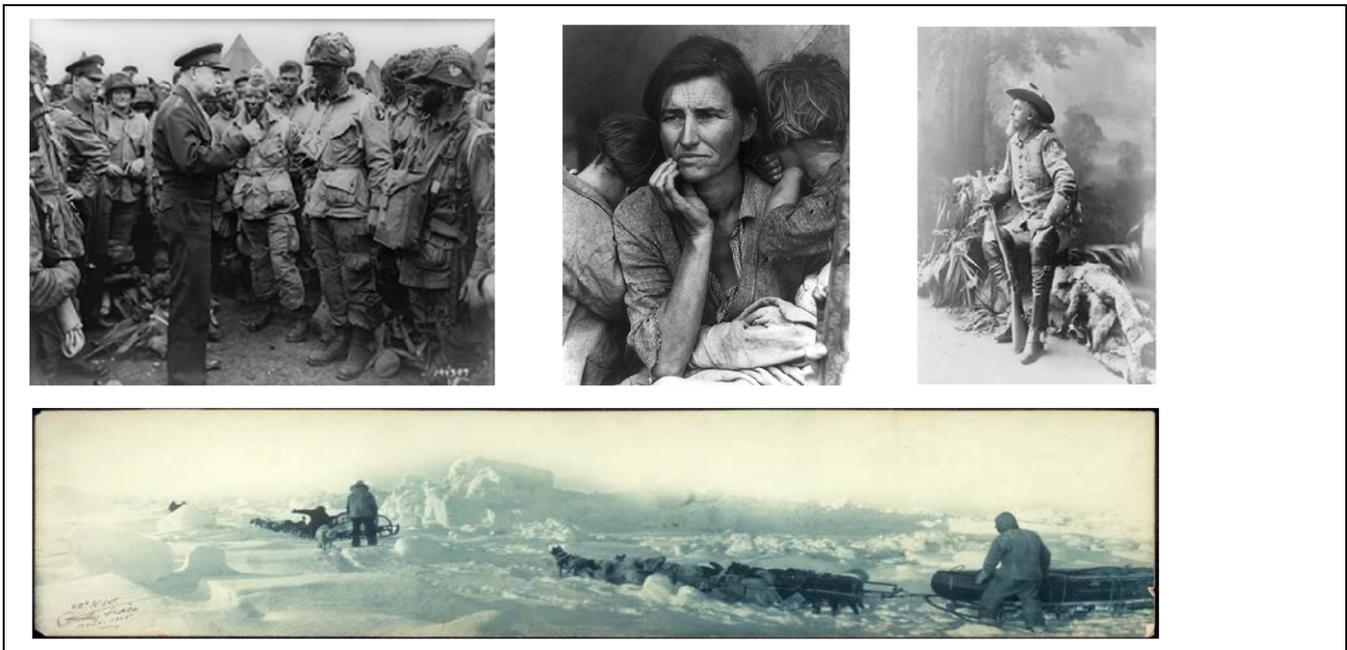


Figure 1: Photos from American Memory (<http://memory.loc.gov/ammem/index.html>)

### Strategies for the 21st Century: Engaging the Interactive Audience

The Library entered the 21<sup>st</sup> century aware of the need to reimagine its traditional roles and responsibilities for new audiences. Building on its earlier digital initiatives, the Library's current digital activities reflect the increasing importance of interactivity in the user experience. These initiatives embrace the core concepts that have come to be known as Web 2.0, with the goal to facilitate interactive information sharing through interoperability and collaboration.

These activities are freeing the Library's wealth of digital content, making it more accessible to a broader community than ever before, both at Loc.gov and in the social networks and Web convergence points such as YouTube, Facebook, iTunes, Flickr, and Twitter where users now gather. The Library's ultimate goal is to facilitate a deep engagement with its content through actions (such as linking, embedding, tagging or rating) that let users interact with the content in the manner of their choosing.

The Library's Flickr pilot project ([http://www.flickr.com/photos/library\\_of\\_congress/](http://www.flickr.com/photos/library_of_congress/)) provides an illustration of how exposing material through new channels can increase the use of content in new and innovative ways. The project was launched in January 2008 with an invitation to the public to comment on and describe approximately 3,000 historic photos. This opportunity to interact with the Library's collections struck a chord with newly engaged users.

In the first 10 months of the project there were 10.4 million views of the photos on Flickr, resulting in enhancement to more than 1,000 Library of Congress Prints and Photographs Online Catalog records with new information provided by the users. Moreover, the average monthly visits to the PPOC rose 20 percent over the five month period of January-May 2008, compared to the same period in 2007.



Empowering users: an active role for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

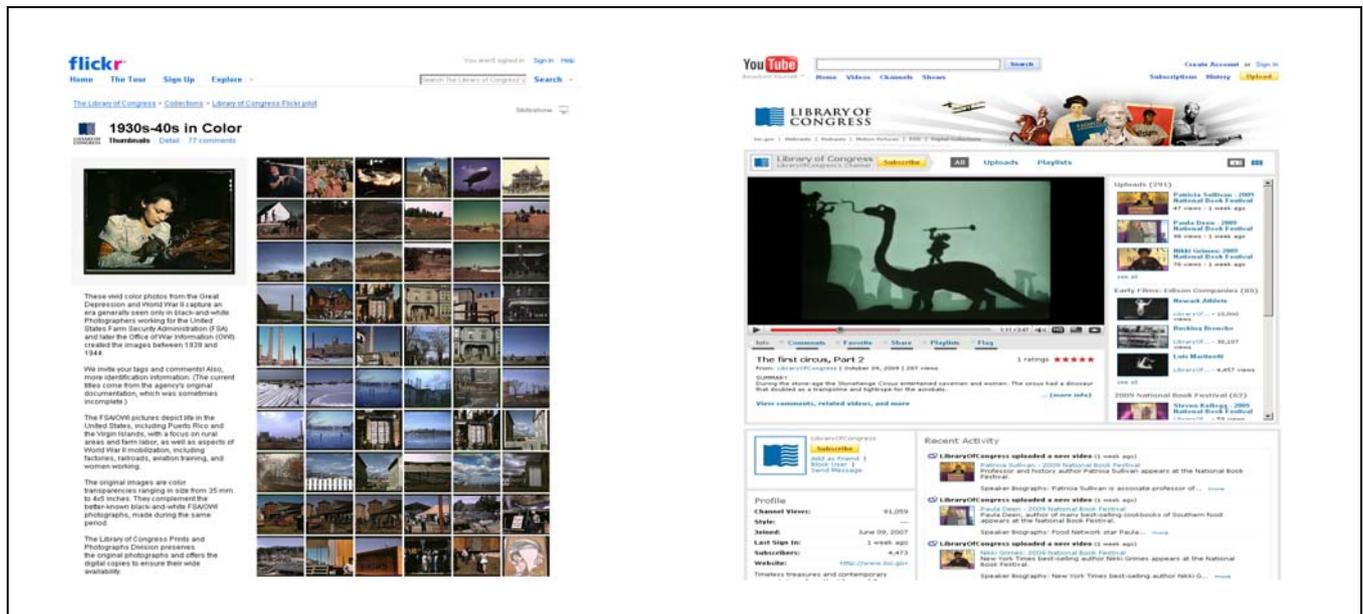


Figure 2: Library of Congress Flickr ([http://www.flickr.com/photos/library\\_of\\_congress/](http://www.flickr.com/photos/library_of_congress/)) and YouTube projects (<http://www.youtube.com/user/LibraryOfCongress>)

The success of the Flickr project inspired the Library to engage in content distribution agreements with Yahoo and YouTube that expose valuable Library content through their channels.

These agreements are models of the way in which government can leverage private sector efforts while ensuring universal access to authoritative content. Any agreement to place content on a non-Library of Congress site must be non-exclusive, and access to the Library's content must be without cost. Notwithstanding advertisements that might appear on various search results pages, an option to control or entirely exclude advertising on the account in close proximity to the Library's content is preferred. The Library also must be clearly identified as the source of the content, and pages with Library content are branded both graphically and through account naming conventions.

The content used in these various pilots is already available on the Library's Web site, ensuring the public can be confident that the material will remain accessible regardless of the long-term viability or future access policies of the non-Library site.

The Library of Congress online video portal on YouTube (<http://www.youtube.com/user/LibraryOfCongress>), established in April 2009, offers select items from the Library's collections of early motion pictures, along with recordings of Library-sponsored events, lectures and concerts. Through this pilot, the Library - home to over 1.2 million film, television, and video items - is sharing items from its collections with users who enjoy video but may not be aware of the extensive video resources available on the Library's own Web site.

The reach of these new distribution channels is exemplified by the usage statistics of a single video posted to YouTube. The *Rosie the Riveter: Real Women Workers in World War II* video received over 5,000 views on YouTube during April 2009, the first month of the Library of Congress channel launch, and has since garnered over 20,400 views in its first year. It took more than 5 years for the video to reach that level of viewership on LOC.gov. Over 60 percent of the YouTube viewers reached the video through an embedded clip on the Wired.com blog noting the launch of the Library's YouTube channel, demonstrating the potential reach of authorized content sharing.

### Network Strategies: Leveraging Experience

In addition to exploring new channels of content distribution, the Library of Congress has engaged with emerging networks of diverse partners. Network strategies allow each participant to leverage the experience, resources and expertise of the entire network community. It has been the Library's experience that networks of collaborating partners increase innovation. The Library has also identified the following benefits to pursuing a network-focused strategy:



- Jointly produced or shared network infrastructure capacity
- Shared product and service innovations and dissemination
- Collective adoption of policy, standards, and best practices
- Increased efficiency in joint problem identification and pursuit of solution paths
- Amassed expert knowledge and knowledge sharing
- Increased efficiencies of pooled capital resources and their distribution
- Increased capacity to target and influence public policy decision makers
- Increased capacity to attract investments in digital preservation

In 2000 the National Digital Information Infrastructure and Preservation Program was created to develop a national strategy to collect, archive and preserve the burgeoning amounts of digital content for current and future generations, a logical outgrowth of the Library's historic mission to "sustain and preserve a universal collection of knowledge and creativity for future generations."

NDIIPP has deeply engaged the Library in collaborative partnerships, with the understanding that digital stewardship on a national scale depends on public and private communities working together. To that effect, the Library has built an ongoing preservation network of over 150 partners from across the nation to tackle the challenge of preserving and providing enhanced access to digital content of importance to the nation.

This approach encompasses a new catalytic role for the Library, with NDIIPP's strategies for success exemplifying the Library's new roles within developing networks of likeminded partners:

- Take early action
- Learn by doing
- Be catalytic
- Act collaboratively
- Engage Multiple platforms
- Leverage Natural, Distributed Networks
- Share Resources

No single organization has the resources and the ability to preserve all the data needed by policy-makers, planners, and scholars. NDIIPP is leveraging the expertise and existing work being accomplished by federal agencies, universities, the private sector, and not-for-profit organizations. By working together, the network can identify and maintain the content, including geospatial data, which will allow complex, temporal analysis of historical data in the future.

Another model of networked collaboration in which the Library of Congress participates is the International Internet Preservation Consortium (IIPC). The IIPC enables the collection of a rich body of Internet content from around the world to be preserved in a way that it can be archived, secured and accessed over time. The group fosters the development and use of common tools, techniques and standards that enable the creation of international archives. In addition to the collaborative work within the consortium, the IIPC also encourages and supports national libraries everywhere to address Internet archiving and preservation.

The twin goals of expanding online offerings through new channels while also embracing networks of diverse partners come together in the World Digital Library, a project being developed with the cooperation of the United Nations Educational, Scientific and Cultural Organization and 32 other partner institutions.

With 55 partners in 35 countries as of October 2009, the WDL makes available on the Internet, free of charge and in multilingual format, rare and unique primary source materials that reflect the history and cultural heritage of the participating countries. The WDL promotes international and inter-cultural understanding and awareness, increases the diversity and volume of cultural content on the Internet, provides resources for educators, scholars and general audiences, and builds capacity in partner institutions in ways that will narrow the digital divide within and between countries.



Empowering users: an active role for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009



Figure 3: The World Digital Library (<http://www.wdl.org/en/>)

### Collaboration and Interaction: Strategies for the Future

These recent collaborative initiatives have increased the volume of content, the variety of formats, the kinds of user communities, the personalization of the content and the ability of the Library of Congress to influence and support scholarship throughout the world.

Future initiatives will explore the how the Library acquires and preserves content that exists in disparate locations around the world, brought together through new collaborative methods of creation. The “crowdsourcing” of news reporting is one example of how the traditional channels of communication are changing as digital technologies enhance user’s ability to both create and consume content. The Library will also be called upon to support new, collaborative methods for learning, education and research.

The technological environment is rapidly changing, and the Library is changing with it, continuing to develop new strategies to showcase its collections based on leveraging opportunities for collaboration and interaction where they exist while grappling with the exponential growth in content and the concomitant increasing user engagement.

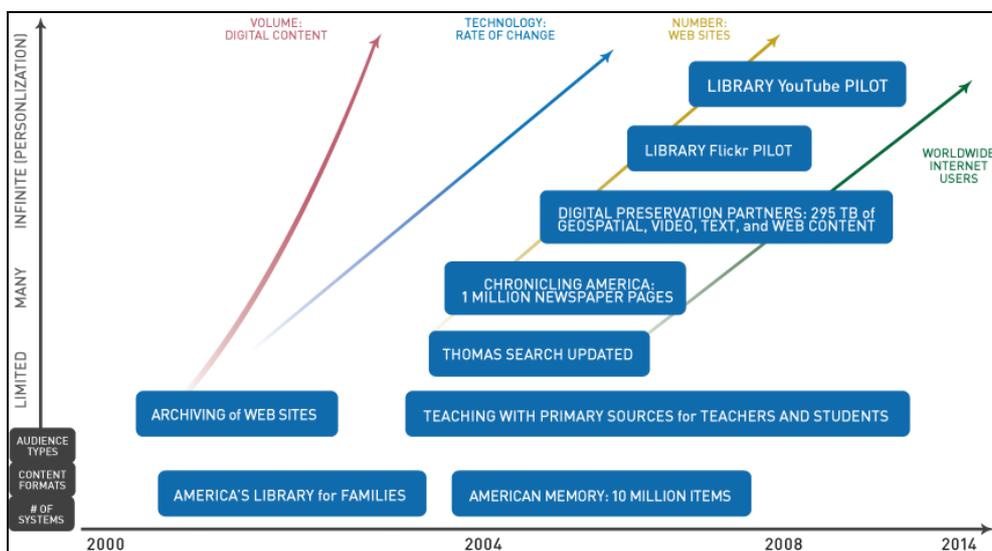


Figure 4: The Library of Congress's Digital Content and Interactivity

**Luciana DURANTI**

## **The Long-term preservation of digital heritage - The Case of Universities Institutional Repositories**

Digital preservation can be defined as the whole of the principles, policies, rules and strategies aimed at prolonging the existence of a digital object by maintaining it in a condition suitable for use, either in its original format or in a more persistent format, while protecting the object's identity and integrity, that is, its authenticity.<sup>1</sup> One might ask why authenticity is an issue for all kinds of digital objects. While it is obvious that authenticity is a necessary requirement in the preservation of records, the value of which as sources of evidence resides in their trustworthiness, why would it be for other types of digital heritage, such as publications, works of art, or games? Although a digital entity that does not qualify as a record is not conceptually linked to the idea of trust, it still needs to have an identity that is certain and indisputable, and its manifestation must be stable, always equal to itself, intact. And here lies the problem, because, traditionally, these qualities reside in the original, the perfect entity, the first complete item to be issued, released or made public (be it a unique one or one of many), or in an authentic copy of the original generated by a person with the authority to do so. However, as the concept of original disappears in the digital world, where can we look for the assurance that what we observe is what it claims to be?

We have no other choice than make inferences based on a variety of circumstances, but primarily on the integrity of the environment in which the digital entities in question reside and of the processes aimed to maintain them and to ensure the accountability of the person or organization responsible for them. This means that institutions must create mechanisms that allow for the determination of authenticity based on the trustworthiness of the source of the digital entities and the chosen method of their transmission through time, and then adopt the necessary strategies to preserve them in a sustainable way.

The selected preservation methodology must allow the preserved entities to continue to be readable and useable regardless of any technological changes to the underlying hardware or software environments, and the preserving organization to account for these changes, so that the entities may continue to be migrated to newer platforms as needed to avoid technological obsolescence.

To illustrate the issues digital repositories of cultural heritage will have to deal with, I am going to discuss an InterPARES 3<sup>2</sup> Case Study called "cIRcle." cIRcle is a digital repository for the management, dissemination and preservation of the intellectual output of a university and its community members.

A university institutional repository is defined as "a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its members."<sup>3</sup> Although their creation has been predicated in large part on the requirement of making available to the public research products that have been developed with the support of public money, and on the benefit to the university of showcasing its research, institutional repositories (hereinafter IR) have emerged in North America and Western Europe primarily because they are regarded by the university communities as a means of having access to products of scholarship and research, and as a locus for preserving such products and maintaining access to them over the long term.<sup>4</sup> Therefore, in the past few years, the IRs have accumulated not only preprints and post-prints of articles, books, theses and dissertations, but also raw data files resulting from research, working papers, course syllabi, class notes, handouts, students' papers, committee meetings agendas and minutes, unpublished conference presentations and several other types of documentation that fall under the category of personal and university *records*, not only *publications*, and are preserved by the university archives and special collections. This mix of documentation and data creates severe challenges to the IR's continuing access and preservation (the very reasons why they exist) from several points of view.

The InterPARES research project has demonstrated that it is not possible to preserve digital materials, but only the ability to reproduce them. Reproduction involves different activities at different times in the life of the

1 See the InterPARES Project Terminology Database at [http://www.inter pares.org/ip2/ip2\\_terminology\\_db.cfm](http://www.inter pares.org/ip2/ip2_terminology_db.cfm).

2 InterPARES 3 is the third phase of the InterPARES Research Project (1999-2012). While the first two aimed to develop theory and methods of authentic digital long term preservation, the third is testing the findings of the previous two in real situations. See [http://www.inter pares.org/ip3/ip3\\_index.cfm](http://www.inter pares.org/ip3/ip3_index.cfm).

3 Clifford A. Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," ARL Bimonthly Report 226 (2003): 1-7, available at <http://www.arl.org/bin-doc/br226.pdf>.

4 See Clifford A. Lynch and Joan K. Lippincott, "Institutional Repository Deployment in the United States as of Early 2005," D-Lib Magazine 11 (September 2005), available at <http://www.dlib.org/dlib/september05/lynch/09lynch.html>; and Gerard van Westrienen and Clifford A. Lynch, "Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005," D-Lib Magazine 11 (September 2005), available at <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>.

material. In the initial few years, it may consist simply of retrieving and reassembling the digital components that constitute the object to generate a copy or, if the object is technologically complex, as in the case of interactive and/or dynamic documents from the visual and performing arts and from the sciences, it may require its *re-creation*. However, when the digital format becomes obsolescent, it is necessary to either migrate the digital object to a newer technology by changing its architectural structure or, in some cases, to emulate the behaviour of the old technology to access the object. Regardless, throughout the existence of the object, ongoing copying and transformative migration<sup>5</sup> are required for reasons of security (which is based on redundancy) and of continuing access. These activities raise several issues, among which the paramount ones are those of authenticity and intellectual rights. The authenticity of digital material is dependent upon the maintenance through time of its identity and of its integrity<sup>6</sup>. The intellectual rights of the copyright owner are attached to the authentic version of the digital object and, specifically, to its documentary form, which is defined as the rules of representation that govern the expression of the ideas of the author in a stable and fixed manner.<sup>7</sup>

Intellectual rights comprise several types of rights, but among them the ones that are affected by long-term preservation by means of constant transformative migration or emulation are the two major groups of intellectual rights: economic rights and moral rights. Economic rights are those that enable the copyright owner (not necessarily the author or creator) of a work to make commercial gain from the exploitation of that work.<sup>8</sup> Moral rights are those rights that the author or creator retains (regardless of whether the author still retains the economic rights) over the integrity of a work (rights of reputation) - such that no one, even the copyright owner, is allowed to distort, mutilate or otherwise modify the work in a way that is prejudicial to the author's honour or reputation; the right to be associated with the work as its author by name or under a pseudonym and the right to remain anonymous (rights of attribution); and the right to refuse to allow the work to be used in association with a product, service, cause or institution in a way that is prejudicial to the author's honour or reputation (rights of association).<sup>9</sup>

A recent census of college and university IRs in the United States has found that 70.8% of them do not have a policy for licensing content. In addition, there is no mention in the literature concerning IRs of the issue of authenticity through time, and none of them appears to have strategies in place for long-term preservation.<sup>10</sup> This is probably due to the fact that it is the belief of those who manage an IR that its content exists somewhere else, which is a safe presumption for preprints and post-prints, but certainly not for all those digital objects that are unique and often qualify as records, such as the official copy of theses, professors' and students' papers, etc.

Given the situation described above, it is necessary:

to identify in which way digital preservation strategies as recommended by the major international research projects on the subject may infringe existing intellectual rights (economic and moral) legislation as it applies to published and unpublished material;

to establish what long-term preservation measures would be possible in the context of the existing legislation and to test them on IRs in course of development to assess their impact on the continuing authenticity and accessibility of the digital material; and to determine what changes to the law are required to ensure that the proper long-term digital preservation strategies can be applied so that the research output of universities can remain attributable and accessible in its authentic form for as long as needed.

In order to do so, the InterPARES 3 project has selected as a case study an institutional repository called cIRcle, at the University of British Columbia (UBC)<sup>11</sup>. As stated in the brochure publicizing it to the UBC faculty and students, cIRcle assembles various communities and collections. Communities are UBC departments, labs, research centres, schools and other administrative units. Within cIRcle, each community oversees one or

5 Transformative migration is defined as "The process of converting or upgrading digital objects or systems to a newer generation of hardware and/or software computer technology" (InterPARES 2 Project, "Terminology Database: Glossary," available at [http://www.interpares.org/ip2/ip2\\_terminology\\_db.cfm](http://www.interpares.org/ip2/ip2_terminology_db.cfm)). The effects of transformative migration of the digital materials in an institutional repository are an important consideration insofar as any new additions or modifications to an existing work (even a work already in the public domain) can trigger new copyright considerations.

6 Luciana Duranti (ed.), *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* (San Miniato, Italy: Archilab, 2005).

7 Luciana Duranti and Kenneth Thibodeau, "The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES," *Archival Science* 1 (2006): 13-68.

8 Michael O'Hare, "Copyright and the Protection of Economic Rights," *Journal of Cultural Economics* 6(1) (1982): 33-48.

9 Mira Sundara Rajan, "Moral Rights in Information Technology: A New Kind of 'Personal Right'?" *International Journal of Law & Information Technology* 12(1) (2004): 32-54.

10 Elizabeth Yakel, Soo Young Rieh, Beth St. Jean, Karen Markey and Jihyun Kim, "Institutional Repositories and the Institutional Repository: College and University Archives and Special Collections in an Era of Change," *The American Archivist* 71(2) (2008): 344.

11 See <https://circle.ubc.ca/>.

more of its own collections, which contain items submitted to the IR. As currently envisioned, cIRcle's operational goal is to be able to accept, preserve indefinitely and provide continued readability and accessibility to virtually all published and unpublished digital objects created in any file format by or on behalf of the University, its faculty, staff or students - including preprints and post-prints of academic journal articles, other items such as theses, dissertations, departmental publications, technical reports, bulletins, conference proceedings, course notes and other learning objects, and raw research data. cIRcle has yet to develop, articulate and implement a maintenance plan that addresses this ambitious goal, and it has not attempted yet to address the issue of the protection of economic and moral rights in the context of long-term preservation. This situation makes of it the ideal candidate for the development of a preservation strategy for IRs that is sensitive to intellectual rights issues and for the testing of such a strategy.

As of November 6, 2009 there were 14,073 items in cIRcle totaling 130GB. This material is stored in DSpace, which is a database with a set of services to capture, store, index, maintain and make accessible a variety of entities in a digital format over the internet utilizing a controlled set of workflows and access permissions. Dspace is an open-source application, freely accessible at [sourceforge.net](http://sourceforge.net), one of the largest open source software repositories on the net. It is written in Java, providing broad based support and compatibility with a broad base of internet browser; for a database back end it uses either Oracle – the industry leader – or Postgres, an open source relational database. The fact that DSpace is an open source application is good because of the authenticity issue.

It was stated earlier that the preservation of authenticity, to which all intellectual rights are linked, requires the protection of the identity and integrity of the material. Identity is not difficult to maintain overtime if the appropriate set of metadata is attached to the various digital entities and kept inextricably linked to them. Integrity is problematic not just to protect, but also to prove, because one has to rely on the integrity of the environment in which the entities reside. It is very hard to assess the integrity of an environment that is proprietary. In contrast, open source satisfies the legal requirements of objectivity, transparency, verifiability and repeatability for any process that is carried out in a digital environment<sup>12</sup>.

In DSpace, the records themselves are embedded in a hierarchical folder structure based on the collection. Contained within each folder is the original bitstream, a full text extract of the contents (used for searching), a thumbnail of the record for web presentation, and copyright information on the record. DSpace uses Preservation Services modules to verify the integrity of the stored files (utilizing a checksum to look for file corruption or alteration) and media filters to define file conversations. Records are accessed through the web via a persistent web address that allows researchers to link directly rather than having to use a database search every time. The bitstream format contains information on how the material in that format is to be interpreted, allowing for control and granularity. For example, .doc may refer to more than one version of word, each of which presumably has different characteristics and functionality. Each bitstream format also has support level associated with it indicating how likely it is that the format will be accessible into the future given the toolset currently available to the system<sup>13</sup>.

Digital forensics experts value open source, which, at the same time, allows modification and encourages dissemination, thereby making it possible to submit the software together with the digital entities presented as evidence, so that their accuracy can be tested promptly by anyone at any time. This is especially true when conversion or migration occurs, because it would allow a practical demonstration that the software could not simultaneously manipulate the content of the files while copying them and that nothing could be altered, lost, planted, or destroyed. Finally, open source is preferred because of the possibility of exchange of evidentiary material between the parties in the course of e-discovery.

Why should we care about issues of evidence and discovery? Because it is more than certain that, if an author feels that his or her intellectual rights have been infringed by the preservation measures taken by the institutional repository, he or she will want to see the issue solved in court. Undoubtedly, cIRcle procedures are in conflict with copyright legislation also regardless of preservation methods. This is because one has to consider that acting within copyright is different from policing copyright. Items are generally posted to cIRcle by

12 See for example Brian Carrier, "Open Source Digital Forensics Tool. The Legal Argument," available at [http://www.digital-evidence.org/papers/opensrc\\_legal.pdf](http://www.digital-evidence.org/papers/opensrc_legal.pdf), p. 7; and Erin Kenneally, "Gatekeeping Out Of The Box: Open Source Software As A Mechanism To Assess Reliability For Digital Evidence," *Virginia Journal of Law and Technology*, 6 Va. J.L. & Tech. 13 (2001). Available at <http://www.vjolt.net/vol6/issue3/v6i3-a13-Kenneally.html>, para. 34-35. See also Andrea Ghirardini and Gabriele Faggioli, *Computer Forensics* (Milano: Apogeo, 2007), p. 230.

13 The three levels are Supported, Known and Unsupported. Supported are those file types that the institution has reasonable level of confidence that they have the tools and/or techniques available to progress the files through future technology changes. Known formats are those that are recognized by the institution and attempts are being made to create or obtain the tools necessary for future migration/access. Unknown file formats are those that will be preserved at the bitstream level only; it will be up to the researcher to obtain the software/tools necessary to view the files.

the author of a digital entity or the author's representative. Each submission requires the depositor to authenticate his/her authority to submit this work. cIRcle staff don't have enough time to verify copyright ownership for each item submitted, so cIRcle has to rely on the declaration of the depositor in order to operate. This is an act of faith, but it is a necessary one. Provided that cIRcle removes work upon notice of an infringement, and provided that cIRcle did not publish the infringing work knowingly, cIRcle should be protected from prosecution. Another issue has to do with materials that are scanned and uploaded to an IR. cIRcle's retrospective theses project involves the deposit of digitized theses and dissertations originally archived in print. It can be difficult to contact the authors of these items to obtain their permission to deposit. When authors cannot be contacted after due diligence in attempting to notify them, cIRcle may choose to proceed with publishing their work online. They do so assuming implicit permission by dint of the university's prevailing stewardship and provision of access to the item. Should the author request removal of the work from cIRcle after it has been published online, cIRcle remains obliged to honour the author's wishes and remove the work from its holdings.

But these are minor issues with respect to the problems presented for intellectual rights protection by the preservation strategies that cIRcle will have to adopt in the presence of a legislation that is still much behind the development of technology. cIRcle will have to begin separating the protection of the moral rights and that of the economic rights. To do so, its strategy will need to distinguish data integrity, which means that the content of the entities in the repository is not modified accidentally or intentionally during the regular maintenance and use activities, from duplication integrity, which means that the process of creating a duplicate of the data for preservation does not modify either intentionally or accidentally the form and composition of the original entity. This reproduction would either be based on the principle of non-interference, which involves a non-transformative conversion, or on the principle of identifiable interference, which means that the method used does alter the entities but the changes are identifiable<sup>14</sup>. The application of both principles, by ensuring the creation of authentic copies, would allow for the protection of moral rights, which cannot be renounced. As it regards economic rights, any preservation activity would infringe the law as it stands now, thus, the only solution at this time is to obtain the permission of the copyright owners.

The InterPARES Project has made a submission to the Commission of Industry Canada and the Department of Canadian Heritage responsible for updating the Canadian copyright act, requesting that specific attention be given to the problems presented by the long term preservation of authentic digital entities.<sup>15</sup> In the meanwhile, it is conducting an inventory of all the items in cIRcle to identify their nature and characteristics, content, current licence, attached digital rights management, etc. in order to develop an intellectual property policy and a preservation plan consistent with it. The research conducted on cIRcle and its results will be accessible on a dedicate web site named "University Institutional Repositories: Copyright and Long Term Preservation," accessible at <http://uir-preservation.org/><sup>16</sup>.

14 E. Casey, "Error, uncertainty and loss in digital evidence," *International Journal of Digital Evidence* 1. 2 (2002).

15 Copyright Consultation statement available at: <http://copyright.econsultation.ca>

16 For more information on copyright and moral rights as related to preservations see: Austin, G. W., "The Berne Convention as a Canon of Construction: Moral Rights after Dastar" *NYU Annual Survey of American Law* 61 (2005): 101-139; Burkitt, D., "Copyrighting Culture – The History and Cultural Specificity of the Western Model of Copyright," *Intellectual Property Quarterly* 2 (2001): 146-186; Christie, A., "Reconceptualising Copyright in the Digital Era," *European Intellectual Property Review* 17(11) (1995): 522-530; Davies, G., *Copyright and the Public Interest*, 2d ed. (London: Sweet and Maxwell, 2002); Dreier, T. K., "Adjustment of Copyright Law to the Requirements of the Information Society," *IIC* 29(6) (1998): 623-639; Dreier, T. K., "Authorship and New Technologies from the Viewpoint of Civil Law Traditions," *IIC* 26(6) (1995): 989-999; Frow, J., "Public Domain and Collective Rights in Culture," *Intellectual Property Journal* 13(1) (1998): 39-52; Gervais, D., "Canadian Copyright Law Post-CCH," *Intellectual Property Journal* 18 (2004): 131-167; May, C., "Why IPRs are a Global Political Issue," *European Intellectual Property Review* 25(1) (2003): 1-6; *New York Times Co. v. Tasini*, 533 U.S. 483 (2001).14; Padfield, Tim, *Copyright for Archivists and Users of Archives* (London: Facet Publishing, 2004).

Rigamonti, C. P., "The Conceptual Transformation of Moral Rights," *American Journal of Comparative Law* 55(1) (2007): 67-122; *Robertson v. Thomson Corp*, 2006 SCC 43.15; Sterling, J. A. L., "Philosophical and Legal Challenges in the Context of Copyright and Digital Technology," *IIC* 31(5) (2000): 508-525; Vaver, D., *Copyright Law, Essentials of Canadian Law Series* (Toronto: Irwin Law, 2000).

## **Daniel TERUGGI**

### **Ethics of Preservation**

It has been widely assumed that digitisation of analogue contents is the most effective process for preserving contents from their natural and inevitable physical decay. From a technical point of view, and independently of the risks incurred within the digital preservation process, this assumption is true since digital replication permits to reproduce contents without altering them. However, from a more strict ethical approach on the nature of the content to preserve, it is important to question the continuity of the original content through its technological transformations and to analyse the concept of content equivalence.

Within the domain of text, the possibility to separate the information from the physical object containing it has been part of a long known strategy to reproduce written material without any major losses (except copying errors). With analogue media and images (as well as sounds), the fact that the content is imbedded in the carrier conditions the quality of the information and creates new replication paradigms, difficult to analyse and quantify efficiently. Independently of procedure standards, which may guarantee the quality of the process, it becomes extremely difficult to evaluate what quality means, mainly for images and sounds whose initial quality may be considered as very poor regarding today's technology. The problem goes further beyond the digitisation process, when we encounter contents, which have gone through a time decay process that has deteriorated the original state of the content. A major question arises then: which was the original state of a decayed content? Should we try to recover the initial state or even improve it?

The discussion happens also in the Digital world, where migration processes bring small deformations which may be important through time; where transcoding is a necessary action to maintain contents accessible, however it may introduce artefacts that can be annoying with time.

### **Content representation**

Content representation is associated to the concepts of original and copy; the original is the work of a content creator (artist, author, or whatever name is applied) who is the originator of the work and claims its ownership (several authors may also be concerned plus other right-holders). A copy is a reproduction of the original, which conserves some or all of the features of the original content or object. A representation is a more or less accurate description of something; representations are used as equivalents of the original object even if the representation is very poor (a cinemascope film viewed on a small television set with transmission problems is still considered as viewing the film). Representations are very useful objects to help us deal with contents or to simplify the access to them.

The difficulty is that: in some cases the distance between the original and its representation may be clear for our perception (it is evident for all that it is not the same thing to go to the Louvre to see the Mona Lisa by Leonardo, than to see the image on a cookie package); however in other situations the difference is not appealing for our perception: a photocopy of a book is considered as being the same thing as the book, a highly quality degraded photography is considered as having the same value as the original photography itself. Depending on the domain originals or copies, or representations; don't have the same implication and value: if we take four types of cultural domains as Museums, Libraries, Archives and Audiovisual collections, the concept of original or copy has different implications:

- Museums tend to contain physical objects; each object has an author or several authors, which can be known or unknown. Often high value is associated to the objects, which are mainly visible in the Museum's site. In this context copies are representations of the objects that serve to inform the visitors about the object; however viewing the physical object is the main objective for the Museum.

Copies are considered reductions of the original and our perception identifies them as representations of it. There are also physical copies of objects, and they can be very tricky to identify, however this is outside of the scope of this analysis.

- Libraries tend to contain written material (books or manuscripts), the books are the carrier for information; however the information can exist independently of the book and be transmitted through copies. Since writing is a coding system for information, the copy is as good as the original (mistakes may be introduced in the process)

Publishing is based on the principle of making multiple copies of the same information providing the original information is not modified, whatever way the information is presented (or reformatted in shape, typography or font size)

- Archives contain collections of objects, which may include physical objects, written material, audiovisual elements, etc. The main characteristic is that objects are related together through a collection, which may contain many authors and an owner of the collection. The links among the objects are as important as the objects themselves and the concept of original is related to these links that give coherence to the collection. Within the collection each individual object has the same copying issues as those seen in other domains.

Parts of the archival material may be copied, however the main concepts are the integrity of the whole collection and the links among the objects. These links are a knowledge that is added to the material; copies can be made of the structure as well as the objects.

- Audiovisual collections contain media (film, video, audio) in which information is linked with the media ("the media is the message" as Marshall McLuhan said). The notion of author is highly variable (several right owners generally). The main characteristic is that the deterioration of the media will have as a consequence the deterioration of the information or content. Copying an original normally gives a highly similar equivalent of the original (with loss in the analogue world and theoretically no loss in the digital world).

A copy is a new entity of the work; however there may be slow and imperceptible changes and degradations. A strongly degraded copy is still considered as a representation of the original.

### **Living with physical decay**

Time is a perpetual threat for any kind of object: books, archives, media, physical objects decay slowly or faster and that is the fate of our physical world. Probably the most efficient way to conserve information is to regularly copy it on a new carrier until it starts decaying and the process of copying is started again. We have no original books from antiquity, however the writings of Plato or Aristotle are still available through the slow and tedious process of successive hand-made copies until the arrival of print. These copies probably introduced small or important errors, to which the process of translation should be added, enlarging the divergences with the original source. We can still access those writers and understand their words and ideas, and sometimes track the different versions and transcriptions of the work. This process, when applied to digital contents and even to analogue to digital transfer, is called migration and it is largely accepted that it is the best solution in order to keep digital contents accessible. As with book copying, it is important to keep track of all the migration actions in order to evaluate eventual deformations or degradations. Concerning physical objects, specific actions have to be undertaken to minimize the action of time regarding their integrity. Restoration has to be done when parts of it fall apart and specific temperature and humidity as well as light environments have are foreseen in order to reduce the action of elements on them.

Preservation actions have to be organised and planned because protecting objects under special conditions, doesn't stop the action of time and, concerning analogue media for example, there is no way of stopping the decay process. The digital domain brought the hope for a new world in which there would no more be information transmission problems since copies were exact replicas of the original content. Just by regularly replicating would bring eternal life to digital contents. Reality shows that the replication process is perfect in theory but in practice many small distortions may appear. This has brought the need for new words to characterise needed actions: authenticity, consistency, replication, integrity...

A second major problem is the format in which the content is codified, which should be able to be accessed in the future. If no specific transfer actions are undertaken in order to maintain the feedback environment needed in order to access information; content may be well preserved however no access will be possible since the "decoder" that permits information to be interpreted by our senses, no more exists. Some words are associated to this situation and actions, as: transcoding, obsolescence, emulation, virtual universal formats...

To summarize the problems, digital copying gives identical replicas of documents; however it may introduce small differences and the document may not be accessible in the future because of compatibility aspects. Preservation as a paradigm has changed in the digital world: keeping a digital object in a secured and controlled physical environment, doesn't guarantee by any means that the document will be accessible in the future. Preservation ceases being "passive", to become "active", which means continuously checking the document and migrating it to a new environment or format when necessary. The major problem addressed is longevity and the capacity to organise preservation in such a way that any document may be accessible in any future.

## How to ethics intervene?

Ethics study which properties, if any, are responsible for the truth or validity of something. This kind of loose definition is closely related to the concept of author, and the moral rights that being an author confers: “you should not modify my work” is one of the strong aspects of authorship and has implications in the use and diffusion of contents (unless it is put on Creative Commons with that property!). Under this condition, some questions should be addressed concerning analogue to digital migration, as well as digital transcoding or restoration. Within the Audiovisual domain (which includes cinema), where information and carrier are closely interrelated and dependent, these issues may have strong implications.

Going from analogue to digital represents a huge paradigmatic change in the way information is coded. Even if a film or video and its digital equivalent look very similar, they are completely different from a technical point of view. Our perception considers both as being equivalent, which is the starting point for an analogue to digital high-quality migration. Almost always, information is compressed in order to optimise storage space and transmission bandwidth; this process is applied as long as it doesn't affect the visual rendering of the original image. Even if the image is not compressed and the digital definition is as high as the analogue definition, the technical principles underlying image representation are different. Curiously the “authenticity” of the digital replica is not questioned and authors have easily accepted migration, considering it as an added value in terms of longevity (the process is accepted with no restrictions in the video and sound domain, however within the film domain only nowadays it is considered that the digital replica of a high quality cinema film is equivalent or at least acceptable when comparing it with the original 35 or 70mm film).

Preservation exists since the beginning of film, radio and television. It happened accidentally, when carriers were kept somewhere and then rediscovered and reused; or voluntarily, when content were organised in collections associated with documentation and conserved in specific environments. In any case, these contents have undergone the challenge of time and our knowledge about preservation comes mainly through the experience accumulated while trying to keep them accessible. Since the beginning of audiovisual preservation, the quality of images and sounds has continuously improved. At the same time, our senses have become adapted to high quality audiovisual information and, even if images of the past are associated with archives and that gives them an emotional value, their quality is very poor from a technical point of view. However a question always remains: am I viewing the image as it originally was, or am I viewing a degraded image of the original image? This question has no clear answer, since in some cases the action of time is perceptible in what is called image defects (scratches, drop-outs, dirty splits), in other (many) cases, a slow and progressive degradation of the image is difficult to detect and there is no way of knowing what the original quality was.

This has implications in the “restoration” domain, where the ambition is to restore the initial quality of an image or a sound. Defects are corrected in order to achieve what is considered as being the original quality. This process can go beyond the hypothetical “initial quality” and the image can be improved thus making it look better than it originally was. The quest for the lost quality has gone beyond the initial ambition; in the age of High-Definition, of 3D images, even quite recent images can seem of very poor quality in comparison with today's productions. There is then a tendency to improve the original quality or to upgrade it in order to be acceptable when used in new contexts.

When images or sounds have a high an “artistic” value (as in cinema and music) it is considered that restoration should not alter the original quality of the content (including its initial technical defects). When addressing other kind of contents, where the notion of author is less present (television or radio programs, recordings of famous voices or moments) there is a tendency to try to improve the quality to a higher point than it initially was. In other words aging brings extra value to films and music, and implies bad quality for other domains.

## Two different conceptions

Even if law establishes quite clearly what author rights are and the conditions under which contents should be used; there are in fact two kinds of ethics: institutional ethics, where institutions have precise missions regarding content preservation and access with a very high respect to author's rights and the context in which contents are used. On the other hand there is what users daily do with contents; they circulate with little control and concern about author rights and it is considered that contents should be free for all within the web space. Ethics here are completely different!

Authors are more concerned with the circulation of their contents and the respect of their rights than with the quality of what circulates or with the ethic implications of going from analogue to digital. A certain way of thinking has prevailed in the last years concerning quality: it is considered that bad quality copies protect contents, since the originals are of a much higher quality, meaning that the original is safe somewhere and if bad copies circulate it has lesser implications. This doesn't prevent users from downloading bad quality copies and to be rather satisfied of viewing a very bad quality copy of a new film instead of going to the cinema. Bad



qualities were the consequence of narrow bandwidths, but as the quality of transmission improves, so does the quality of the images that circulate, so the bad quality protection will be without much sense in the future.

As quality advances authors and content holders may be more concerned about quality and the thin difference that exists in many domains between originals and copies will progressively fade out. The concept of author and the concept of work are also evolving and will need in the future adaptations to continuously changing technical, legal and user environments. Systems for controlling the circulation of contents on the web are quite widely used today, however the main concern could be in the future to be capable of identifying what is circulating and thus finding efficient ways to remunerate authors instead of trying to block the circulation of protected contents.

The role of institutions in this changing environment is difficult to anticipate. They have a role of keepers of cultural contents and they need to develop strong preservation strategies in order to guarantee that the original contents or certified copies are preserved forever. In parallel, they have to interact and exchange with the external world while being the warrants of authenticity and of respect to authors and other right holders.

## **Andrea GRANELLI**

### **Learning processes on the Net: more information or noise?**

The world is undergoing several radical changes.

The effects of the Financial Crisis scourging the economy are amplifying the magnitude of phenomena - dating back to early 90s – bound to the evolution of digital technologies and to the widespread of the Internet, which have had a profound effect on people behaviour and on social dynamics.

Said trends impacted strongly global markets, either by changing the competitive context or by opening up local markets to a worldwide dimension, introducing new challenges both for goods producers and for service providers: companies have to deal with global audiences and worldwide competitors never seen before, restructuring their offering to meet the ever-changing needs of their customers.

Goods producers are facing a massive introduction of new technological solutions at a pace which makes hard to exploit them adequately. Service providers are confused by a smart, active audience with changing tastes and needs; more specifically, content providers and media companies are involved in a deep crisis tied to the failure of copyright-bound business models in favour of an economy based on almost-free and/or user generated contents.

This change in competitive context, in short, has opened up a panorama both of big opportunities and of tremendous threats.

From the point of view of economic operators, especially the content providers and those active in the so-called *knowledge economy*, the challenge is to learn how to cope with the high rate of change and heterogeneity of customers, namely to produce content that is capable to engage them, becoming memorable and useful.

Besides mere technical and engineering issues - not pertaining the object of this document - there's the need to find clues of alternative ways to relate to customers and markets, both in the analysis of problems and requirements, and in the ways new technologies are used and perceived.

It's my opinion, in fact, that – as of today - users (and customers) play a role in innovation processes that is much important than ever, because of their new source function.

For explanation's sake, we introduce here 4 kinds of innovation:

- Product Innovation;
- Process Innovation;
- Organisational Innovation;
- “Customer-enabled” innovation.

Product, Process and Organisational innovation are the traditional processes involving technicalities and internal aspects of companies management; “Customer-enabled” innovation, vice-versa, leverages the fertility of the “common wisdom” in order to find clues of flaws or improvements that can be made to a product or service.

This can be achieved mainly in two ways, both improved by the use of bidirectional mass communication channels:

- analyzing the way customers use and perceive a given technology or solution, as a source of invaluable insight for the inception of new products, or for the design of successive iteration of the same product or service – which we call “Use Innovation”;
- analyzing the way groups of individuals react to complex problems in order to discover smart solutions, to build further improvement opportunities, or to overcome inherent flaws of globalisation (e.g. Co-housing, Carpooling, Community-based agriculture, Micro-nurseries, Energy conservation and waste recycling procedures) – which we call “Social Innovation”.

The analysis of consumers has to be updated according to some basic reflections about the psychology of the consumer, trying to avoid common misconceptions about an “average man” or a “model man” (Le Corbusier) with an homogeneous rational mind; the audience has to be segmented and analyzed, bearing in mind that, along with a rational side, people's choices are driven by a powerful emotional thrust. Therefore, it's necessary to embrace the most advanced instances of the ethnographic research method (e.g. Filmed street interviews, psychoanalysis of the verbatim of online surveys, shadowing ...) to increase the understanding of dynamics and evolution of purchase processes and real-world location browsing. New technologies, moreover, allow the inclusion of a broader set of tools in the box of researcher: sensoristics and biometric analysis, fields that are centric for the development of non-canonical natural interfaces, is a valuable aid for the researchers willing to “read” between the lines of a consumer's choice.

Technology, though, may get in the way of consumers' needs satisfaction, if not correctly managed. Some philosophers (Paul Virilio, among others) strongly criticize the widespread technological positivism, stating that

technologies have inherent costs implicit in their use, and incidents are integral to their very existence. Even if somehow catastrophic, this vision points out the strong influence that technological progress has on the boundaries of the traditional categories of thought, opposing, for example, reality to virtuality – indeed drawing complex and potentially dangerous scenarios.

People are hit by these changes in several ways, the most prominent of which is what is generally referred to as *information overload*.

In fact, the ease of access to growing amounts of information – enabled by the Internet – mixed with the inherent flaws of the Internet as an ecosystem relying upon machines, brings a heap of unexpected side effects, ranging from the aforementioned copyright-bound issues, to the dangers tied to monopolies in information brokering.

The belief that the Internet can act as a collective memory has to be carefully reviewed, keeping in mind some issues:

- the information overload poses problems concerning information retrieval and the ability to control search tools; the risk derives from the fact that one big provider of search tools is capable to manipulate circulating information, and to decide what has to surface over the sea of the Internet;
- the ease of production of digital contents is contributing to the information overload by adding a lot of “noise”, i.e. heaps of semiotic waste piling up throughout the Internet;
- the growth of low-cost consumer electronics industry introduced a risk factor connected with supports durability; while a clay tablet – if properly conserved – may last centuries, a low-end hard drive may fail in the arc of a week.

We propose, as an antidote to the situation described above, the widespread adoption of the methods traditionally belonging to the field of design – specifically Product Design. Companies like IDEO taught the world an important lesson: good design is to create experiences, not just products. This phrase is the key of a revolution that puts the consumer at the nexus of the design process, as only with the adoption of the viewpoint of the consumer, is possible to fully evaluate the performance of a given product or service. If the Key Performance Indicator is the customer satisfaction, the only measure of performance should be the customer himself.

The method that is emerging in today’s design field involves three main features:

- Experience as the new paradigm in product and service design
- Storytelling as a tool to drive intra-psychic semantic processes
- Interface as a solution to reduce the complexity of the hyper-connected world

Experience is the key trait to understand today’s dynamics of product value: the high cost, for example, of a coffee served over a table in Piazza S. Marco, Venice, is justified only considering the value of the whole experience, which comprehends the context in which it takes place; the only difference between products seemingly identical, moreover, is the experience they deliver – often through features not related directly to the product, for example the packaging.

A “good” experience, created by leveraging perceptive, cognitive, symbolic and semantic aspects, interacts positively with the customer’s mind, with several interesting side effects. A natural attitude of humans – amplified by the continuous exposure to advertising – is to pursue the part of the world that promises pleasuring experiences – both sensually and intellectually: in a world overcrowded with information and stimuli, the attention is a limited resource; hence, the importance – for companies and designers – of the ability to catch it through positive experiences.

The focus on experience can help in achieving success because of its importance in mnemonic processes:

- a pleased customer will have the desire to repeat the experience; a loyalty bond with the producer (or provider) is forged;
- a pleased customer will share with friends and relatives his positive experience; the customer himself becomes a living advertising.

In short, a correctly designed experience allows designers to coherently leverage both the rational and the emotional side of the customer.

Core of the experience creation process, is the ability to tell a story in an engaging way, i.e. to acquire a storytelling proficiency. Novelists throughout history showed both that “only a great novel may be able to express the multiplicity of human experience, the paths of our interiority, the behaviours inside a society” (Edgar Morin), and that a novel draws its imaginative power from its ability to join distant concepts, images, things. The latter feature is what ultimately connects the activity of novelists and designers: the discovering of an unexpected proximity causes a pleasure in the discoverer that is rationally inexplicable. Physicist (and piano player) Victor Weisskopf wrote: «What is beautiful in science is the same that is beautiful in Beethoven. There is a fog of events and suddenly you see a connection.». As we can see, many sense generating activities (research, music, literature) share this relationship with proximity, distance and storytelling. We are “storytelling



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

beings”: narration allows us to better understand ourselves and others, and to build, accept and share experiences.

The user experience of a digital technology is undoubtedly tied to its interface. Basically a component that enables the use of the technology itself, the interface is a key element common to each and every tool, artifact or interactive system. The interface is only apparently the surface layer where information are exchanged and functions are triggered: it mirrors the deep structure of the underlying object and its ultimate purpose is to ease the interpretation and use through hints and suggestions (affordances, as Donald Norman calls them). This way the interface become an ubiquitous entity, ideally connecting the layer closer to the user with the designer's original intent, in a loopback process prolific enough to reveal new approaches to an existing technology. Said that, it's easy to understand why the Interaction Design field is so important as of today, and why the use of a given technology is so influenced by the nature of its interface. Social use of technologies, in fact, depends on how the interfaces expose and facilitate community building and interpersonal interaction. In general existing interfaces are the result of two opposed thrusts: one toward minimalism and feature subtraction, the other toward complexity and completeness; Jean Marie Floch, semiotic and advertiser, notes that these trends (recurring also in commercial and advertisement iconography) are directly connected to the two archetypical aesthetic patterns borrowed from architecture and art: the Classic and the Baroque. Traces can be found in the vast majority of graphic expression throughout recent history.

A field in which the production of “good” content and the ability to foster absorption and re-use of that content is rather important, is that of the so-called e-learning. Key to this process is the development – both as a concept and as a digital construct – of a digital-self capable to collect and rearrange the stimuli spotted by a given student/user.

The digital-self is, in our vision, a unique point of access to all online activities carried out by an individual, and to all the knowledge that person stored online. This structure, though, has to be constructed bearing in mind that – because of the risks described above – the more autonomous the digital self is, the more reliable it will be in the future: Internet, in fact, can provide neither a storage solid enough to guarantee continuous access to materials stored remotely, nor the means to efficiently organize and rearrange contents. An unmanageable digital self is useless, because it lacks the recombining features that make it a prolific structure: the recombination itself, in fact, help contents to stick to a user's memory.

In the e-learning process the relationship between recombination and recalling is very important. Every process of learning (reading, listening to a lecture, visit a museum, ...) should leave structurable traces in the “digital self”, to be subsequently revised and reclassified: the reclassification of these “memory traces”, decontextualized and disembodied, will follow the specific associative structures of the digital self, thereby increasing the “awareness” of semantic relationships and incremental accumulation of knowledge.

From the above reflections derive the innovative approach of the Experience Roma project. Experience Roma aims to bring innovation in cultural tourism by supporting the user/traveller all along his travel experience: from the planning, to the fruition, to the sharing of travel memories. The approach followed by the project is e-learning oriented and based on three main features:

- Segmentation of languages according to segmentation of target: each item residing in the database (POI, Routes, ...) is described using language in three different expressive styles (basic, historical and artistic, anecdotal) for the three types of users (general tourist, sophisticated tourists / specialist, junior user/ student);
- User generated content: the content generation system expects the participation, beyond site administrators, of the user community and a of group of brokers who select and enrich the contributions of users and prepare them for publication;
- Multimedia storytelling: contents are narrated through multimedia elements of various kinds, interactive maps, virtual laboratories and learning environments dedicated to deep studies.

The project, presented nationally – during a dedicated event in Rome, and through a book – and abroad – at the British Museum during an event related to cultural tourism – will be online in the course of 2010.



*Empowering users: an active role  
for user communities*

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## ***Helen TIBBO***

### **User- Based evaluation in the Web 2.0 world**

Increasingly cultural heritage institutions are providing access to their holdings and a variety of user services through their website. Feedback from users and potential users is essential to the sound development and ongoing improvement of these efforts. Professor Tibbo discussed the range of activities cultural heritage institutions can undertake to elicit such feedback in today's web environment including surveys, interactive websites, and web analytics tools. Her focus and examples explored such repositories serving higher education.

## Daniel J. COHEN

### Engaging and Creating Virtual Communities

In 1882, a young anthropologist from Washington, D.C., went west to collect objects for the Smithsonian Institution. He found a carved round shell, roughly eight centimeters across with two holes in it, buried in a small hill in St. Clair County, Illinois.

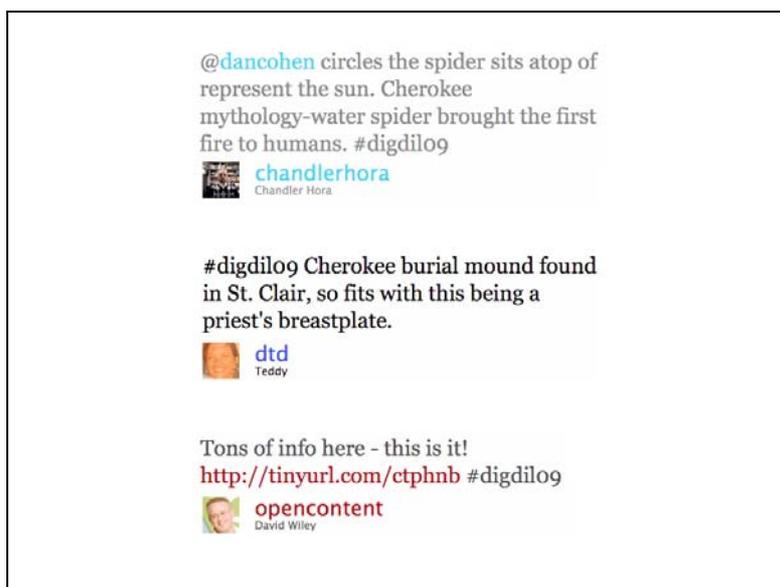
Confused about what the mysterious object was, the anthropologist brought it back to Washington and presented it colleagues in a salon for the exchange of ideas and knowledge. He took an isolated piece of history and placed it into a community that he hoped would illuminate it.

Over a century later I reproduced the anthropologist's activity digitally by presenting the same object online, to see what readers of my blog and my followers on Twitter could make of it, individually and by talking to each other. A different kind of analytical community: a virtual one.

I asked those following me online to work together to figure out what the object was. Participants in the experiment could post live comments on Twitter, and others could follow along by searching for the experiment's hashtag. (A hashtag is a hopefully unique string of characters that enables a search of Twitter to reveal all comments at a specific conference or on a particular subject.) I encouraged everyone to talk to each other and leverage each other's knowledge. And I gave the virtual community only one hour to solve the mystery of this carved shell.

What happened along the way was as interesting as the result. First, Twitter was remarkably effective in spreading word of the experiment. Indeed, in the first five minutes about a dozen others on Twitter retweeted (rebroadcast) my historical challenge to their followers. This multiplier effect meant that within minutes many thousands of people heard about the experiment; over 1,900 actually viewed a facsimile of the shell on my blog in that hour.

Once the race was on, solvers took two distinct paths toward a solution. The first path was the one I was trying to encourage: some quick thoughts about facets of the object, followed by open, online scholarly debate. I mentioned that the object was made out of shell but was found far away from water in the Midwest (of the U.S.), which led to some interesting speculation about origins and movement of Native Americans, Europeans, and Africans. Others focused on the iconography of the spider; what could it symbolize and in which cultures was it used? These were decent lines of inquiry that one could imagine in the Victorian anthropologist's salon or in the back pages of an 1882 academic journal.



Incredibly, it took much less time than an hour for a solution: nine minutes, to be exact, for a preliminary answer and 29 minutes for a fairly rich description of the object to emerge from the collective responses of roughly a hundred participants. Solution: the object was an ornamental gorget from the Cahokia tribe.

Although my experiment was admittedly a bit of a stunt, I hope the implications of it are clear. When connected in the proper way to cultural heritage materials, virtual communities can be synthesized and catalyzed to produce helpful interactions and spawn new knowledge.



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

Those in the cultural heritage world should understand that virtual communities are already developing, often without our engagement. These communities, using services like Twitter, Facebook, and dozens of smaller, more focused outlets for social media, are self-organizing and often have a radical, do-it-themselves spirit. These virtual communities have no regard for traditional boundaries, such as institutional lines or barriers erected by academic credentialing. And they are currently working to create new functionality organically out of the way their particular communities function and exchange knowledge.

It is extremely useful for cultural heritage professionals to understand this new world enabled by social media and to think about how it fits in with the goals and mission of their work. Much can be gained by looking at some of the experiments that are going on online in this realm. I present here just a few examples, ones that I feel are most relevant to the work of museums, libraries, and universities.

Since I began with a discussion of my own experiment on Twitter, let me stay with that popular service for a moment. Jay Rosen is a professor of journalism at NYU who is very active on Twitter ([http://twitter.com/jayrosen\\_nyu](http://twitter.com/jayrosen_nyu)), providing what he calls a “flying seminar” to his many followers by pointing them to important publications and commenting upon current events. Although detractors see Twitter as nothing more than 140-character messages, Rosen has seen it as a place to build a community of discussion and debate. His activity on Twitter shows that social media is so flexible that it is what you make of it. Rosen calls what he does on Twitter “mindcasting” rather than “lifecasting” - that is, he uses Twitter to project and discuss what he is thinking about in his academic work rather than to inform others of his physical whereabouts or what is going on in his non-academic life.

Moreover, the use of Rosen’s Twitter posts can be aggregated to get a good sense of what the journalism community is tracking at any one time. Rosen has set up a page called 40 Twits that shows which articles that he has highlighted are getting special attention by his readers.

Since Rosen set up his 40 Twits page, Twitter itself has launched features that make it easier to visualize the activity of a particular virtual community - to get the pulse of that community. For instance, I recently used Twitter’s “list” functionality to bring together all scholars working in digital humanities (whether they are in universities, libraries, museums, or the technology field - again, across boundaries). This master list creates a unified stream of posts and makes it easier for scholars in this area to connect with each other.

In addition, the digital humanities Twitter list can, like Jay Rosen’s Twitter followers, be used to provide a snapshot of aggregate interest. I recently set up *Digital Humanities Now* (<http://digitalhumanitiesnow.org>) to take the combined activity of the 350 people on my digital humanities Twitter list and show what they are reading and commenting upon. For those outside of the field, visiting the *Digital Humanities Now* site provides at a glance a good sense of what the discipline is interested in.

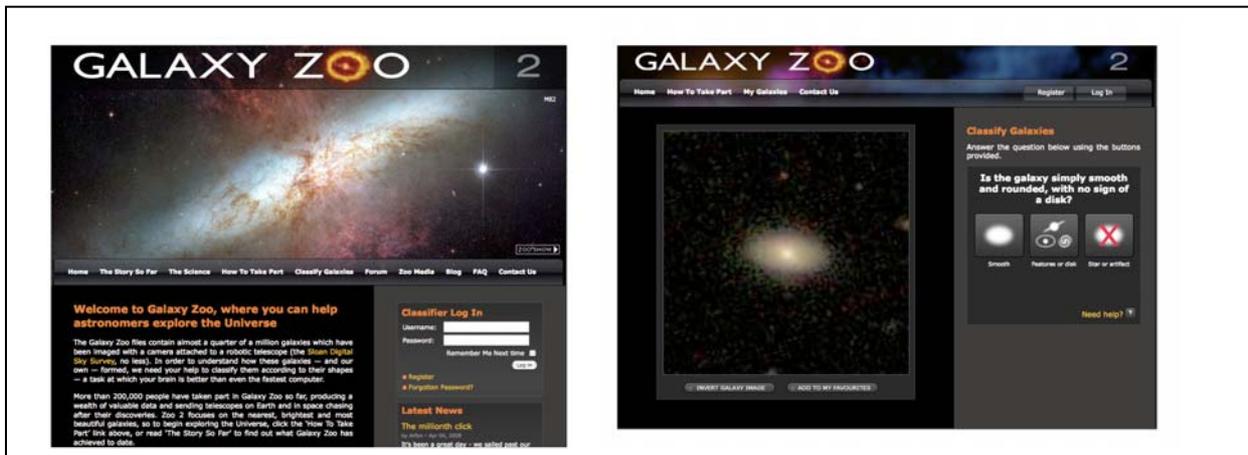
Cultural heritage practitioners should also look beyond the humanities for inspiration on how to use social media. A good example from the natural sciences is Galaxy Zoo (<http://galaxyzoo.org>), a website created by astronomers. These scientists currently are having to deal with a flood of information from telescopes and other sensors, and they obviously have limited personal time to sort through the millions of images and petabytes of data to find heavenly objects worthy of further study.

The astronomers behind Galaxy Zoo rather shrewdly decided to enlist the help of amateur astronomers and other enthusiasts of the night’s sky to see what they could do together as a virtual community. Remarkably, more than 200,000 people have taken part in the community, helping to classify tens of thousands of galaxies. Doing aggregate work that no single astronomer or even a large group of professional astronomers could do, these “Zooites” have confirmed that galaxies do not prefer to spin clockwise or counterclockwise, and they have found unusual objects that astronomers have marked for more careful inspection. By curating their community well (and sharing in the joy of discovery), the astronomers behind Galaxy Zoo have produced a valuable lesson in crowdsourcing for academic purposes.

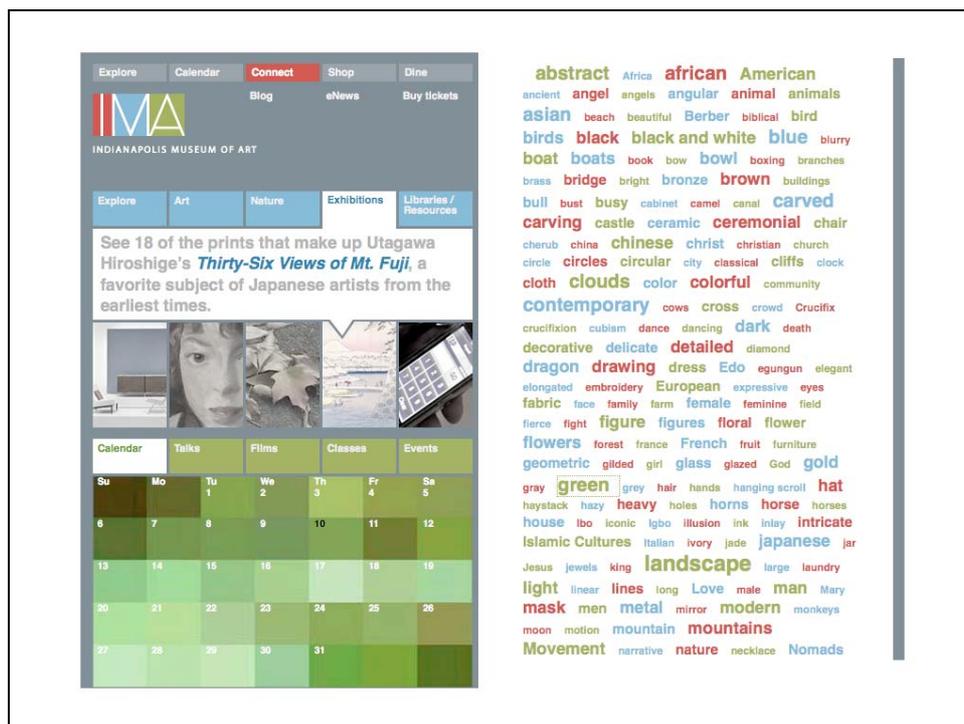


Empowering users: an active role for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009



Museums, of course, have also engaged in their own crowdsourcing experiments. A project and associated software called Steve (<http://www.steve.museum/>) is being used by some museums to permit visitors to assign tags, or descriptive keywords, to works of art. The Indianapolis Museum of Art displays its Steve tags as another way to navigate its collection, and the community has now created scores of tags about virtually every aspect of IMA's holdings.



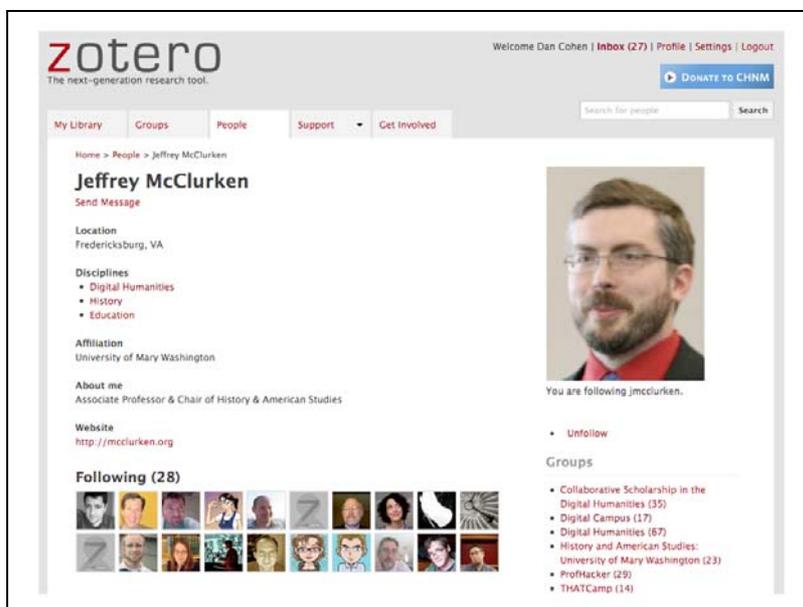
Unsurprisingly these tags sometimes differ from the controlled vocabulary and academic descriptions given to those very same artworks by IMA's curators. It is instructive for cultural heritage professionals to compare the outcome of Steve tagging and a museum's formal descriptions. For Robert Indiana's iconic "LOVE" painting, from the 1960s, the gallery label focuses on its geometry and frequent reproduction, the latter theme echoed by the amateur community in tags such as "omnipresent" and "postage stamp." But the community goes on attach some value judgments about the painting: "overexposed, overrated," as well as some helpful tags for those scanning the entire IMA collection for "word art."

What can we learn from these experiments in social media, crowdsourcing, and virtual community-building? First, there is still a yearning for validation and authority, which cultural heritage institutions are in a good place to capitalize on. Second, social media is still an annoyance for many people, often because it appears as an

unseemly mix of personal and professional life. (Thus Jay Rosen's urge to "mindcast" rather than "lifecast.") Finally, much of the work in this area is going on with the "secondary" products of scholarship rather than the primary ones. Virtual communities can be extraordinarily useful for the classification of galaxies or artworks, or adding context to a primary source; it is probably unreasonable to expect the crowd to produce a great novel together.

At the Center for History and New Media and George Mason University we have been engaged in our own experiment to build scholarly communities online through the exchange of bibliographical (i.e., secondary) information. The Zotero project, consisting of client and server software used by hundreds of thousands of scholars and researchers in virtually every field of endeavor in forty languages, has entered a 2.0 phase in which we are focusing on social media and the combined power of groups.

Zotero 2.0 allows researchers to establish online profiles and to expose information about themselves and what they are studying. As on Twitter they can follow and be followed by others. Should they choose to do so, they can share their bibliographies and associated scholarly materials.



The screenshot shows the Zotero user profile for Jeffrey McClurken. The profile includes a header with the Zotero logo and navigation links. The main content area shows the user's name, location (Fredericksburg, VA), disciplines (Digital Humanities, History, Education), affiliation (University of Mary Washington), and a list of groups they are following, such as 'Collaborative Scholarship in the Digital Humanities' and 'Digital Campus'.

A year ago we added the capability to create and join groups of any size and theme. These Zotero groups can be private or public, and to date nearly every discipline and subdiscipline is represented in thousands of macro- and microgroups. Some groups are loose networks of common interest; others are strongly curated by their members or their owners.

Regardless of how they are structured, these virtual communities are engaged in substantial exchanges of knowledge, and can be used for the creation of new collaborations and understanding.

Similar communities are just starting to arise within institutions and between institutions. At the City University of New York, the CUNY Academic Commons, which launched in 2009, attempts to unify online the many CUNY campuses through a common portal and the exchange of social media such as blog posts. HASTAC, the Humanities, Arts, Science, and Technology Advanced Collaboratory, brings together thousands of scholars working at hundreds of institutions to discuss the ways in which digital media and technology can transform their work and enable new forms of collaboration.

Social media can also be used by cultural heritage institutions as a live news feed to reach and engage a large audience. In the United States, the Library of Congress and the Smithsonian are both very active on Twitter, Facebook, and other social media outlets. The New York Public Library, perhaps ahead of its time, devotes an entire section of its website to its presence in social media and the ways that its audience can interact with it through those media.



Empowering users: an active role for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

The screenshot shows the New York Public Library website interface. At the top, there is a navigation bar with the NYPL logo and a search bar. Below this, a large red banner reads 'Connect'. The main content area is titled 'Blogs, Videos & Publications'. On the left, there is a sidebar with various links like 'Chats @ NYPL', 'Literacy Journal', and 'Schomburg Center Junior Scholars Program'. The central part of the page features a 'Connect with NYPL' section with social media icons for Facebook, Twitter, YouTube, and Flickr. To the right, there is a 'DONATE NOW!' button and an 'ASK NYPL' chat option. The bottom right corner includes a 'FROM OUR BLOGS' section with a quote about paper scalptors.

Ultimately, these experiments in engaging and creating virtual communities lead one back to the objects of research and scholarship themselves. When I visited the Smithsonian as part of their “Smithsonian 2.0” effort to address social media and interactivity, I went behind the scenes and saw Abraham Lincoln’s handball and paraphernalia from the presidential inauguration of Barack Obama. Just like the ceremonial Cahokia shell from the nineteenth century, these objects could have constellations of researchers and the general public gathering around them online. Cultural heritage institutions have always curated physical objects; in the twenty-first century they will also have to curate virtual communities around these objects.



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## **John UNSWORTH**

### **Computational methods in humanities research**

#### **What are computational methods in the humanities? What's the difference between using a computer and using computational methods?**

Because the computer is a general-purpose modeling machine, it tends to blur distinctions among the different activities it enables and the different functions it performs. Are we word-processing or doing email? Are we doing research or shopping? Are we entertaining ourselves or working? But even though to an observer all our activities might look the same, the goals, rhetoric, consequences, benefits, of the various things we do with computers are not the same. I would bet that everyone here uses a web browser, a word-processor and email as basic tools in their professional life, and I expect that many of you are also in the humanities. Even so, you do not all do humanities computing – nor should you, for heaven's sake – any more than you should all be medievalists, or modernists, or linguists. However, if you are in any of these disciplines, one of the many things you can do with computers is to use computational methods, in which the computer is used as tool for modeling and analyzing humanities data and our understanding of it. Today, I simply want to point out that such activity is entirely distinct from using the computer when it models the typewriter, or the telephone, or the movie theater, or any of the many other things it can model.

There are any number of tools for modeling and analysis, depending on the nature of the source material: xml is a way of modeling text; mpeg is a way of modeling audio and video; GIS is a way of modeling geographic data, with other kinds of information layered on top of it; and we have various ways of modeling other kinds of information. The point, in each case, is that there should be some way of validating the model – some way of determining whether it is internally consistent, and some other way of determining whether it corresponds accurately to important features of the thing it models, even though the selection of those features and the importance given them will, inevitably, reflect the subjective interests and purposes of the person doing the modeling. Still, a model is a form of knowledge representation, and knowledge is always situated – in a person, and with a purpose – so, beyond accurately expressing those features of the object on which all observers can agree, the measure of success is not objective accuracy, but rather expressive completeness.

In addition to expressing the perspectives and purposes of the modeler, new perspectives on familiar materials can become available to others, as a result of the creation of digital primary resources. As an example here, I offer The William Blake Archive, which presents full-color images, newly transcribed texts, and editorial description and commentary, on all of Blake's illuminated books, with non-illuminated materials (manuscript materials, individual plates and paintings, commercial engravings, etc.) now coming on line. The Blake Archive makes it practical to teach Blake as a visual artist, by the simple fact of the economics of image reproduction on the web, and this is a fundamental change from the way I was taught Blake, through Erdman's text-only synthetic edition (which is also, by the way, available on the site).

There's a deeper impact of digitization, though, beyond increased access: that deeper impact is realized by those who do the digitization, provided that they are subject-area experts who are aware of the complexity of the source materials. In the act of representation, seemingly simple questions, like "is this poem a separate work, or is it part of a larger set of poems?" can be unavoidable – requiring some decision at the level of markup, for example – and they can also raise issues that are critical to understanding the work in question. However we may decide such questions, we are both informed and constrained by our own decisions, when subsequent and related issues arise. Likewise, with images, when we digitize, we choose file-type, compression, color-correction, and other settings based on what we consider valuable and significant in the image – and when our chosen strategy is applied across a large body of images, or when others come to our digital surrogate with purposes we hadn't shared or predicted, we are bound to confront the fact that our surrogate has been shaped by the perspective from which it was produced. In this sense, the real value of digitization for humanities scholarship is that it externalizes what we think we know about the materials we work with, and in so doing, it shows us where we have overlooked, or misunderstood, or misrepresented significant features of those materials.

No better example of this struggle between materials and intentions could be found, I think, than the documentation on the "Editorial Commentary" pages of the British Library's Nineteenth-Century Serials Editions project (<http://www.ncse.ac.uk/commentary/index.html>), which lay out the choice of materials, problems raised by multiple editions in serials, the construction of a "datamap" and a "concept map" for the materials, structural "segmentation policies," and the metadata schema that evolved during the course of the project team's effort to analyze and represent its six 19th-century serials. I'll quote just briefly from a now



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

disappeared “work in progress” page that was once on the NCSE site, now no longer even in the internet archive, for its description of developing the NCSE datamap, in order to explain what I mean by this deeper impact of digitization. The datamap is a map of “data fields” in which the content of the NCSE primary materials will be represented, and it maps the relationships between those fields. Once an initial sketch of the map was prepared, it was tested against the primary sources in “a page turning exercise in which the team assimilated new data fields occurring in the source materials into the map and also reconfigured the map as appropriate.” The team that went through this exercise noted that “this work required interpretation at every stage, our abstract conceptualisation of the source materials becoming increasingly concretely represented in the map as it was developed.” Even so, the data don’t always obey the map:

‘The creation of the map has flagged up some potential challenges in the way in which our data might be rendered. As is evident from the map there are instances where relationships between its fields skip levels. (e.g. department items) and some items ‘float’ and can exist at almost any level (e.g. price). The dilemma facing ncse is thus whether to enforce an artificial framework upon the sources (top-down) or to attempt to adapt the framework to the sources (bottom-up).’

For me, this is very reminiscent of the exercise of developing the original SGML Document Type Definition for the Rossetti Archive, in doing which we went through an iterative process of modeling the components of Rossetti’s paintings and poetry, an exercise that forced an explicit discussion of the nature of these materials, the relations between their parts, and the rules that could be deduced to govern the markup that would represent these. I guarantee that, in both of these cases, unless we had been digitizing the materials in question, and unless the scholar-expert had been party to that digitization, these discussions would never have taken place, and this explicit specification of the scholar’s understanding of the materials would never have emerged. But these are the benefits of the early stages of digital humanities - the handmade phase, if you will, where the focus tends to be on scholarly editing as the analytic activity enabled by modeling the source material in digital form.

Beyond modeling, and beyond the hand-made phase of digitization, what does it mean to speak of computational methods? The word “method” implies a way of doing something; there should be something that can be computed on the basis of the representation, whether that’s a matter of information retrieval, algorithmic transformation, statistical profiling or comparison - essentially, I would say “computational methods” involve some kind of analysis, and that analysis produces some kind of (reproducible) results. Those results are not, themselves, the end of the story: in the humanities, empirical results are most likely to be the beginning of the story - the evidence for an argument, the occasion for an essay, which still needs to be argued and essayed, in the same way we’ve always done.

### **What are the conditions that call for computational methods?**

In the handmade phase, we could choose digitization, but we could choose not to digitize as well: scholarly editions, for example, can still be produced without digitizing the source materials. However, when we move from handcraft to industrial-scale digitization, we are required to consider computational methods in a different light. The primary condition that calls for computational methods is the availability of a large amount of data in digital form, with the possibility of reprocessing that data into other, purpose-built, representations. With respect to humanities research that focuses on text, we are certainly in that industrial phase: Google Books, as of October, had scanned about 10 million books. The HathiTrust, which is the shared digital repository that stores materials scanned out of the collections of some of the major research libraries in the U.S., had about 4.5 million volumes as of last month. Only some of this material is public domain, but the Google Books Settlement provides for the creation of at least two research centers that will provide access to the in-copyright material, for researchers in various disciplines who want to do “non-consumptive research” with it (where non-consumptive” means, basically, that you’re not supposed to be taking material out of the research environment).

As Franco Moretti points out, in *Graphs, Maps, Trees*, humanities scholarship normally focuses on a “minimal fraction of the literary field”:

“. . . a canon of two hundred novels, for instance, sounds very large for nineteenth-century Britain (and is much larger than the current one), but is still less than one per cent of the novels that were actually published: twenty thousand, thirty, more, no one really knows—and close reading won’t help here, a novel a day every day of the year would take a century or so... And it’s not even a matter of time, but of method: a field this large cannot be understood by stitching together separate bits of knowledge about individual cases, because it isn’t a sum of individual cases: it’s a collective system, that should be grasped as such, as a whole.”

I think that what Moretti calls “the quantitative approach to literature” acquires a special importance when millions of books are equally at your fingertips, all eagerly responding to your Google Book Search: you can no



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

longer as easily ignore the books you don't know, nor can you grasp the collective systems they make up without some new strategy—a strategy for using computational methods to grapple with profusion.

However, in order to exercise these strategies, in order to use computational methods, it is almost always necessary to be able to reprocess texts into new representations—transforming them, for example, into database representations, or indexes, or adding information about parts of speech, normalized spelling, etc. Particular purposes require particular representations, and different data-types will offer different features for analysis, but the basic point is that in order to do more than search and browse, it is almost always going to be necessary to reprocess the data, and one would usually wish to begin that reprocessing with the richest form of the source material.

### **What is the potential of such methods? What kinds of research questions can be addressed computationally?**

When working with texts, computational methods can help us answer questions having to do with any number of empirical features of those texts and their authors, including vocabulary, syntax, grammar, sound, structure, reference, location, genre, gender, metaphor, intertextuality, and many other things. For example, we might examine

- historical trends in the use of language (for example, is there a golden age of the passive voice?)
- distinctive patterns of language that are characteristic of an author, by comparison to other authors of the same period (for example, what are the words that Jane Austen avoids, by comparison to her peers?)
- features that distinguish one genre from another, or one mode from another (for example, comedy vs. tragedy, or sentimentalism vs. realism)
- features that distinguish male from female authors in the same period
- the role of certain ur-texts, like the Bible, in shaping later texts
- authorship attribution, for example in multi-authored works like encyclopedias and so on.

Given the ability to reprocess texts, these questions can generally be answered at a level of specificity that would be impossible for a human reader to achieve, simply because the computer can keep track of empirical evidence at a very granular level. The role of the human interpreter is to understand and validate the methods by which the evidence is produced, and then to make sense of that evidence, in an argument.

Similarly, when working with other kinds of raw material—music, images, maps, 3D models, etc.—whatever empirical features that material offers will be available to computational methods, and those methods will support whatever meaningful questions can be asked on the basis of such features. Taking the whole process full-circle, one form of validation may eventually be production. In music composition, for example, computers have been able to learn algorithmic composition, using the features that characterize a particular composer well enough to produce new compositions that are plausible as works of that composer (see, for example *Computers and Musical Style* by David Cope, professor emeritus of music at the University of California at Santa Cruz; or listen to <http://bit.ly/4QARWn> for a couple of samples of the work of his program, named Emily Howell). In music, this is a matter of getting the syntax right; in language, this form of validation will be more difficult, because of the semantic component, but one day, we may see new works of fiction in the manner of famous now-dead authors, produced by computers. This is really just the generative inverse of analysis.

### **What has been the impact of such methods? Have computational methods changed the way we study and teach the humanities?**

Certainly, both the hand-made and the industrial phases of digitization have had profound impacts on how we study and teach, or if they haven't, they should. The profusion of texts makes it all the more important that we teach students to understand the importance of editions, and to distinguish between reliable and unreliable editions. The presence of true scholarly editions in electronic form makes it possible to provide both students and researchers with unprecedented depth of access to the process and variety of artistic production (think back to the Blake example, or look at some of the scholarly editions produced by the University of Virginia Press—Melville's *Typee* with its manuscript, and an analysis of the process of revision that led to the final text. To take a different kind of example, there have been a number of interesting digital humanities projects based on correlating textual data with maps, for the purpose of analysis. I have first-hand experience with several of these, including the Valley of the Shadow project, which mapped military records and information from diaries and newspapers to produce interactive battle maps of some of the major campaigns in the American Civil War, and The Salem Witch Trials project, which mapped documentary records from the trial records to produce an interactive record of the location and spread of witchcraft accusations in the Bay Area colonies. In the case of the Civil War, these maps, combined with other data, helped to produce insights about the daily lives of individual during the civil war that no research to date has been able to match: read Ed Ayers's book, *In the*



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

Presence of Mine Enemies to see exactly what kind of impact the coordination of very granular information from many different source can have, on the telling of history—this is a book that couldn't have been written without digitized primary source material. Ben Ray, in the Salem Witch Trials, was able to use the combination of maps and trial records to ascertain that a popularly held belief about the geographic concentration of accusers in one part of town, and accused in the other, was simply not true—and that, moreover, there were more accusations of witchcraft outside of Salem, in the larger colony, than inside it. These are some results that derive more or less directly from the act of digitization, by domain experts.

In my more recent experience, I've been working to develop tools that leverage digitized representation for the purpose of machine-aided analysis—in this case, text-mining. Over the last four years, I have worked with faculty, students, and computer experts at half a dozen different institutions in the United States and Canada, and at the National Center for Supercomputing Applications, to develop MONK, a workbench for text-mining across literary collections.

The full release of The MONK Project, available by authentication to about 50,000 faculty and 400,000 students at a dozen universities in the Midwest, includes about a thousand works of British literature from the 16th through the 19th century, provided by The Text Creation Partnership (EEBO and ECCO) and ProQuest (Chadwyck-Healey Nineteenth-Century Fiction), along with Martin Mueller's edition of Shakespeare (thirty-seven plays and five works of poetry), plus over five hundred works of American literature from the 18th and 19th centuries, provided by libraries at Indiana University, the University of North Carolina at Chapel Hill, and the University of Virginia.

MONK stands for Metadata Offer New Knowledge, and the metadata MONK provides is at the word level (part of speech, lemmata, position in the text, n-grams, etc.) for each of the 150 million words in this corpus. Behind the workbench interface, MONK's quantitative analytics (naive Bayesian analysis, support vector machines, Dunning's log likelihood, and raw frequency comparisons), are run through a toolkit developed at NCSA, called SEASR. Users typically start a project with one of the toolsets that has been predefined by the MONK team. Each toolset is made up of individual tools (e.g. a search tool, a browsing tool, a rating tool, and a visualization), and these tools are applied to worksets of texts selected by the user from the MONK datastore. Worksets and results can be saved for later use or modification, and results can be exported in some standard formats (e.g., CSV files).

In the process of designing MONK, we worked with humanities doctoral students and junior faculty who had specific research questions they wanted to answer, using these tools. For example, Sarah Steger was interested in sentimentalism in British fiction, and specifically, what distinguished sentimental from non-sentimental fiction, at the level of vocabulary—and, by extension, at the level of subject matter. She started by running naïve bayes routines on a training set of 409 mid-Victorian novels that she classified as either sentimental or unsentimental. The larger testbed, to which the software applied Sarah's training data, was 3,921 novels; ultimately, the software returned 1,348 chapters as sentimental, and it had detailed information about the language use that was characteristic of the sentimental chapters, and that distinguished them from non-sentimental chapters. She was able to get fairly definitive results on the words that separate sentimental from unsentimental fiction, as well as learning that Dickens seems to be the archetype of sentimentality in this period of British fiction.

### **What are the limitations of such methods? What research questions cannot be addressed computationally?**

In general, research questions that are wholly intuitive in nature, or that do not make use of empirical evidence in source material, will not lend themselves to computational methods. Aesthetic appreciations, likewise, don't benefit much from these methods. Arguments that depend on the performance of the critic may be assisted by evidence of the sort that these methods can provide, but they may not, as well.

Reflecting on our experience in the MONK project, where we based our analysis on meticulously prepared texts with in-depth linguistic information, my colleague and co-investigator Martin Mueller gave the following, fairly exhaustive, account of the limitations of our methods in the MONK project. He said,

"The computer has no understanding of what a word is, but it follows instructions to 'count as' a word any string of alphanumerical characters that is not interrupted by non-alphabetical characters, notably blank space, but also punctuation marks, and some other symbols. 'Tokenization' is the name for the fundamental procedure in which the text is reduced to an inventory of its 'tokens' or character strings that count as words. This is an extraordinarily reductive procedure. It is very important to have a grasp of just how reductive it is in order to understand what kinds of inquiry are disabled and enabled by it. A word token is the spelling or surface of form of a word. MONK performs a variety of operations that supply each token with additional 'metadata'. Take something like 'hee louyd hir depely'. This comes to exist in the MONK textbase as something like

hee\_pns31\_he louyd\_vvd\_love hir\_pno31\_she depely\_av-j\_deep



*Empowering users: an active role  
for user communities*

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

Because the textbase 'knows' that the surface 'louyd' is the past tense of the verb 'love' the individual token can be seen as an instance of several types: the spelling, the part of speech, and the lemma or dictionary entry form of a word."

### **Conclusion:**

What's really changed? Well, perhaps nothing, for humanities scholarship that isn't primarily interested in modeling its source material in order to understand its structure or ontology, or scholarship that isn't especially interested in the evidence that is offered by that source material, for empirical arguments. But if your scholarship depends, to some extent at least, on empirical evidence, or if you are interested in the features that the computer can "understand," or if you are interested in correlating different kinds of evidence, along some shared dimension, then computational methods could change your work entirely, could lead to new answers to old questions, or even better, to altogether new questions. And as we approach, in the next decade, a time when all the books (not archives, but books) in research libraries are digitized, it may become harder to ignore the capabilities that computational methods offer to the scholar and the teacher. This is, in fact, how change has come to other disciplines—on the heels of a transformation of the bulk of their data from analog to digital. Our day is coming soon—in fact, it's already upon us, so it's time to begin thinking about how to cope with it.



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## **Ingrid PARENT**

### **Internet – driven convergence between libraries, archives and museums: an opportunity, an inevitability or both?**

I am delighted to be here and would like to thank the Fondazione Rinascimento Digitale for their invitation to speak at this special event. My talk will be different from preceding ones, since I will focus on convergence – specifically, Internet-driven convergence involving libraries, museums and archives. Is this an opportunity? Is it an inevitability? Or perhaps both? So today I would like to share some thoughts and examples with you about the changing nature of information and users in a digital age and how cultural institutions are responding, and how they will need to change in order to meet these challenges.

I believe that the collections of the future will integrate access to traditional print, museum and archival materials, with digital and three-dimensional objects, media formats onsite broadcasting and telecommunications. This extraordinary but achievable vision can only be realized if we bring libraries, archives and museums together through either physical convergence or through virtual convergence using the power of the Internet.

#### **The role of technology**

As we near the end of the first decade of the 21st century, we increasingly live in an age of connections, collaborations and convergence – between people, between professions, between countries, between ideologies. And technology helps foster the connections that make convergence possible, linking people and ideas across borders and boundaries. For libraries, archives and museums (LAMs), technology enables us to envisage collaborations and partnerships and projects that never would have been possible before the age of the computer, the Internet, digitization and so on. And technology allows libraries, archives and museums to bring together their collections and expertise in a way that can be accessed, ideally, from a single, central point by the user – who isn't concerned with our professional differences or demarcations, but simply wants quick and easy access to information.

This is quite a shift in mindset – and technology is a key factor behind such developments.

Most information searches begin with a search engine, and not with a library catalogue. Google – not JSTOR, or WorldCat, or any one of a number of other academic databases – is the default search option of choice for our users, regardless of what we, as information professionals, may think of this situation. And our users are no longer passive; they are using the information they find to create and share more information.

So we have all this content, all this technology, all these users. Yet we are almost drowning in the very information landscape we have created.

The challenge for libraries, archives and museums is to combine their expertise and offer solutions to such overload in a way that makes sense for all types of users all over the world.

However – it's also important to take a step back and realize that technology should not be viewed as the answer for everything. Technology should not drive digital projects. It is an enabler, but it is not an end-point.

Now, I'd like to look at the term "libraries, archives and museums" or LAMs. This is a kind of shorthand that lumps three closely related, yet distinct professions together – and while it may be convenient, we need to keep in mind that even as we use this term, the key commonalities and differences between the professions must be understood. If not, we run the risk of undermining trust and impeding serious dialogue across these communities.

So – how are we distinct? The key differences between libraries, archives and museums reside in the nature of their respective collections and in the culture and orientation of their professions.

Libraries have tended to focus on the description of metadata of books and serial publications, and the development of resource-sharing networks and practices such as interlibrary loans to serve users.

For archives, work has centred on establishing provenance and context, and the development of collective documentary approaches to establish the authenticity and meaning of archival documents and records.

Museums have focused their efforts on classifying or grouping individual objects based on their material, usage or some other common physical or social aspect, and the interpretation and use of objects in public exhibits designed to illustrate or explain a particular story, event, or other aspect of natural or human history.

So while the general public may not particularly care whether their information comes from a library, an archive or a museum, the professionals of these different institutions do. Again, a common understanding and appreciation of why each profession does things differently will determine the success or failure of convergence.



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

While LAMs are different in various respects, I think it's essential to remain focused on our commonalities. Libraries, archives and museums are all collection-based, they are service-oriented organizations, some more than others, and generally they are not-for-profit. They tend to have mandates that support learning, education, scholarly study and research, the advancement of knowledge, and the collection, preservation, organization, use and enjoyment of the collections in their care and custody.

These characteristics mean that LAMs are infrastructure-heavy. They need buildings that are big enough to house them, as well as micro-environments for various formats of material, and the security to accommodate collections, exhibitions, staff and public areas. They also share two other important features: a high demand for information technology and management, and skilled workforces to support collection management and outreach activities.

We all deal with some kind of evidence or object. The systematic discovery or protection of this evidence, the need to inventory and describe it, the problems of physical conservation, the importance of interpreting and presenting it for the general public, its value both in education and in scholarly study – these are issues we all face, and ones that bring us together. Comparing the solutions tried by different disciplines provides useful inspiration. Because all forms of evidence are equally important, contributing their unique insights to the mosaic that forms our social memory, our heritage.

So where are we at? What is the state of collaboration and convergence among the LAM community in today? To me, the answer is clear: the true convergence of libraries, archives and museums with each other is still in its infancy. In fact, if I may use an analogy, it is not unlike the silent movies of the early 20th century, when a moving image was projected on a screen, and a real, live orchestra provided the music and sound. Image and sound, two different media, were collaborating, but not yet fully integrated into an altogether new, organic whole, such as exist today.

On a regional, national or international level, a highly converged system or network of “libraries, archives and museums” today remains the exception, not the norm. Indeed, I think that in many cases, silos rather than synergies continue to define the LAM landscape. There may be a number of reasons for this situation. Appropriate levels of funding may not be available to foster effective collaboration and convergence. Individual institutions may be bound by mandates and statutes that set a restrictive set of operating guidelines and norms. The appropriate technologies may not be available or affordable. And contradictory cultures and attitudes at different institutions can pose enormous barriers.

However, in spite of these obstacles, there are some exciting examples of convergence.

Library and Archives Canada was formed from the merger of the former National Library and the National Archives in 2004. Why? It wasn't for political or economic reasons. It was because separate institutions in the context no longer made sense. The two legal mandates were highly compatible, and the holdings were complementary. But most importantly the digital environment was blurring the distinctions between their holdings more and more, and the highly corresponding skills and competencies of the staff were not really being used to their full potential.

The lessons to be learned in any type of merger is not to drag the past into the future – that is, not to adamantly adhere to practices that may have worked well in the past, but no longer are suited to the realities of today and tomorrow. So – be flexible, be open, be willing to change, and take some risks.

Other examples of convergence of information resources:

- The Smithsonian Institution
- University of Calgary's Libraries and Cultural Resources
- Columbia and Cornell University Libraries Cooperation
- University of British Columbia Library provincial digitization project
- Europeana
- World Digital Library
- LAMMs at the International NGO level

I bring all these examples to light because I think what they demonstrate is that, yes, technology is a great enabler of convergence, of collaboration, of innovative partnerships that can take shape in many ways.

But what is more important is a willingness to recognize our commonalities. The more this kind of communication takes place, the more opportunities will present themselves, including new models of collaboration and convergence, new kinds of training and education, partnerships, preservation, collection management approaches, and search and online access offerings. Working together also lends LAMs a stronger and more articulate voice to advocate for funds and user rights, and to promote strengths and qualities.

However, there remain several challenges to advancing convergence or collaboration. There is obviously the technology challenge, always keeping up to date with new ways to create, share and preserve digital



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

information resources. But I would argue that the biggest challenges lie more on the non-technical side, in the organizational and cultural attitudes that we have developed as information professionals.

Successful convergence across disciplines also demands a new set of diverse skills and competencies, people who combine subject-matter expertise with a knowledge of digital technology and information management. We need to create a new LAM professional, someone who can function in this increasingly cross-border landscape. And if our institutions encourage such environments, this, in turn, can help us all attract and retain younger professionals – people who bring new skill sets, perspectives and attitudes to the table.

There is also a need to create guidelines for organizations that are working on converged collections, for example, ensuring that they are catalogued and accessible via a single website.

And finally, convergence and collaboration means that libraries, archives and museums will need to master the challenges of virtual exhibitions, multimedia, and fast and efficient digitization in order to take full advantage of the potential of digital technologies. We will need to become better communicators across many languages to promote joint projects.

The Web enables us to recall that the past was a holistic place – the documents, artifacts, paintings, books, historical sites, photographs – they all existed together, informed each other and collectively formed part of the holistic context of any historical action. The Web enables us to overcome the territorial boundaries that have arisen over decades – when we broke up the past, and put some in museums and some in archives, some in libraries, some in historic sites.

The Web is all about convergence, and it enables us to overcome these boundaries and reassemble the past.

At the beginning of this talk, I posed the question: is Internet-driven convergence A) an opportunity? B) An inevitability? Or C) Both? The answer, as I see it, is probably both, although at this stage, more of an opportunity than an inevitability. Internet-driven convergence is a huge opportunity for libraries, archives and museums, and it is inevitable, given the relentless march of technology and the increasing interdependencies that are bringing us all closer together – whether we like it or not. It is our task, then, to take this technology and apply it in innovative ways, all the while remaining sensitive to our differences, confident of our commonalities and strengths, and focused on the users who will benefit enormously from our efforts if we are successful.

## **Manuela SPEISER**

### **Digital Libraries and Digital Preservation: EU-Research Perspectives**

EU-funded research on digital libraries and digital preservation deals with leading-edge information and communication technologies for expanding access to and use of Europe's rich cultural and scientific resources. It also investigates how digital content created today will survive as the cultural and scientific knowledge of the future.

#### **Three areas of research**

The rapid pace of change of electronic devices and formats for recording, storage and use represents a threat to long-term accessibility of these resources. With the increasing proliferation of digital content this risk is imminent equally for businesses, the public sector and individual users. Digital preservation research aims at concepts, techniques and tools for ensuring availability of digital resources over time, while guaranteeing the integrity and authenticity of the information as originally recorded.

Leading edge technologies can enhance users' experiences with cultural and scientific digital resources. ICTs for capturing, rendering, modelling and visualising cultural artefacts support study and creative use of artefacts, and their aggregation into virtual collections. ICT-funded research establishes the basis for scalable and interoperable digital library platforms supporting digitisation and retrieval of heterogeneous content, in multimedia formats, from distributed collections and across languages.

These broadly are the three areas targeted by "Digital Libraries and Digital Preservation" research in the ICT programme:

- Digital preservation
- Digital libraries
- Cultural heritage

#### **Target Outcomes**

The ICT Work Programme 2009-2010, which is the basis for ICT Call 6, identifies six target outcomes related to the three domains, each associated to a specific funding scheme:

a) Scalable systems and services for preserving digital content, handling end-to-end workflows for different types of digital resources, guaranteeing their long term integrity and authenticity. The feasibility of solutions should be demonstrated in large scale testbeds (IP).

b) Advanced preservation scenarios:

b1) Methods and tools for preserving complex objects, addressing the life-cycle of composite digital information instances (STREP).

b2) Intelligent digital curation and preservation systems able to learn, reason and act autonomously, integrating tools and methods to support the complex decision making processes for appraisal, selection and management of diverse collections of digital resources. The system should ensure that the representation of the objects and their embedded semantic knowledge are preserved in order to support their future re-use. Appropriate verification scenarios should be integral component of the work (IP).

c) Innovative solutions for assembling multimedia digital libraries for collaborative use in specific contexts and communities, enhancing scholarly understanding and experiences of digital cultural heritage (IP).

d) Adaptive cultural experiences exploring the potential of ICT for creating personalised views of various forms of cultural expression, reflecting individual narrative tendencies (i.e. adapt to the background and cognitive context of the user) and offering meaningful guidance about the interpretation of cultural works (STREP).

e) Interdisciplinary research networks bridging technological domains (e.g. computing models, knowledge representation, visualisation and graphics), information and archival sciences, and social and cognitive sciences (NoE).

f) Promoting the uptake of EC-funded research enabling the deployment of new ICT-based cultural and memory preservation services, leveraging the impact of associated national initiatives; identification of future 'Grand Challenges'; establishment of a pan-European network of living 'memory centres' for validations, demonstrations and showcases (CSA).

#### **Expected Impact**

- Significant advances in the ability to offer customisable access services to scientific and cultural digital resources, improving their use, experiencing and understandings;



Empowering users: an active role for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

- Reinforced capacity of organisations to preserve digital content in a more effective and cost-efficient manner, safeguarding the authenticity and integrity of these records;
- Significant reduction in the loss of irreplaceable information and new opportunities for its re-use, contributing to efficient knowledge production;
- Strengthened leading edge research in Europe through restructuring of the digital libraries and digital preservation research landscape; leveraged impact of research results.

Research themes	Digital preservation	Digital libraries	Cultural heritage
Target outcomes in the ICT Work Programme 2009-2010	a) Scalable systems and services for preserving digital content (IP)	c) Innovative solutions for assembling multimedia digital libraries (IP)	d) Adaptive cultural experiences (STREP)
	b) Advanced preservation scenarios: - b1/ complex objects (STREP) - b2/ intelligent systems (IP)		
	e) Interdisciplinary research networks (NoE) f) Promoting the uptake of EC-funded research (CSA)		

The table presents how the research themes are linked to the target outcomes and funding schemes.

### At a glance:

#### Objective

ICT-2009.4.1: Digital libraries and digital preservation

#### Funding schemes

- Small or medium-scale focused research actions (STREP)
- Large-scale integrating projects (IP)
- Networks of Excellence (NoE)
- Coordination and support actions (CSA)

#### Indicative budget

€ 69 million in total

- IP/STREP: € 56 million with a minimum of 50% to IPs and a minimum of 30% to STREPs
- NoE and CSA: € 13 million

#### Call

- ICT Call 6; submissions possible from
- 24 November 2009 to 13 April 2010

#### Web site

[http://cordis.europa.eu/fp7/ict/telearn-digicult/digicult\\_en.html](http://cordis.europa.eu/fp7/ict/telearn-digicult/digicult_en.html)

#### Contact

[info-digicult@ec.europa.eu](mailto:info-digicult@ec.europa.eu)



*Empowering users: an active role  
for user communities*

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

**Wednesday 16th December, Plenary Session**  
*Invited lectures*



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## **Jill COUSINS**

### **Europeana: of the user, for the user**

Europeana is an online portal that contains 6 million digitised objects from Europe's museums, libraries, archives and audiovisual sources. It connects you to Europe's cultural heritage by providing access to items from over 1,000 heritage collections, including:

- Images - paintings, drawings, maps, photos and pictures of museum objects
- Texts - books, newspapers, letters, diaries and archival papers
- Sounds - music and spoken word from cylinders, tapes, discs and radio broadcasts
- Videos - films, newsreels and TV broadcasts

Some items in Europeana are world famous, from cultural institutions like the Rijksmuseum in Amsterdam, the British Library in London and the Louvre in Paris. Others are hidden treasures. The collection is continually being expanded and will reach 10 million objects in 2010.

From the beginning, Europeana has concentrated on creating a user-centric design. Focus groups, surveys, log-file analysis and user profiles are just some of the techniques used to ensure the site satisfies user needs.

Europeana is supported by funding from the European Commission and member states. Originally known as the European digital library network – EDLnet – it is a partnership of 100 representatives of heritage and knowledge organisations and IT experts from throughout Europe. They contribute to the Work Packages that are solving the technical and usability issues.

The project is run by a core team based in the national library of the Netherlands, the Koninklijke Bibliotheek<sup>1</sup>. It builds on the project management and technical expertise developed by The European Library<sup>2</sup>, which is a service of the Conference of European National Librarians<sup>3</sup>.

Overseeing the project is the EDL Foundation<sup>4</sup>, which includes key European cultural heritage associations.

---

<sup>1</sup> <http://www.kb.nl>

<sup>2</sup> <http://www.theeuropeanlibrary.org>

<sup>3</sup> <http://www.nlib.ee/cenl/about.php>

<sup>4</sup> <http://version1.europeana.eu/web/guest/edl-foundation>



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## **Rossella CAFFO**

### **Digital Libraries programs in Italy**

My paper intends to explain the strategy at national and European level adopted by the Italian Ministry of Cultural Heritage and Activities in order to coordinate digitisation initiatives and to promote access to cultural information.

The Italian Ministry's strategy is focusing on the integration of existing information systems, on the recovery of databases not in line with the current international standards, the creation of web sites and of cultural portals.

The Central Institute for the Union Catalogue of Italian Libraries and for Bibliographic Information (ICCU), directed by me, is the institute coordinating all cataloguing and documentation activities carried out by Italian libraries.

The ICCU promotes and develops programmes, studies and scientific initiatives concerning cataloguing, inventories and digitisation of the bibliographic and documentary heritage preserved in State libraries and other Italian public and private institutions.

The Institute coordinates the development and dissemination of the librarian cultural heritage in order to define a national system of services.

Through the Technologic Observatory of Cultural Assets and Activities (OTEBAC) the ICCU also promotes the harmonisation of digitization standards and the management of digital resources across all sectors of cultural heritage.

The ICCU:

- coordinates, promotes and manages the catalogue, the network of the National Library Service (SBN), the InterLibrary Loan services and documents' delivery;
- coordinates, promotes and manages the national databases concerning the census and bibliography of manuscripts, antique books and the Italian library database;
- manages the procedures of local applications in accordance with the National Library Service;
- promotes and coordinates the production of national regulations and the dissemination of international standards and cataloguing rules, assuring the uniformity of the catalogue and the production of the bibliographic control;
- participates at international level in the production and updating of bibliographic standards and formats;
- promotes and coordinates the production of standards and guidelines for digitisation of bibliographic and documentary heritage, with respect to the phases of archiving, management, conservation and access to digital resources;
- coordinates the monitoring of digitisation projects and oversees the publication and dissemination of digital resources, integrating them with SBN;
- manages education and training activities in its own areas, offering traditional and e-learning courses;
- takes part into international projects concerning the dissemination of information and digitisation of cultural and scientific heritage, such as ATHENA, CERL, DC-NET, DPE, MICHAEL, TEL, Europeana;
- carries out editorial activities.

The activities, researches and technical actions are promoted by ICCU in accordance with the general directives of the Ministry of Cultural Heritage and Activities and of the General Directorate for Library Heritage, Cultural Institutes and Copyright.

### **What follows is a summary presentation of the most important projects monitored by ICCU.**

Internet Culturale: digital catalogues and resources of Italian libraries

(<http://www.internetculturale.it>)

Internet Culturale is a multilingual portal (in Italian, English, French and Spanish) that allows users to access documents and digital resources of Italian libraries and to find information on their activities.

Through the portal users can:

- search within the national online loan service;
- visualize and download images of linked digital collections;
- access to the InterLibrary Loan service;
- search in the digitised historical catalogues of Italian public institutions;
- search in the Italian libraries database.

In the section devoted to special catalogues users can find information on:

- Italian editions of the 16th century (EDIT16);
- manuscripts in Latin alphabet owned by the libraries participating the Manus national census;



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

- the bibliography of manuscripts produced by the Institutions joining the BibMan project;
- the description and digital reproduction of Greek palimpsests.

Internet Culturale also allows browsing, with the help of three-dimensional technologies, of animations, hypertexts, through virtual exhibitions dedicated to famous characters (e.g. Svevo, Verdi, Totò) and tourist cultural itineraries or gourmet journeys.

Internet Culturale also devotes a large section to musical heritage allowing the search by institutions, projects, collections, authors and by rare items such as autographs (there are for example 272 items on Donizzetti, 70 items on Puccini, 25 items on Bellini), by musical itineraries.

Contributions come from State Libraries and Monumental Libraries, Libraries of Conservatories, Museums, Civic Institutions, State Archives, and important musical institutions.

The digital objects are almost all preserved in the ICCU repository with a number of approximately one and a half million images that form one of the richest online database.

The digitisation of the historical catalogues of libraries took place in the framework of the Italian Digital Library Project and includes the scanning, in image format, of over 200 Italian public library catalogues, making available about 7 millions images accessible through Internet Culturale.

This collection gathers digital reproductions of catalogues, in cards or in volumes, organised in different ways (alphabetic, topographic, systematic, mixed) and relative to various materials (printed editions, manuscripts, special materials). The headings were taken in the form in which they appear in the catalogue (without standardization) in order to respect the "historical authenticity" of the catalogues.

Information about libraries and about each digitised catalogues (history, organization, special ways of searching) are also available.

Internet Culturale is integrated with Culturaitalia, the aggregator of contents and access point to databases, websites and digital collections involving all sectors of cultural heritage.

### **Cultura Italia (<http://www.culturaitalia.it>)**

Culturaitalia proposes a guided access to the world of Italian culture. Thanks to innovative IT solutions, the portal gathers and organises millions of information-related elements on the resources that make up the country's vast cultural universe, and makes them available to all web users.

The data on cultural resources are not produced by the Portal, but are provided directly by the entities who own and manage the resources. All the operators in the cultural sector – public administration and private companies – upload only the "metadata", that is, the descriptive information of the resources in their possession. Culturaitalia offers to users the opportunity to consult and search information on Italian cultural resources within a single site.

The user, through the Portal, accesses a database of "metadata" which gathers and organises the information arriving from all providers participating in Culturaitalia. Users can discover resources of all types which make up the country's extensive cultural heritage (museums, photographs, libraries, archives, galleries, exhibits, monuments, videos, discs, etc.), can carry out searches for scientific research or just out of simple curiosity. Culturaitalia is an "open" system in that it grows and develops in sync with the new information contained in the resources that enrich its database. The Portal does not, per se, contain resources on the Italian cultural heritage, but rather proposes itself as a starting point for a guided search towards other sites.

The Portal offers a service to users, who will have at their disposal a single location from where to start their own search itineraries online, in terms of Italian culture, and the operators in the field, who can take advantage of a high-quality showcase to promote their own contents. Once the resources of interest are located in Culturaitalia, the user can consult them directly at the data source, by heading to the provider's site or by contacting them via other channels, to complete their process of analysis and understanding.

The Portal is an answer to the needs of an expert public as well as to the needs of the general public. Culturaitalia offers to specialised users, such as students, researchers, and those employed in the cultural sector, the opportunity to carry out targeted searches that correspond to very specific interests using a very advanced software. For non-specialised users, such as citizens and tourists, the Portal can stir curiosity and offer opportunities to discover or find out more about cultural resources available in the territory, mostly thanks to its editorial content (thematic itineraries, articles, highlights, events, columns) published to spotlight the stores of "metadata" present in the website database.

The project is promoted and managed by the Ministry of Cultural Heritage and Activities (MiBAC) with the scientific consultation of the Scuola Normale Superiore of Pisa. Culturaitalia is supported by many organisations and institutions that belong to the world of Italian culture and provide the relevant information, that is to say the real "primary resource" of the Portal. Precisely because of its ability to integrate in one system the information-related elements of many different entities, Culturaitalia is a leading project in Europe and has been used as a reference for many other countries hoping to promote similar initiatives.



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

CulturalItalia is the national aggregator which feeds Europeana, the European Digital Library. Summarizing, CulturalItalia aggregates contents at Italian level, while Europeana aggregates contents at European level. Both projects use the same standards and share the same approach and philosophy. Both projects involve all heritage sectors.

MICHAEL, Multilingual Inventory of Cultural Heritage in Europe, <<http://www.michael-culture.org>>

The MICHAEL and MICHAEL Plus projects were funded through the European Commission's eTen programme, to establish a new service for the European cultural heritage.

The MICHAEL project was a partnership between France, Italy and the UK to deploy a cultural portal platform that was developed in France. MICHAEL Plus then extended the MICHAEL project to the Czech Republic, Finland, Germany, Greece, Hungary, Malta, the Netherlands, Poland, Portugal, Spain and Sweden. The two projects were closely aligned. The projects focussed on the integration of national initiatives in digitisation of the cultural heritage and interoperability between national cultural portals to promote access to digital contents from museums, libraries and archives.

The projects have established this international online service, to allow users to search, browse and examine descriptions of resources held in institutions across Europe. We hope that the technical standards and sustainability model that we have established for the project will mean that more countries will provide their contents to the portal in future times.

Through the multilingual MICHAEL service people are able to find and explore European digital cultural heritage materials by using the Internet.

MICHAEL's objective were:

- A European cultural heritage inventory, available to all and providing access to cultural heritage resources.
- Sustainable management for the project to continue.
- Endorsement and implementation at a national government level, in order to underpin further funding as required.
- A methodological and technical platform, which makes it easy to add new national instances of MICHAEL, thus improving the content and user bases.

The technical results of the MICHAEL project can be listed as follows:

- The MICHAEL data model for multilingual digital cultural heritage inventories
- An open source technical platform for national instances built on Apache Tomcat, Cocoon, XtoGen, XML etc.
- Interoperability protocols for national instances to contribute data to the European service
- European MICHAEL search portal
- Methodology and model which is easy to deploy in additional countries.

MICHAEL currently includes some 10.000 digital collections from 4000 cultural institutions in Europe representing millions of di data.

The MICHAEL Consortium (19 states and 40 partner) has set up an international association known as MICHAEL-Culture AISBL, to grant sustainability and to allow for further service developments. MICHAEL-Culture is a member of the European Digital Library Foundation, that manages and monitors the implementation of Europeana.

The Italian instance of MICHAEL is interoperable with CulturalItalia.

When considering experiences at European level, I wish to illustrate two other projects coordinated by ICCU.

ATHENA (Access to cultural heritage networks across Europe)

(<http://www.athenaeurope.org>)

In November 2008, the 'Network of Best Practice' was launched. This is a new project known as the eContentplus programme developed by the MINERVA network.

Its partners come from 20 EU Member States together with 3 non-European observers. 109 major museums and other cultural institutions are directly associated with the project, and 20 European languages are represented. It is coordinated by the Italian Ministry of Cultural Heritage.

ATHENA's objectives are to:

- Support and encourage the participation of museums and other institutions not yet fully involved in Europeana;
- Produce a set of scalable tools, recommendations and guidelines, focusing on multilingualism and semantics, metadata and thesauri, data structures and IPR issues. These will be used by museums to



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

support internal digitization projects and to facilitate the integration of their digital content into Europeana;

- Identify digital contents that are present in European museums;
- Contribute to the integration of the different sectors of cultural heritage with the overall objective to merge all these different contributions into Europeana. This will be carried out in cooperation with other projects more directly focused on libraries and archives;
- Develop a technical infrastructure that will enable semantic interoperability with Europeana.

ATHENA will:

- Bring together relevant stakeholders and content owners from all over Europe;
- Evaluate and integrate standards and tools for facilitating the inclusion of new digital content into Europeana;
- Enable the user of Europeana to have a complete experience of European cultural heritage;
- Work with existing projects (Europeana, and Michael are both present in ATHENA);
- Develop links and joint activities with other relevant projects in the Europeana 'cluster' (for example EuropeanaLocal).

DC-NET - Coordination action contributing to European Research Area Network (<http://www.dc-net.org>)

DC-NET is an ERA-NET (European Research Area Network) project, financed by the European Commission under the e-Infrastructure - Capacities Programme of the FP7.

The main aim is to develop and to strengthen the co-ordination of the public research programmes among the European countries, in the sector of the digital cultural heritage.

This scope will be pursued by the participating Ministries of Culture by endorsing a Joint Plan of Activities, to be initiated already during the project time-frame, through a wide and intensive programme of seminars, workshops, meetings and Presidential conferences dedicated to the encounter of the digital cultural heritage sector with the technological research and the e-Infrastructure providers in Europe.

The main objective of the DC-NET project is to develop and to strengthen the co-ordination among the European countries of public research programmes in the sector of the digital cultural heritage. The project will integrate the research capacities of the participant member states, will identify their communalities and will valorise existing programmes and projects in order to initiate the deployment of a wide and comprehensive European e-Infrastructure that will increase the research capacities of the digital cultural heritage community.

The DC-NET project will contribute to the coordination of the research priorities of Ministries of Culture, their Agencies and other cultural bodies (museums, libraries, archives, audiovisual, archaeological sites, etc.) across Europe in the area of the eInfrastructures targeted to the digital cultural heritage. The project will coordinate the manner in which cultural actors can and should engage with national and European e-Infrastructures to generate innovative services, tools and data sets to support the research of multidisciplinary communities. A programme of seminars, workshops, meetings and conferences will involve all the relevant stakeholders. A plan of joint activities for e-Infrastructure-enabled research in the sector of digital cultural heritage will be generated and the joint activities will be initiated.

ICCU's backoffice in the field of digitisation currently is mainly devoted to three activities:

- Development of the campaign "Join CulturalItalia" targeted to cultural institutions.

The workflow consists in five main phases: a) Identification of databases to be harvested; definition of the agreements between MiBAC and the cultural institution; 3) definition of the data amount and level of description to be made available to CulturalItalia; 4) the metadata harvesting; 5) the quality check.

- Development of institutional cultural websites

MiBAC is continuing to diffuse all guidelines and publications produced by the MINERVA project in the field of cultural websites.

Furthermore, it promotes "Museo & Web", the open source content management system developed by MiBAC in the framework of MINERVA.

Currently MiBAC institutions have developed more than 300 websites; half of them made with Museo & Web.

In the last years, institutions changed the way they perceive websites.

According to Maria Vittoria Marini Clarelli, director of the National Gallery of Modern Art (GNAM), a museum's website represents a tool to prepare/complete the visit; a way to facilitate the cultural mediation; a free of charge way to access; a window. Feedbacks proved that users expect that an institutional website should be trustworthy and reliable, that contents should be maintained up to date, that the website should grow in time.



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

A well done website changes the institution's way to communicate; produces more service and qualitative information; stimulates more internal communication and coordination, as well as benchmarking with other institutions.

- Supporting institutions in the activities of digitisation

As regards digitisation, there are still several open issues: high costs, insufficient know-how (with the exception of librarians), insufficient knowledge of resource description standards and harvesting protocols, a general need of training, often a low quality of metadata, products available only off-line (CDs-DVDs), a generalised use of proprietary software.

Currently ICCU set up an Italian cross-domain working group for the harmonization of ALM sectors metadata standards in a cross-domain perspective. The Working group is composed of ICCU experts working together with the Central Institute for cataloguing and documentation (ICCD) and the General Directorate for Archives of the Italian Ministry of Cultural Heritage.

But, another real barrier to the free dissemination of contents are Copyright and Intellectual Property Right (IPR). This issue is common to several European Member States.

MiBAC, in accordance with the European strategy, will soon adopt 'Creative Commons' licences in the framework of CulturalItalia.

The European project ARROW (<http://www.arrow-net.eu>), coordinated by AIE (Associazione italiana editori, Italian Publishers Association), is trying to clarify, at a European level, issues concerning orphan works and out of print works, in the field of publishing.

As far as multilingualism is concerned, the barrier of European linguistic diversity is dramatically evident regarding common access points to distributed databases. Investments are insufficient and tools available are not yet suitable.

Partnership with privates is another issue to be investigated. It could be a way to guarantee the online presence of cultural works under copyright (books) or intellectual property right (i.e. contemporary artworks). Of course agreements must be checked carefully by the institutions involved. In fact, after Google's recent digitisation proposals to several European libraries, national governments decided to share a common position in order to guarantee public access and reuse of contents susceptible of Google digitisation.

On the contrary, many efforts are carried on in the field of digital preservation. ICCU made digital repositories available for institutions, while the Central National Library of Florence, the Central National Library of Rome, and the National Library of Venice are working in close cooperation with Fondazione Rinascimento Digitale at the initiative Magazzini Digitali ((Digital Repositories) <<http://www.rinascimento-digitale.it/magazzinidigitali.phtml>>).

An important aspect to be taken into consideration are users. Institutions are used to consider the users of cultural websites, digital libraries and portals as ALM professionals (librarian, archivist, curators, etc.) scholars and experts, information

scientists, students, tourists, creative or just curious people.

Institutions should take the users' age and the distinction between digital immigrants and digital natives into account.

Digital immigrants are people who are "old" enough to have attended the world of "yesterday" but "young" enough to have lived in the world of "today". So far they played a central role in knowledge and culture building and in developing the Internet. Generally speaking, up to now, websites, digital libraries and cultural portals have been conceived by digital immigrants.

Digital natives are born after 1980. They study, work, write and interact in a different way from their "antecedents"; they read blogs instead of newspapers; often they only get to know each other on the Internet; they have never written a letter, but they chat or text friends; many of them have never got in a library, they download music illegally, believing that it's legal; they handle easily and skillfully digital contents creating new ones. Summarising, the main aspects of their lives are mediated by digital technologies.

Is the institutional web ready to satisfy the digital natives' needs? It is necessary to monitor and analyse the audience behaviour in order to get responses to this issue.

Recently CulturalItalia launched an online survey to measure the users' satisfaction, based on the Minerva Handbook on cultural web user interaction. The results will be published in spring 2010.

I wish to end my speech bringing at your attention a successful story: the experience of SBN, the Italian library network.

SBN (<http://www.sbn.it> - Servizio bibliotecario nazionale) is an infrastructure of national services for users promoted by the Ministry of Cultural Heritage and Activities, coordinated by ICCU and acting in cooperation



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

with Regions and Universities. State libraries as well as local, university and private libraries operating in different sectors are linked to the SBN.

The libraries participating in SBN (approximately 4000) are organised in 69 Nodes distributed across the national territory, connected to a central system (the Index), which sets up the general catalogue of libraries belonging to the network.

In 2002, with the launching of the Index evolution project, the rationalisation, integration and renovation of the Index central database was developed and the opening to other systems and management of different levels of cooperation was also envisaged.

The main aims of the project were:

- the technological renewal of hardware and software
- the opening of the SBN Index to management systems of a non-SBN library using the most widespread bibliographic formats (UNIMARC, MARC21);
- the management of diversified levels of cooperation;
- the development of new activities such as for example derived cataloguing and cataloguing of special materials;
- the development of government and monitoring functions of the system and of the increase of the databases.

Since 1997 the Index has been available to users through the OPAC (Online Public Access Catalogue) system that allows access to the contents of the catalogue, with research methods that are user-friendly and articulated, and to use its connected services.

The OPAC database currently contains about 11 millions of different pieces of information and 50 millions of localisations; it can be consulted through two interfaces (<http://opac.sbn.it> and <http://www.internetculturale.it>) with approximately 42 millions searches carried out in 2009.

Other important features of the OPAC SBN system across the Web are:

- the service of SBN ILL Interlibrary Loan;
- the integration with the local OPACS;
- the access to the cards of the identified libraries ;
- the presentation of search results in various formats among which UNIMARC and USMARC formats;
- the UNIMARC export of individual bibliographic items;
- the possibility of search and presentation for authority entries regarding authors included in the SBN Index (currently 36.000);
- the possibility to operate as a Z39.50 client and therefore to interrogate other Z39.50 catalogues at national and international level.

All the Italian initiatives mentioned in my speech have a common distributed approach, share coordination structures at local, regional and national level, have close links with national digitization strategies, benefit from the active participation of hundreds of cultural institutions (at all levels and in all sectors) and, last but not least, have in general a cross-domain approach with museums, archives and libraries.

All these efforts have generated a number of benefits for all involved stakeholders: local administrations, institutions and end-users.

The modalities to access information have exponentially increased, interoperability is undergoing rapid changes, local initiatives are enhanced thanks to new scenarios.

But many efforts are still necessary: more coordination and cooperation among stakeholders; more training; advocacy campaigns for funding; more cooperation among cultural institutions and research institutions.

At European level, the new Commission ambition is to realise a European digital agenda including a legislative action. The aim is to create in Europe "a modern and legal platform, competitive and "consumer friendly" for a unique market of online creative content".



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## **Stefano VITALI**

### **The SAN Portal: a common gateway to Italian archival resources on the Web *The National Archives System***

During the second half of the 20th century, in Italy as well as in many other countries of the developed world, the archival landscape underwent deep changes that helped reshape its territorial organization and the balance between its various institutional components. At the base of those changes are some typical phenomena of contemporary society on which archive literature has widely dwelled: the advent of new historiographic trends, addressed at documentary sources that were generally neglected or undervalued up to that time; the interest in recovering individual and collective memories, as well as local and territorial traditions; consequently, the new attention in archives as vehicles of memories and traditions and the rise of new expectations in them from an ever-growing, culturally and socially heterogeneous public; the increasing importance and activism of local and regional institutions and of various cultural and social organizations who headed up a crisis in the traditional balance between central and local powers, favoring the second as opposed to the first.

In Italy the outcome of these changes was the gradual exhaustion of the historical records concentration and keeping model established after the Unity, which was based on the centrality of a network of provincial State Archives, directed by State central administration located in Rome. A different model, gradually affirmed for the past few decades, is based on the coexistence of many archival holders (archives of municipalities, provinces, regions, cultural institutions, research centers, business archives, etc.) and on the development of various initiatives for collecting, describing and promoting archives, supported by a large number of public and private bodies.

This dissemination of archival institutions and initiatives presents many problems of connection and coordination and urges archivists and State, regional and local administrators to establish new forms of collaboration, which are particularly important in a phase like the current one where financial resources continue to decrease and the risk of dispersing them without lasting outcomes is increasing. Therefore, as claimed by many parts, it is urgent developing sustainable models of archival polycentrism capable of improve its positive features, while limiting the most critical, not to say negative, ones, such as an excessive dispersion and fragmentation of archival institutions and projects, which probably users of archives find unreasonable and which, at the same time, risk to put in danger the preservation of the extremely rich and precious Italian archival heritage.

The Second National Conference of Archives, held in Bologna from 19 to 21 November 2009, was dedicated to the need to “make a system” for governing the archival polycentrism. It was an important moment for debates and proposals where representatives from State archives administration, regions, local authorities, cultural institutions and other public and private organizations discussed the need for coordination and cooperation on different fronts: from the construction of common repositories for archives to common strategies for digital preservation, from archival education to initiatives for promoting and publicizing archives. The result was the design of a National Archives System (SAN) to be built in the near future with one’s own national government bodies, regional coordination committees and common repositories and archival services at the local level<sup>1</sup>.

### **The National Archives Portal and the Archives Resource Catalog**

The core component of the National Archives System will be a web portal, which has momentarily been given the name *Portale Archivistico Nazionale* (National Archives Portal) or PAN, which should present itself as the integrated access point to national archival resources, irrespective of the juridical status and affiliation of the institutions or organizations that developed them. The Portal, whose planning and realization began in 2009 by the Directorate General for Archives of the Ministry for cultural heritage and activities, will be divided into multiple sections and will carry a lot of complex contents, such as:

- an “Archipedia”, which is a sort of encyclopedia of archive definitions and concepts compiled in cooperation with the portal’s users and including a glossary which will provide simple and concise explanations of archival technical terms for novices and unskilled users ;
- a database of bibliographic resources related to Italian archives;
- research guides, virtual tutorials and other materials for different typology of audience (for example: teachers, students, genealogists, historians, etc.), including novices and unskilled ones;

<sup>1</sup> See the Conference web site <<http://www.conferenzanazionalearchivi.beniculturali.it/index.php?it/1/home>>, where all the preparatory documents and the final one can be accessed.

- specific thematic sections or sub-portals dedicated to describe and make available different typologies of records (business, genealogical, cartographical and fashion archives etc.);
- editorial and multimedia contents, including news, virtual exhibits, photo galleries and so on and so forth for illustrating the multiple aspects of the world of archives and records;
- a digital archive that allows access to digital reproductions of fonds and series published on web sites of local, regional or national archival institutions;
- Web 2.0 tools for communicating with the users of the portal, allowing them to collaborate in creating its contents and offering them the possibility to build communities on specific topics and research projects.

The central component of the PAN will be the Catalogo delle Risorse archivistiche (Archival Resource Catalog) or CAT that proposes to be a coordinating and integrated access point to archival descriptions stored on archival databases and systems developed at the regional and local level, respecting their autonomy and specificity.

In fact, one way that the archival polycentrism mentioned above has been manifested in Italy, was the development of many software and systems for producing digital descriptions of archives and for publishing those descriptions on the Web. Systems have been developed at the national level by the State archives administration (such as the General Guide to the State Archives<sup>2</sup>, the State Archives Information System or SIAS<sup>3</sup>, the Unified System of Archive Supervising Agencies or SIUSA<sup>4</sup>, the Mediterranean Historical Multimedia Archives<sup>5</sup>; by some provincial State Archives (the State Archives of Florence<sup>6</sup>, Milan<sup>7</sup>, Bologna<sup>8</sup>, Naples<sup>9</sup>, Venice<sup>10</sup>; by some regions (like Lombardy<sup>11</sup>, Emilia-Romagna<sup>12</sup>, Piedmont<sup>13</sup> or Umbria<sup>14</sup>; by other territorial entities (like the Province of Trento Historical Archives<sup>15</sup>; by individual cultural or political institutions (like the Piedmont Institute for History of the Resistance and the “Giorgio Agosti” Contemporary Society<sup>16</sup>, the Giangiacomo Feltrinelli Foundation<sup>17</sup>, The Senate of the Republic<sup>18</sup>, The Chamber of Deputies<sup>19</sup> and many others) or by groups of ‘federated’ cultural institutes (the Institute of the Resistance Network<sup>20</sup> or the project Archives of the 1900s<sup>21</sup>).

The presence on the Web of these multiple systems is not only the result of the way in which historical archives have been computerized in Italy, nor is it just the result of the archival polycentrism mentioned above. Actually the dialectic between ‘local’ and ‘national’ systems of archival descriptions reflects a deeper logic that has to do with the double meaning associated with archives today. On one hand archives are products of specific historical and geographical contexts and hence, sources for knowledge of their history and vehicles of specific memories and identities; on the other, they are tout court cultural heritage, thus carriers of universal significance and values that cannot be closed within restricted territorial areas, but must acquire a national, and possibly international, visibility. Therefore, not only is this multiplicity of systems not casual, nor does it constitute a sort of limits to quickly overcome, possibly through their centralization into one single system, but contrarily, it represents an undoubtedly valuable resource. Nevertheless, it is true that in the previous years there has been a growing need to establish connections, data exchanges, increasing levels of interoperability

2 See its first version at <<http://www.maas.ccr.it/h3/h3.exe/aguida/findex>> and the new one, published online in 2009 at <<http://guidagenerale.maas.ccr.it/>>.

3 See at <<http://www.archivi-sias.it/>>.

4 See at <<http://siusa.archivi.beniculturali.it/>>.

5 See at <<http://www.archividemediterraneo.org/portal/faces/public/guest/>>.

6 See the Florence State Archives Information System or rSiASFI at <<http://www.archiviodistato.firenze.it/siasfi/>>.

7 See the Online Guide to the State Archives of Milan at , <<http://archiviodistatomilano.it/patrimonio/guida-on-line/>>.

8 See the Archival heritage of the State Archives of Bologna at <<http://patrimonio.archiviodistatobologna.it/asbo-xdams/>>.

9 See the Archival heritage of the State Archives of Naples at <<http://patrimonio.archiviodistatonapoli.it/xdams-asna/>>

10 See the Online Guide to the State Archives of Venice or SiASVe at <<http://www.archiviodistatovenezia.it/siasve/cgi-bin/pagina.pl>>.

11 See the section dedicated to the description of archives on the cultural heritage portal of Lombardy Region at <<http://www.lombardiabeniculturali.it/archivi/>>

12 See the Ibc Archives Portal at <<http://archivi.ibc.regione.emilia-romagna.it/ibc-cms/>>

13 See Guarini Web for archives at <<http://www.regione.piemonte.it/guaw/MenuAction.do>>

14 See .DOC- information System for the archives of the Umbria Region at <<http://www.piau.regioneumbria.eu/default.aspx>>

15 See the online inventories on the web site of the Historical Archives of Trento Province at <[http://www.trentinocultura.net/catalogo/cat\\_fondi\\_arch/cat\\_inventari\\_h.asp](http://www.trentinocultura.net/catalogo/cat_fondi_arch/cat_inventari_h.asp)>

16 See ArchOS. Integrated System for the archival Catalogs at <<http://metarchivi.istoreto.it/>>

17 See Foundation online archives at <[http://www.fondazionefeltrinelli.it/feltrinelli-cms/cms.find?flagfind=quickAccess&type=1&munu\\_str=0\\_6\\_0&numDoc=95](http://www.fondazionefeltrinelli.it/feltrinelli-cms/cms.find?flagfind=quickAccess&type=1&munu_str=0_6_0&numDoc=95)>

18 See the Online Archives Project of the Senate of the Republic which includes descriptions and digital reproductions of archival fonds held also by other institutions at <<http://www.archivionline.senato.it/>>

19 See web site of the Historical Archives of The Chamber of Deputies at <<http://archivio.camera.it/archivio/public/home.jsp?f=10371>>

20 See the Guide to Historical Archives of the Resistance Institutes at <<http://beniculturali.ilc.cnr.it/insmli/guida.HTM>>.

21 See Archives of 1900 s - Memory on the Net at <<http://www.archividelnovecento.it/archivinoventocento/>>.

between local, regional and national archival systems so that a great deal of reflections and discussions has taken place on how to create fruitful forms of collaboration<sup>22</sup>.

By developing integrated access tools and offering essential information on the nation's archival heritage, CAT wants to build an answer to such a need and represent a tool for joining many existing systems together without substituting them; on the contrary it gives them greater visibility and enhances their specific characteristics. An operation like this is possible - even with the diversity of the software tools used and of some aspects of descriptive formats - since the systems developed in recent years share the same conceptual model and a common adoption of the international standards of archival description. These systems are generally based on an architecture that, besides the descriptions of archival materials created in conformity with ISAD (G), include the separate description of creators (corporate bodies, families, persons) in accordance with ISAAR (CPF) and the description of custodians of archives (archival institutions, but sometimes other entities or families and persons as well), according to the International Standard for the Description of Institutions with Archival Holdings (ISDIAH).

Therefore, the CAT will sketch a general map of the national archival heritage capable of providing an initial orientation to researchers and guide them towards more informative resources available in the systems that will participate in the National Archives Portal. It will contain descriptive records of custodians, of fonds or archival aggregations, finding aids and creators. It will be populated and updated through procedures that privilege procedures of data harvesting based on the OAI-PMH protocol. Other ways of importing will not be excluded, as for example the upload of XML files in a specific area of the Portal or the direct entering data into the CAT database, using an ad-hoc on-line interface. The purpose of these multiple implementation procedures is to let even the less technologically equipped holder of archives participate in the project.

Custodians of archival materials will be univocally identified and essential information on each of them will be provided in order to produce an authority list of all entities, institutional and non-institutional, that in Italy hold and provide access to historical archives. In addition to the essential identification data of each custodian (name, location) and a brief description of its history and current status, the availability of a reading room with regular opening hours and a reference service provided by skilled archivists will be indicated. Description and identification data, if available, will be acquired directly by the systems that participate in the PAN. The portal's editorial staff will correct, update and standardize the descriptions, if necessary, in order to provide users with reliable and up-to-date information. The CAT record for each custodian will contain, along with a hypertextual link to its web site, links to other descriptions which can be found in any of the systems participating in the PAN.

Regarding archival aggregations, the CAT will include the highest levels of description of each archive (fonds, or groups of fonds). It will also include lower levels of description such as sub-fonds or even series if created by specific creators, different from those of the fonds which they belong to.

The CAT will provide also concise information on the availability and main features of finding aids, existing on paper or in digital formats, for the archival aggregations described. An appropriate hypertextual link will address users to the digital ones directly accessible on the Web.

The selection of the information elements to be included in the CAT database for describing archival aggregation, creators, finding aids was prompted by a sort of principle of subsidiarity and economy and aimed to identify just those elements which are really essential. For wider and deeper descriptions, users will be addressed to the harvested systems. Consequently, only the elements considered mandatory according to the international descriptive standards together with few others regarded as such in Italian archival tradition have been included. A maximum number of characters in open-text fields will be provided in order to avoid redundant information. The description imported into CAT database from the provider systems will be published without any correction or modification.

Each CAT record will contain a direct link to the corresponding record in the original system, from which the data has been imported. Following the link the user will access the complete description of the archival aggregation, creator or the whole finding aid - if it exists - in its original context. The descriptions of archival aggregations will also be linked to relevant digital reproduction projects which will be made available on the Portal.

Each imported record is to be connected to a CAT record describing the original system in order to provide information about the provenance of the data. Moreover, since it cannot be excluded that the same archival aggregation, finding aid or creator is described in more than one of the archival systems exporting their data into the CAT database, provision of information about original systems will help users to correctly interpret and contextualize such multiple descriptions of the same entity.

<sup>22</sup> See for the example *Verzo un Sistema Archivistico Nazionale?* special issue of "Archivi e Computer", XIII (2004), 2, edited by Stefano Vitali.

Nevertheless, in order to provide users with information of increasing quality and for adding value to records imported into the CAT database, the descriptions of creators will be linked to an authority file of creators which will be progressively implemented by the central staff of the Portal. This authority file should be not only the primary access point for research and navigation in the CAT, as well as the tool for connecting records imported from the provider systems, but also a national point of reference for identifying corporate bodies, persons and families and recording their authority names and descriptions. Therefore, in the future, local systems will not have to produce their own description of creators, but, if they would like, they could directly refer to the national authority file for creators published on the Portal. Finally, this authority file will represent a bridge towards analogous authority files developed in catalogs and descriptive systems of other cultural heritage domain, such as the National Library System.

### **SAN standards definitions and characters**

The CAT architecture, the descriptive elements to be included and the exchange formats between the CAT itself and the systems which will provide the data have been developed by some ad hoc workgroups appointed at the beginning of 2009 by the State-Regions Joint Technical Commissions for defining archival standards. Representatives from different archival institutions and administrations and from various regions took part in the discussions. Besides metadata sets for the various entities described in CAT<sup>23</sup>, the workgroups defined export-import formats, protocols and procedures and developed standards and methodologies for the preparation of the descriptions of corporate bodies, persons and families to be included in the authority file of creators. During the first few months of 2010 metadata sets will also be released for digital resources which will be made available on PAN, according to the same approach used in CAT. In the Portal's digital archives, thumbnails of the images and essential information, including hypertextual links to the original systems, will be imported, allowing users to research among the available digital resources, make a preliminary selection and then be directly addressed to the harvested systems for quality viewing of the digital reproductions of archival documents.

For exporting the descriptions from the existing systems to the CAT, an XML exchange format has been developed. It has been named "SAN exchange format". It is based on three schemas, each of one includes a subset of elements of the Encoded Archival Description or of the recently released Encoded Archival Context (Corporate Bodies, Persons, Families)<sup>24</sup>.

In particular, the schema for archival aggregations is based on EAD and include only elements from the <did> element. Due to some specific requirements of the SAN exchange format, many of the EAD elements used had to be partially adapted and modified. The schema for finding aids is based on EAD and includes only elements from the <header> element. Also in this case some of the elements have been adapted for meeting the specific requirements of the SAN exchange format. The schema for creators is based on EAC-CPF. In this case, the standard worked much better and no specific adaptation was required to develop a schema completely compliant with the characteristics of the SAN exchange format. Finally, a fourth ad hoc schema has also been developed for importing some essential information on custodians from the systems which can provide such information. As mentioned above these descriptions, after been imported, will be amended, standardized and updated by the editorial staff of the Portal<sup>25</sup>.

In developing the schemas, an approach was adopted that can be defined as "recordcentric". Every item description exported will generate just one XML record, uniquely identified by the identity code of the description in the original system. Relationships between items - even those of hierarchical nature between archival aggregations - will be explicitly recorded, making reference to the identity code of the description/s of the related item/s.

The XML schemas for the four entities (custodians, creators, archival aggregations, finding aids), integrated with control information necessary for correctly implementing the import procedures, have been accommodated into a whole export-import format, which is available to all those responsible for the existing archive descriptive systems who wish export their data into the National Archival Portal and contribute to the effort in constructing a single access point to Italian archival resources on the Web<sup>26</sup>.

23 See Sottocommissione tecnica per la definizione dei metadati [...], Tracciati descrittivi del CAT: soggetti, <[http://ims.dei.unipd.it/data/san/metadati/docs/2009-04-17\\_Documento-conclusivo-sui-tracciati-CAT.pdf](http://ims.dei.unipd.it/data/san/metadati/docs/2009-04-17_Documento-conclusivo-sui-tracciati-CAT.pdf)> (provisional url).

24 See relevant documentation on the web site of the project at <<http://eac.staatsbibliothek-berlin.de/>>

25 See Sottocommissione tecnica per la definizione dei metadati [...], Tracciati descrittivi e schema XML di esportazione-importazione dai sistemi aderenti al catalogo delle risorse archivistiche (CAT) Versione 1.1, (ottobre 2009), <[http://ims.dei.unipd.it/data/san/metadati/docs/2009-11-08\\_tracciati\\_export\\_import\\_1-1.doc](http://ims.dei.unipd.it/data/san/metadati/docs/2009-11-08_tracciati_export_import_1-1.doc)> (indirizzo provvisorio).

26 The overall schema and relevant documentation are provisionally accessible respectively at <<http://gilgamesh.unipv.it/cat-import/cat-import.xsd>> and <<http://gilgamesh.unipv.it/cat-import/cat-import.html> - id6>.



### **Conclusions: Why the PAN experience is important**

The realization of the National Archival Portal and its Catalog of archival resources is destined to represent an important turning point in the ways with which the Italian archival community has until now conceived and constructed its own presence on the Web. For the first time the community will have at its disposal a system built from the ground up through data exported from various providers. A system that wants to be completely placed on the inside of the Web's new horizons based on cooperation, interoperability and data reusability.

The Portal is going to change radically the Italian archival landscape, with significant impacts on the existing systems. The need to adapt, even if slightly, their data to make them fully compliant with the SAN exchange format will have inevitable fallbacks on the same systems that will be stimulated towards more elevated levels of homogeneity and standardization in order to allow wider and wider data reuse in contexts different from the original.

On the other hand, the open and cooperative model adopted for developing the SAN standards has already fulfilled largely satisfying and important results. Further consolidation of these results, which must derive from their wide use, will constitute an important premise for taking further steps forward in developing other more complex national exchange formats related to other aspects and components of archival description.

It is worth remembering that if standards and metadata are important because they make it possible for systems to exchange data, they are even more important because they force people and institutions to communicate.



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## **Gianbruno RAVENNI**

### **Cultural Heritage On-line. Empowering users: an active role for user communities**

The Tuscan Region policies for the spread of digital culture go further on the specific policies about cultural heritage and I think it's right talking about them, because are critical for creating the environmental conditions to develop digital technologies for cultural heritage.

This work has begun with the establishment of a network, which function was regulated according to the regional law num. 1, 2004: "Disciplina della Rete telematica regionale Toscana". In the October 2009 a second regional law has been approved, it is the number 54: "Istituzione del sistema informativo e del sistema statistico regionale". In addition to this other intervention programs have been put in place for developing the implementation of digital technologies in local administrations, in local health authorities and in the justice management.

Regarding the specific sector of cultural heritage, we have to mention the important role that Tuscan public Libraries are carrying out in the promotion of digital culture during the last ten years, with the construction of new Libraries, some of them beautiful, and with the growth of services that implement largely digital technologies for users community.

The new Tuscan Libraries are doing a very important work to promote the digital culture and the rights for all to benefit from these technologies, information and their services.

About cultural heritage sector, the Region and his partners have promoted and supported several projects: some of them are already operative, e.g. The Digital Newspaper Library, others are still in execution, like the project for the digitization of papery catalogues of historical local Archives. We have published on line significant databases such as the D.B. of the ancient book, the mediaeval manuscripts, the Tuscan Lexicon Atlas and so on.

We have recently presented the DANTHE portal, created in collaboration with the University of Florence. This portal allows to access to the on line data bases regarding the cultural heritage in Tuscany.

We looked for and found a constant cooperation with the Ministry for Cultural Heritage and Activities and its structures in Tuscany. The results of this collaboration are: the beginning of the regional pole SBN, the participation in the MICHAEL project and in other projects related to Archives (SIUSA), the "Carta dei vincoli", and an on line data base about all the architectural, archaeological and landscape constraints.

The cooperation and coordination of the projects are of primary importance and represent the first target to be assigned to public administrations and, at the same time, are necessary for the availability of suitable financial resources.

Finally it's very important the integration between public institutions that manage cultural projects, from the State to local administrations, and private subjects like the Fondazione Rinascimento Digitale, that today has organized this Conference, but also other cultural institutions, e.g. universities, research Centres and enterprises.



*Empowering users: an active role  
for user communities*

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

**Parallel sessions I**  
***Digital library applications & interactive Web***



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## **Anna Maria TAMMARO**

### **Digital library applications and interactive Web: from space to virtual place**

Introducing this session during the Conference Cultural Heritage Online I was looking for finding a reply to the following questions: How librarians and technologists can exploit the potential of emerging technologies? How can the cultural institutions transform the way digital libraries provide services and resources?

In trying to reply to these questions, after the conclusion of the session, I reflected on what the speakers have presented. They have described how emerging technologies can support the integration of different digital collections, can facilitate community building and extend connectivity to the ubiquitous user. However the speakers of the session have evidenced that there are challenges on existing delivery models of traditional cultural institutions, which have to change.

Two new roles were emerging from the presentations: a new role of digital libraries, a new role of users This two roles are two faces of the same coin: digital libraries are participating to a diffused culture of learning; users are actively engaged in creation, modification, and distribution of information objects in digital libraries as learners in a new virtual space.

Analyzing the different presentations, I can say that the digital libraries applications, with a focus on their users, move from the paradigm of cultural institutions as place to the paradigm of digital libraries as virtual spaces for learning.

#### **New digital library role**

Most of the speakers said that digital libraries should transcend the “search and access” approach, and serve as collaborative knowledge environments. The metaphor of the virtual space is very simple in technical term. The focus of the new role however is not on the tools and technologies, but on the changing roles of libraries supporting learning services. For example, the VKS (Virtual Knowledge Studio for the Humanities and Social Sciences) in Netherlands, focuses on the new scholarly practices, on the research methods, and on the ways of knowledge creation in the humanities and social sciences, stemming from the introduction of information and communication technologies into those scholarly fields.

Web 2.0 technology is a hot topic at the moment, and public librarians in particular are beginning to feel the pressure to apply these tools. Indeed, Web 2.0 has the potential to transform library services, but all the speakers have demonstrated that only if the policy and strategy for digital library services are ready to be transformed, the digital libraries can afford the new role. In a world where computing power, ubiquity and connectivity are increasing, the digital libraries could follow powerful new visions of their services for facilitating learning. This new digital library role brings cultural institutions into the cultural and social aspect of the technology.

Manzuch (Vilnius University) explained that the new role can be expressed as “communication of memory and involvement of the communities”. This means bridging cultural heritage to the users, communicating heritages using the technology and digital media, sharing memory as part of the community values and cultural conditions. The present services of cultural institutions how are interpreting the cultural heritage? And how the services of cultural institutions are communicating memory?

With the contribution from leading practitioners in all areas – including lecturers, librarians and e-learning technologists – the session has explored the strategic approaches which digital libraries in the world are following. New services are the new access channels and value added services which EuropeanaConnect is going to provide for users of Europeana; or the service of digitisation on demand of the project EOD; or the activities supporting publishing of the EUI Library. “Interaction with the users” is the most advanced functionalities of the new services: transactions are made online, the whole ordering process and service management process (from the point of view of the digital library) is online and also modern payment services such as credit card payment – where needed are integrated in the services.

The speakers of the session have reviewed the technological tools of the digital library setting, but they have especially recommended the institutional policies, the theoretical reflections and the business models that are needed to create a new strategy for digital libraries. Challenges encountered when designing a large-scale application of digital library for a diverse public audience have been explored, including digital rights management, user content moderation, and balancing customizability with simplicity of interface. Also, staff development needs have been stressed, which includes programmes of training courses for the staff of the cultural heritage sector as well as higher and further education, conferences and workshop on best practices for exploiting the potential of the Web.



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## **New users role**

How users can be involved in digital libraries? The first attempts made from digital libraries to include the active role of users in services have been described in the session: building networks of resources also produced by users, involvement of users in the selection of eBooks, participating to control the vocabulary, or putting data in the Web, collaborating in building repository of harvested resources. Users have been asked of leaving comments, annotating, making collaborative discussions, producing video content streams, cooperating with graphically enrich metadata visualisation.

For example, the British Library (BL) is involving users in digitisation and content creation. BL asks to users text correction, improving the quality of the text version, putting their annotations, delivering user-created guides, participating to the discussion board and to the collaborative creation of text extraction.

More expert communities of users are involved in discovering, assessing, organizing critically all the primary resources, or in doing critical editions, not only terminological distinctions, putting annotations and doing linguistic analysis with syntactical tools, realising domain ontology to mark the texts of the documents.

The speakers have evidenced that we do not need pharaonic projects but small scale projects, adapt to the moment and which fit in the different needs of the communities.

Some issues about the learner and teaching and learning interventions of digital libraries were also evidenced. Considering the enduring socio-psychological features of online interaction and the different user engagement, the diversity of user's community needs an interface design which is easy to access. User expectations want the accessibility of interface which should be focused on users intended purposes and the cultural institutions should not miss this opportunity of improving interaction and communication in the contemporary net-based social networks.

The most fundamental need in user's involvement is "trust": who is this person? The presentations gave a lot of information about user's generated content. Trust of communication and the expectations of an active role of users need a collaborative knowledge environment, as virtual space improving and facilitating learning.

## **Conclusion**

Cultural institutions are going back to their roots, to start from their foundations to create new delivery mode of their services. The convergence in the Web of cultural institutions is not enough, if they offer services with no difference with the previous "search and access" paradigm.

Interinstitutional cooperation needs to improve communication of memory and to create a virtual learning space. The focus on learning communities, scholars and other users move the cultural institutions services to the Web, moving from being a physical place to realising a virtual space, with the aim of promoting knowledge in extended services paradigms.

In conclusion the need of user's involvement is not new, the available tools are new and the functionalities can be highly extended, behind the traditional services. The projects in this session were able to use these new tools with creativity and the result was innovation.

## **Brian KELLY, Charles OPPENHEIM**

### **Empowering users and their institutions: a risks and opportunities framework for exploiting the potential of the Social Web**

#### **Abstract**

Following the initial excitement generated by Web 2.0, we are now seeing Web 2.0 concepts being adopted across the cultural heritage sector. Libraries, with their responsibilities for facilitating access to information resources and engaging with their user communities, have been early adopters of Web 2.0, and the term "Library 2.0" is now becoming accepted. Similar approaches are happening in the museums and archives sectors, with the terms "Museum 2.0" and "Archives 2.0" gaining currency.

But how should we ensure that the initial enthusiasms for use of Web 2.0 services and approaches become embedded within the organisation? And are cultural heritage organisations aware of the potential risks associated with making use of externally-provided services such as Facebook, YouTube and del.icio.us, including misuse of such services, associated legal concerns as well as the dangers of making use of services for which there may be no formal contractual agreements?

In this paper the authors argue that the cultural heritage sector needs to recognise that use of Web 2.0 providers does not necessarily provide an environment in which safe, secure and reliable delivery of services to the user community can be guaranteed. But rather than seeking to replicate successful Web 2.0 services in-house, we feel that we are in an environment in which cultural heritage organisations need to take a risk management approach to the use of networked services.

The paper describes a framework which is being developed, which aims to ensure that institutions have considered the risks associated with use of Web 2.0 technologies and services and have identified strategies for dealing with potential risks in order to achieve the goal of balancing the risks and benefits in order to maximise the dividends to be gained by use of Web 2.0.

**Keywords:** Social Web, Web 2.0, risks

#### **Introduction**

The Web 2.0 term has now been widely accepted as a description of a new pattern of ways in which the Web is being used. The Web environment has progressed from the publishing paradigm which characterised what is now sometimes referred to as Web 1.0, in which small numbers of content creators use tools ranging from desktop HTML authoring tools through to enterprise Content Management Systems and corresponding editorial and quality assurance processes to produce content for passive consumption by end users.

In a Web 2.0 environment large numbers of users are creating content using an ever-increasing variety of tools with such content being made available via a wide variety of commercial Web 2.0 services including photographic sharing services such as Flickr, video sharing services such as YouTube and social networking services such as MySpace and Facebook.

The characteristics of Web 2.0 were described by O'Reilly [1]. The key areas relevant to this paper include: (a) application areas including blogs and wikis, social sharing services and social networking services; (b) the ease of reuse of content elsewhere through syndication technologies such as RSS; (c) a culture of openness and sharing, which has been helped through the development of copyright licences such as Creative Commons; and (d) the concept of the 'network as the platform' by which services are hosted on externally-hosted services and accessible over the network, rather than a managed service within the organisation.

The Social Web is closely linked to Web 2.0. But whereas Web 2.0 includes various technical aspects (including technologies such as RSS and AJAX) the focus of the Social Web is very much focussed on the connections between people.

#### **Opportunities**

Why should cultural heritage organisations be interested in Web 2.0 and the Social Web? Answers to this question may include:

- The Social Web is popular.
- Social Web services can provide an opportunity to engage with new user communities and address challenges such as widening participation and social inclusion.
- Cultural heritage organisations, which are concerned with sharing and maximising access to cultural and scholarly resources, can exploit the Social Web to further this key mission.

- The Social Web can be cost-effective, allowing cultural heritage organisations to exploit a technical infrastructure that is already in place and is popular with many users.
- The economic downturn means funding for in-house development work is difficult to obtain.
- Popular Web 2.0 services can be easily used by end users as they can make use of services and interfaces they may already be familiar with.

These opportunities have been identified by many cultural heritage organisations which are already exploiting the Social Web's potential. Some examples of how museums, libraries and galleries are exploiting the Social Web are given below:

The National Library of Wales (NLW) has a remit to collect, preserve and give access to all kinds and forms of recorded knowledge, especially relating to Wales and the other Celtic countries, for the benefit of the public, including those engaged in research and learning. The use of Web 2.0 approaches for Library 2.0 delivery is ingrained in the NLW's 2008 strategy document *Shaping the Future* [2] which outlines the Library's desire to explore collaborative and diverse models using external resources. This will allow the NLW to leverage Web platforms which are heavily focused on user engagement in order to deliver future services.

The Brighton on the Pull Project provided an opportunity for Brighton Museum & Art Gallery to work with target audiences and new ways of researching their collections. The ethos of On the Pull was concerned with taking a step away from the traditional museum exhibition to encourage new visitors and target audiences. The project team explored use of social networking services as a marketing tool in order to get away from the associations with the word 'museum' as a way of breaking down barriers and the connotations the word was found to hold for the focus groups. Music and video clips hosted on YouTube are embedded in MySpace. In addition to MySpace, FaceBook was also used as a marketing tool to advertise events, promote competitions, display promotional images, images of objects from the collections and play music. [3].

The arrival of a pair of nesting peregrine falcons at Derby Cathedral provided Derby Museums with an opportunity to promote the town to a large audience. A Webcam provided live video footage of the nesting of three chicks and an accompanying blog and MySpace account, together with use of Flickr and YouTube for providing access to photographs and video footage resulted in "evidence emerging of visitors coming to Derby specifically because of its peregrines" [4].

## Risks and Barriers

### Identifying the Concerns

UKOLN, a national centre of expertise in digital information management based at the University of Bath has, over the past two years, delivered a series of workshops for the UK's cultural heritage sector. The workshops have provided an opportunity for practitioners in the sector to gain an understanding of the potential of Web 2.0 and to explore its potential. The workshops have also identified barriers to the effective deployment and use of the Social Web. A summary of the various concerns is given below.

**Sustainability Challenges:** There may be concerns over the lack of interoperability of third party services, with dangers that a service may be a 'walled garden', allowing data and content to be added to the service or created within the service but cannot be exported to another environment.

**Technical Challenges:** IT support staff may raise technical concerns related to reliance on third party organisations to deliver services for the organisation. These concerns might include performance and reliability issues, security, backups, etc.

**Interoperability Challenges:** Technical concerns raised may also cover the interoperability of third party services with other systems. This might include integration with existing in-house services and the export and migration of data to other services, including replacement services which might not be currently available.

**Support Issues:** Although many popular Social Web services can be used without formal training or support, use of the services in an institutional context may generate user queries.

**Individual Concerns:** Individuals within the organisation may be concerned with the deployment of Social Web services. Staff within the organisation may be reluctant to use technologies such as blogs and micro-blogging services such as Twitter because of an unfamiliarity with the technologies or the culture and expectations in these technologies or a desire to keep professional and social activities separate.

**Organisational Issues:** Proposals to make use of Social Web services by a cultural heritage organisation may not be universally welcomed by everyone within the institution. This may be regarded as undermining the organisation or a department in the organisation. Such concerns may not be openly articulated, but may lie behind concerns raised listed above.

## The Legal and Related Concerns

There are a number of legal risks involved in creating and using resources hosted on Social Web services. Briefly, they can be summed up as follows:

- Putting materials in that one should not deposit, because the copyright or other Intellectual Property Rights are held by third parties.
- Use of Registered Trade Marks or unregistered trade names without permission.
- Failure to identify someone as an author when they should be so named – this may well be an infringement of their Moral Right of paternity.
- Failure to respect the Moral Rights of authors e.g. derogatory treatment of their work.
- Data or advice that is inaccurate or misleading, and could lead to financial, physical or other damage to third parties if followed.
- Outputs that break the Data Protection Act or infringe personal privacy.
- Outputs that contain illegal materials e.g. materials that are pornographic, encourage terrorism, are defamatory, are in Contempt of Court, break race or sex discrimination laws, etc.
- Outputs that break the Disability Discrimination Act by being unreadable to those with impairments.

## A Risks and Opportunities Framework

### The Tensions

This paper has provided examples of use of the Social Web by cultural heritage organisations which have identified the benefits which the Social Web can provide in enhancing the range and quality of services to the organisations' user communities. However we have also listed a range of concerns which organisations considering making use of Social Web services will need to consider. We will now describe a risks and opportunities framework which has been developed in order to support cultural heritage organisations in making decisions on use of the Social Web.

### A Risks and Opportunities Framework For Addressing The Tensions

A risks and opportunities framework has been developed to support cultural heritage organisations in making effective use of the Social Web [6]. This paper introduces further developments to the framework including a summary of risk minimisation approaches and an inclusion of an evidence base.

Use of the framework involves documenting the following aspects of the proposed use of the Social Web:

**Intended use:** Rather than talking about Social Web services in an abstract context ("shall we have a Facebook page" for example) specific details of the intended use should be provided.

**Perceived benefits:** A summary of the perceived benefits which use of the Social Web service are expected to provide should be documented.

**Perceived risks:** A summary of the perceived risks which use of the Social Web service may entail should be documented.

**Missed opportunities:** A summary of the missed opportunities and benefits which a failure to make use of the Social Web service should be documented.

**Costs:** A summary of the costs and other resource implications of use of the service should be documented.

**Risk minimisation:** Once the risks have been identified and discussed approaches to risk minimisation should be documented.

**Evidence base:** Evidence which back up the assertions made in use of the framework.

When using this framework it should be recognised that there are likely to be biases, prejudices, vested interests and other subjective factors which will affect how the framework is used. Ideally such subjective factors will be openly acknowledged and taken into account, although it is recognised that this may be difficult to achieve.



### Application to In-House Developments and Existing Services

It should be noted that this framework need not only be applied to proposals to make use of the Social Web. In order to minimise the subjectivity of the approach it should also be applied to proposed in-house development work and commissioning IT developments. It can also be applied to existing services in order to identify the risks, limitations and constraints which the organisation is willing to tolerate and accept.

## Legal Risks

### A Risk Assessment Formula for Legal Infringements

The risks and opportunities framework recognises that although there will be risks when seeking to exploit the Social Web, it may be necessary to accept such risks in order to deliver services to the user community. A similar approach can be taken to addressing the risks associated with possible copyright infringement.

The example below relates to copyright infringement, but the same formula applies to all the legal risks identified above. The risk can be calculated as follows:

$$R = A \times B \times C \times D$$

where R is the financial risk; A is the chances that what has been done is infringement; B is the chances that the copyright owner becomes aware of such infringement; C is the chances that having become aware, the owner sues and D is the financial cost (damages, legal fees, opportunity costs in defending the action, plus loss of reputation) for such a legal action. Each one of these other than D ranges from 0 (no risk at all) to 1 (100% certain). D is potentially a high number. It is not easy to calculate the cost of loss of reputation.

Factors to bear in mind:

- If the work is to be used in a commercial context (i.e. to generate financial gain) then a rights owner who later becomes aware of the use of their work may be more likely to pursue an action for infringement of copyright than if the work is being purely used for educational purposes.
- The nature of the content used, for example, the rights in high value content, such as commercially produced films, text, images, music and software, are more likely to be actively enforced by their owners.
- Particularly sensitive subject areas are music, geographic data, literary works by eminent authors, and artistic works including photographs and drawings.
- Is there any track record of the contributor ignoring legal niceties in the past?
- Is there any track record of a particular third party having complained before?

Depending on these factors, the risks will vary. However, a Web 2.0 provider that ignored warning signals (e.g. a contributor who ignored legal niceties in the past is allowed to continue to add more materials without checks being made) is likely to receive an unsympathetic hearing from the Courts. Similarly, a service provider who has failed to educate contributors regarding legal issues will also not be viewed sympathetically by a Court.

Ultimately, it is important that the service provider is proportionate about possible risks whilst at the same time prepares suitable mitigating strategies in the eventuality of a complaint. An apology and promise of swift action to rectify is often sufficient.

### Reducing the Legal Risks

A cultural heritage organisation making use of the Social Web should ensure that it has clear and robust notice and take down policies and procedures with a clear address given for complaints. The notice and take down policy adopted by JORUM would be a valuable starting point [7]. Clear instructions should be given as to where and to whom notification of allegedly illegal content should be sent, along with details of the complainer, the complainer's interest in the matter and where the complainer can be contacted. Processes should be put in place to act expeditiously on such a notification.

### Conclusions

We have seen how cultural heritage organisations are successfully exploiting the Social Web in order to deliver quality services to their user communities. There is, however, an awareness that there are a variety of risks associated with use of the Social Web. The use of a risks and opportunities framework and a risk assessment formula have been described which aim to support the discussions and policy-making processes organisations need to take when formulating policies on exploiting the Social Web.

### References

- [1] What Is Web 2.0? O'Reilly, T. Retrieved September 20th 2009: <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
- [2] Shaping the future: The Library's strategy 2008-2009 to 2010-2011, National Library of Wales. Retrieved September 20th 2009: <http://www.llgc.org.uk/fileadmin/documents/pdf/Strategy2008-2011.pdf>
- [3] On The Pull, English, C. UKOLN Cultural Heritage blog, 2 April 2009. Retrieved September 20th 2009: <http://blogs.ukoln.ac.uk/cultural-heritage/2009/04/02/the-on-the-pull-project/>



*Empowering users: an active role  
for user communities*

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

- [4] When Peregrines Come To Town, Moyes, N, UKOLN Cultural Heritage blog, 18 May 2009. Retrieved September 20th 2009: <http://blogs.ukoln.ac.uk/cultural-heritage/2009/05/18/when-peregrines-come-to-town/>
- [5] Time To Stop Doing and Start Thinking: A Framework For Exploiting Web 2.0 Service, Kelly, B., Museums and the Web 2009 Conference. Retrieved September 20th 2009: <http://www.ukoln.ac.uk/web-focus/papers/mw-2009/>
- [6] JORUM procedures to Deal with Queries, Alerts and Complaints, JORUM. Retrieved September 20th 2009: [http://www.jorum.ac.uk/docs/pdf/takedown\\_procedure.pdf](http://www.jorum.ac.uk/docs/pdf/takedown_procedure.pdf)

**Smiljana ANTONIJEVIC, Laura J. GURAK**

**Trust in online interaction: an analysis of the socio-psychological features of online communities and user engagement**

**Abstract**

This paper is part of the authors' joint project on trust in online interaction, and it contributes to the enhancement of collaborative knowledge environments by advancing our understanding of key socio-psychological features of online communities and user engagement. We first review the history of online communities and user engagement, focusing on the socio-psychological features of trust in online interaction. Next, we discuss the issue of trust with regard to user-generated content and cultural heritage, highlighting the issues of trusting beliefs, trusting intentions, and trust transfer. Finally, we argue that a diachronic understanding of online practices holds the capacity to explain much of what we see online today, and we propose that the power of this digital legacy should particularly be valued and employed in the institutions and contexts promoting and fostering cultural heritage in both traditional and contemporary forms.

**Keywords:** trust; user engagement; Internet; cultural heritage.

**Introduction**

The role of users and user communities in the production of online content has become a hot topic in recent academic and public debate, featuring themes such as social networking, user labor, user co-creation, user-generated content (UGC), social annotation, folksonomies, and the like. As van Dijck [1] points out, "with the emergence of Web 2.0 applications, most prominently UGC platforms, the qualification of 'user' gradually enters the common parlance" (p. 41). In this paper, we show that user engagement and user generated online content have a much longer tradition than the current debate implies, and we argue that a diachronic understanding of online practices holds the capacity to explain much of what we see online today. The idea of engaged users and engaged user communities was in a sense "hard wired" into the very earliest pre-Internet systems, thus providing the basis for what we have today and more importantly, for what we have come to expect and even demand in online systems.

Contemporary notions such as "social Web," "participatory Web," "prouusers" and the like obscure the fact that the earliest computer networks had been built as networks of people, not wires, and have always been social and participatory, even before they became the Web. What Web 2.0 has brought is rather the matter of performance than of essence; it has brought easy to use content-creating applications - such as blogs, wikis, social networking sites, and file sharing platforms - rooted in broadband access, affordable hardware and software solutions, and with the Internet perceived and used as a "new normal" in contemporary way of life [2]. Put differently, Web 2.0 has made visible what has always been there - engaged users and engaged online communities.

**Trust in Online Communities and User Engagement**

The concepts of online communities and user engagement have always been central to the analyses of online interaction, even in the pre-Web days. For example, in their prescient 1968 paper, Licklider and Taylor [3] observed that the very earliest networked computers of their time enabled for formation of communities, noting, in a vision that has become the underlying theme of all research in online community to date, that "[online communities of the future] will be communities not of common location, but of common interest" (pp. 37-38; emphasis in the original). As networked computing moved from the laboratories into universities and the corporate world, researchers such as Hiltz and Turoff [4], Kiesler, Siegel and McGuire [5], and Turkle [6] continued to fine-tune our understandings of the social-psychological features of computer-mediated communication (CMC). And as networked computing moved from internal sites (such as corporations and universities) into the general populace, the study of online communities shifted to a wider range of cases and academic disciplines, providing both optimistic [e.g., 7;8] and pessimistic [e.g., 9;10] interpretations of online sociality.

Gradually, new dynamics involving trust and credibility of online interaction began to emerge, highlighting the issue of user engagement as an inseparable element of online sociality. For instance, in her comparative study of two online protests, Gurak [11] demonstrated that already in the pre-Web days a large number of users could quickly assemble around an idea of common interest, and, by sharing information and providing user generated content, create an efficient and successful online social action. Baym's [12] work on online soap opera fan clubs and Hine's [13] methodologically-oriented study of online activities surrounding the Louise Woodward

case offered additional research-based evidence for the power of user-provided content and online communities, while Bakardjieva [14] showed that users transcended the sphere of personal experience by engaging in collective online practice; she asked why the users did what they did and what it meant to them. The same questions - why do users engage in providing online content and what does it mean to them - have come to the fore in the recent analyses of UGC, such as Baym and Burnett's [15] study of amateur experts' provided online content. This last point is one to which we will return in the next segment, when we discuss the main features of user-generated content, particularly in regard to cultural heritage.

In summary, contemporary online forums facilitate and promote features of online discourse that have been part of online interaction from the outset. The gathering of like-minded people around communities of common interest is key to understanding user engagement in the 21st century. Today's user wants quick, accurate, customizable, smart systems, and they want systems they can trust.

Indeed, all forms of human communication, but particularly online communication, depends heavily on trust. As Seligman [16] has stated, "[t]he existence of trust is an essential component of all enduring social relationships" (p. 13). Drawing upon the work of Luhmann, Seligman notes the relationship between trust and confidence, the latter based on whether "one can rely or place confidence in the other's words or commitments or acts" (p. 21; emphasis in the original). Trust, in these terms, then, "involves a vulnerability occasioned by some form of ignorance or basic uncertainty as to the other's motives," which Seligman notes as a particularly interesting concept in the Internet age, because of the "fundamental opaqueness toward the will of another" (p. 21). In other words, while trust and confidence in others is a foundational concept in all forms of human communication, it is a particularly interesting one in the digital age. We could not have functioning online communities without trust; indeed in broad terms, trust is "an important dimension of civic culture" [17, p. 14]. And we could not have any level of user engagement without both trust in the system and confidence in the motives of the system itself as well as the motives of other online participants.

Numerous studies and commentaries have been written on the issue of trust in digital environments in relation to issues such as usability [18], e-commerce [19], interface design [20], credibility [21], and other areas. Bailey, Gurak, & Konstan [22] have noted that "[t]rust plays a critical role when a user assesses the believability of online information content or when selecting an exchange site to purchase a product from. Users will not believe or participate in a transaction with those whom they do not trust" (pg. 311). Bailey, Gurak, & Konstan thus define trust as "the perception of the degree to which an exchange partner will fulfill their transactional obligations in situations characterized by risk or uncertainty" (p. 313), and they posit seven dimensions of trust in digital settings: attraction, dynamism, expertness, faith, intentions, localness, and reliability (p. 315).

More applicable to this paper, however, is recent work on trust in digital repositories [23]. Summarizing the literature regarding technical considerations when building a trusted digital repository, Prieto notes issues such as persistent access, content migration, resource discovery, data collection/quality, and so on (p. 596). He also notes the importance of users feeling that the content within a digital repository is itself trustworthy (p. 596). Yet he then makes this point: the role of the digital repository's stakeholders (whether they are referred to collectively as a community or individually as depositors or users) is key to establishing trust. Put another way, he says that "the repository can be trusted because it has been deemed an appropriate place into which content can be contributed or from which content can be retrieved for purposes of research, study, enrichment, or personal enjoyment" (p. 596). Prieto then goes on to describe the kinds of incentives that might need to be put in place in order to move the academic world from trust that is rooted in print to trust in digital repositories (p. 597).

To this end, we feel that user-generated content, a key feature of the Web 2.0 age, would build on the features of online community as well as digital trust, and could be of key importance to online systems designed for cultural heritage in the 21st century.

### **Trust in User-Generated Content and Cultural Heritage**

User-generated content refers to online content produced by end-users. Online material is considered user-generated when/if it is publicly available, created outside of professional settings, and includes a user's creative effort (i.e., the user did not simply copy and paste the content, but rather has added his or her own creative value to it); [24]. The practice of creating and providing user-generated content is usually considered to stem from the following motives: obtaining public acknowledgement; earning peer recognition; building reputation in a community; expressing oneself; developing skills that can become a profession; having fun; sharing knowledge/contributing to a common idea. In the case of UGC related to cultural heritage, sharing knowledge and contributing to a common idea is often seen as key stimuli, although other motives also play the role in this type of user engagement. For instance, both professional and amateur subject specialists are often prompted to contribute specialized local and/or minority content in different languages, and/or to engage in cultural heritage tribute online activities [25].



User-generated content that contributes to the cultural heritage sector is deemed to have both positive and negative aspects. For instance, UGC is often considered to complement and enhance institutionally provided content by offering novel information on specific—often local—cultural phenomena, as well as by offering novel ways of presenting and/or interpreting those phenomena. UGC is also regarded as a means of transforming static content authority into dynamic, multisided knowledge platform, which has the capacity to engage the public as an acknowledged knowledge-provider [26, p. 20]. Finally, users' active online engagement with cultural phenomena is considered to have broader implications for the perception of cultural heritage, by fostering understanding of culture as an ongoing process, not as closed, historic experience completed and rooted in the past [25]. However, UGC is also considered to introduce unverified and/or difficult to verify popular knowledge (sometimes called "crowdsourcing") into the cultural heritage domain. Participatory culture prompted by Web 2.0 applications especially raises the question of trusted and reliable content with regard to sensitive multicultural issues and/or issues stirring intercultural debates. Similarly, UGC prompts the question of intellectual property (IP), either in the sense of granting users IP rights over their online creations, or in the sense of insuring that UGC acknowledges authorship of the original source(s) when needed.

Both positive and negative aspects of user-generated cultural heritage content bring us back to the subject of trust, by invoking the issues of trusting beliefs, trusting intentions, and trust transfer [27, p. 21]. While trusting beliefs refers to a person's perception of a certain actor and/or source as trustworthy, trusting intentions encompass a person's willingness to make him or herself vulnerable in accordance with such a perception. With UGC, trusting intentions encompass users' willingness to make themselves vulnerable to cultural information provided by other users. Vulnerability in this case implies exposure to potentially incorrect, incomplete, misleading, biased, or in some other way corrupted content. On such occasions, the issue of trust-transfer often comes to the fore, requiring that the burden of establishing trust be transferred from the user to an external proof source. In case of UGC, the proof source is usually institutional and/or contextual. For instance, if a piece of UGC is provided within the official website of an acknowledged cultural institution, the cognitive mechanism of trust-transfer associates such a piece with the given organization, reassigning the proof burden from the user to the institution. Similarly, if a particular user-generated post is provided within an acknowledged online resource, such as Wikipedia, and/or within an online community known to the post recipient, the trust-transfer process shifts the proof burden to the given online context. User-generated content thus reflects a complex interplay among users, institutions, and online contexts, or, as Ridge [28] points out, it reflects the issues of sharing authorship and authority with regard to cultural heritage. To address those issues, both the users and the cultural heritage institutions should be empowered to efficiently employ online systems designed for cultural heritage in the digital age. User empowerment in this sense implies providing customizable, smart, and easy to use systems, which enable users to employ their knowledge, creativity, interest and other positive stimuli by creating, sharing, and discussing cultural content. Similarly, institution empowerment implies enabling cultural heritage organizations to employ the power of user engagement, while at the same time avoiding previously mentioned downsides of UGC through sets of relevant procedures [28]. In other words, developing trust both within and among users and institutions is key to empowering these major actors in contemporary processes of cultural production and preservation.

## Conclusion

Research on digital technology and culture has always made clear that the power and potential of the Internet lies in the unique dynamics and possibilities for online communities and user engagement. Nowhere is this concept of user and communal power more clearly visible than with today's social networking applications and Web 2.0 forms. Contemporary users expect to play an active role in shaping their online experience and want systems that are smart, customizable, and cross traditional boundaries. Consequently, cultural heritage institutions should transcend the "search and access" approach, and in contract should serve as collaborative knowledge environments that promote users actively engaged in creation, modification, and distribution of information objects. To achieve this, cultural heritage institutions have the potential to employ valuable knowledge on user practices harvested in the field of Internet Studies over the past thirty years. While technologies and forms of online interaction have been and will keep changing in a blink of the eye, the main socio-psychological features and dynamics of user engagement have been and most likely will remain rather steady across the platforms and contexts of use. The power of this digital legacy should particularly be valued and employed in the institutions and contexts promoting and fostering cultural heritage in both of its traditional and contemporary forms.

## References

- [1] Dijck, van J. (2009). "Users like you? Theorizing agency in user-generated content." *Media, Culture & Society*, Vol. 31 (1), pp. 41-58.

- [2] Pew Internet & American Life Project. (2005). Internet: The mainstreaming of online life. Retrieved June 27, 2007, from [www.pewinternet.org/pdfs/Internet\\_Status\\_2005.pdf](http://www.pewinternet.org/pdfs/Internet_Status_2005.pdf).
- [3] Licklider, J. C. R. and Taylor, R. (1968). "The Computer as a Communication Device." *Science and Technology*, April: pp. 21-41.
- [4] Hiltz, S. R., & Turoff, M. (1978). *The Network Nation*. Reading, MA: Addison-Wesley.
- [5] Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social and psychological aspects of computermediated communication. *American Psychologist*, 39(10), pp. 1123–1134.
- [6] Turkle, S. (1984). *The second self: Computers and the human spirit*. New York: Simon & Schuster.
- [7] Rheingold, H. (1993). *The virtual community: Homesteading on the electronic frontier*. New York: Harper Collins.
- [8] Reid, E.M. (1991). *Electropolis: Communication and Community On Internet Relay Chat*. Retrieved October 15, 2009, from <http://irchelp.org/irchelp/misc/electropolis.html>
- [9] Doheny-Farina, S. (1996). *The Wired Neighborhood*. New Haven, CT: Yale University Press.
- [10] Fenerback, J. and Thompson, B. (1995). *Virtual Communities: Abort, Retry, Failure?.* Retrieved October 15, 2009, from <http://www.well.com/user/hlr/texts/Vccivil.html>.
- [11] Gurak, L. J. (1997). *Persuasion and privacy in cyberspace: The online protests over Lotus market place and the clipper chip*. New Haven, CT: Yale University Press.
- [12] Baym, N. (1999). *Tune in, log on: Soaps, fandom, and online community*. Thousand Oaks, CA: Sage.
- [13] Hine, C. (2000). *Virtual Ethnography*. London: Sage.
- [14] Bakardjieva, M. (2003). "Virtual togetherness." *Media, Culture, & Society*, Vol. 25 (3), pp. 291-313
- [15] Baym, N.K., and Burnett, R. (2009). "Amateur Experts: International fan labour in Swedish independent music." *International Journal of Cultural Studies*, Vol. 12(5), pp. 433-449.
- [16] Seligman, A. B. (1997). *The problem of trust*. Princeton, NJ: Princeton University Press.
- [17] Sztompka, P. (1999). *Trust: a sociological theory*. Cambridge, UK: Cambridge University Press.
- [18] Bedi, P. and Banati, H. (2006). "Assessing user trust to improve web usability." *Journal of Computer Science*, Vol. 2 No. 3, pp. 283-7.
- [19] Klang, M. (2001), "Who do you trust? Beyond encryption, secure e-business," *Decision Support Systems*, Vol. 31 No. 3, pp. 293-301.
- [20] Pu, P. and Chen, L. (2007), "Trust-inspiring explanation interfaces for recommender systems," *Knowledge-Based Systems*, Vol. 20 No. 6, pp. 542-56
- [21] Fogg, BJ & Tseng, H. (1999). "The elements of computer credibility". *Proceedings of ACM Conference on Human Factors and Computing Systems*, pp. 80 – 87.
- [22] Bailey, B.P., Gurak, L. J., and Konstan, J. (1998). Trust in cyberspace. in *Human Factors and Web Development*. ed. Ratner, J. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 311-321
- [23] Prieto, A. G. 2009 From conceptual to perceptual reality: trust n digital repositories. *Library Review* 58 (8), pp. 593-606.
- [24] OECD. (2007). *Participative Web and User-Created Content: Web 2.0, Wikies, and Social Networking*. Retrieved October 15, 2009, from <http://213.253.134.43/oecd/pdfs/browseit/9307031E.PDF>.
- [25] Harrison, R. (2009). "Excavating Second Life: Cyber-Archaeologies, Heritage and Virtual Communities." *Journal of Material Culture*, Vol. 14 (1), PP. 75-106.
- [26] Coppola, P., Lomuscio R., Mizzaro, S., Nazzi, E., and Vessena, L. (2008). "Mobile Social Software for Cultural Heritage: A Reference Model." *2nd Workshop on Social Aspects of the Web (SAW 2008), BIS 2008 Workshop Proceedings*, pp. 69-80.
- [27] Stewart, K.J. (2003). "Trust Transfer on the World Wide Web." *Organization Science*, Vol, 14 (1), pp. 5-17.
- [28] Ridge, M. (2007). *Sharing authorship and authority: user generated content and the cultural heritage sector*. 2007 Web Adept - UK Museums on the Web, Museums Computer Group conference. Retrieved October 15, 2009, from <http://www.miaridge.com/projects/usergeneratedcontentinculturalheritagesector.html>

## **Max KAISER, Jeanna NIKOLOV-RAMÍREZ GAVIRIA, Veronika PRÄNDL-ZIKA**

### **EuropeanaConnect - Enhancing user access to European digital heritage**

#### **Abstract**

This paper provides an outline of the main goals of the project EuropeanaConnect, one of the two core technical projects in the Europeana project cluster which will develop the current prototype of Europeana into a fully operational service until 2011. Europeana is Europe's digital library, archive and museum portal, launched in November 2008.

**Keywords:** Europeana, semantic web, multilingual access, mobile access, value-added services, licensing framework, audio content

#### **Introduction**

The prototype of Europeana (<http://www.europeana.eu>), Europe's digital library, archive and museum portal, was launched in November 2008. The European Commission is funding the development of Europeana as part of its i2010 Digital Libraries initiative to create an integrated access channel to Europe's distributed digitised cultural heritage resources.

The launch of the Europeana prototype created such a large-scale public interest that the site had to be taken down for a few weeks for a substantial hardware and load balancing reconfiguration. Back online in a test mode since December 2008 and with full functionality since April 2009, the Europeana prototype currently gives access to about five million digitised objects of cultural value from over 1.000 European libraries, archives, museums and audio-visual collections.

Within the next two years Europeana will be developed into a fully operational service. By mid 2010 Europeana is expected to provide access to 10 million items; the target for 2012 is 25 million items. The European Commission is funding a number of projects in the eContentplus and ICT PSP programmes. Within the next years, these projects will develop the technology and services for Europeana and digitise and aggregate content for the portal.

Two core projects will be responsible for the implementation of the technical infrastructure of Europeana: The project Europeana v1.0 (<http://version1.europeana.eu/>) is led by the European Digital Library Foundation. This 30-month project which started in February 2009 will solve key operation issues related to the implementation of Europeana like building the back end services needed to manage the delivery and access of Europeana content and managing the channels that will enable other environments to use the content made interoperable by Europeana via web services and APIs. This project will also develop and manage partnerships to ensure a rich content flow from national and domain aggregators. The second core project is EuropeanaConnect (<http://www.europeanaconnect.eu>) and will be described in the remainder of this paper.

#### **EuropeanaConnect**

EuropeanaConnect is closely associated with the project Europeana v. 1.0 and is a 30-month Best Practice Network co-funded by the European Commission in the eContentplus programme. Co-ordinated by the Austrian National Library, EuropeanaConnect will deliver core components and value-added services for Europeana. It will also act as a content aggregator for audio content and will significantly enlarge the music content available in Europeana. The project runs from May 2009 to October 2011 and brings together 30 partners from 14 European countries, including universities, research institutes, libraries, audio archives and a partner from the publishing industry.

#### **Europeana Semantic Layer**

One of the main goals of EuropeanaConnect is to build a layer of semantic data which will be the basis for all semantic processing in Europeana. Semantic processing would, for example, enable a relationship between "Mona Lisa", "Lisa del Giocondo", "La Joconde" and "Leonardo da Vinci" to be recognised during the search process. A user entering just one of these terms would see also results for the linked concepts, e.g. would automatically find other paintings created by Leonardo da Vinci. EuropeanaConnect will provide the technologies to semantically enrich huge amounts of digital content in Europeana. This will enable semantically-based content discovery and make Europeana content more accessible, reusable and exploitable. EuropeanaConnect will create a network of semantic resources starting from controlled vocabularies and other similar resources which will be provided by content owners. These will be made ready for semantic processing



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

mostly by using SKOS (Simple Knowledge Organization System, a W3C standard) which in turn will be used as the primary level of user interaction with Europeana. Unlike in traditional library catalogue-driven models of user interaction, users of Europeana will be able to explore the data space by using the semantic nodes (i. e. concepts organised and available as web resources) as primary elements for searching and (faceted) browsing.

### **Multilingual Access**

Multilingual and multicultural aspects are at the heart of making Europeana's digital contents effective and exploitable for users across all European countries. Metadata and digital content should be searchable and presentable independent of the language of the searcher or the object descriptions. In other words, users should be able to enter search terms in their own or preferred language, which are translated on the fly during the search process, and have hits returned in a range of selected languages. In order to enhance the multilingual access capabilities of Europeana, EuropeanaConnect will develop a multilingual infrastructure and a set of translation tools that will process queries and object data and produce a suitable multilingual representation for the user:

- a unique repository of language and translation resources for multilingual processing of objects within Europeana, which will possibly be one of the most varied of its kind in Europe;
- a suite of tools for multilingual mapping of controlled vocabularies in at least five European languages and a strategy for incorporating more languages;
- a suite of integrated translation modules for querying and browsing processes which will plug into the general search and retrieval infrastructure of Europeana.

Not every European language is equally well developed in terms of multilingual resources for translation. EuropeanaConnect will start with a core set of languages (English, French, German, Italian, Spanish) for which expanded multilingual translation capabilities will be implemented by the end of the project. During the course of the project, the core set will be expanded, based on the development experiences, with a set of additional languages integrated with less linguistic treatment based on available resources and content.

### **Tools and methodologies for user-driven development**

EuropeanaConnect will provide a better understanding of what current and future web-users really need and expect of the multi-lingual and multi-channel service which Europeana must become. The project will provide the methodology to monitor and evaluate the user interactions with Europeana. Deep Log Analysis will allow for the evaluation of significant information-seeking behaviour of Europeana users. This method will be used by Europeana as a routine for understanding how the Europeana portal is used and what its users really expect from it. The usage-monitoring methodologies and the results of Deep Log Analysis will inform the implementation iterations of the Europeana portal (Europeana v1.0) and will thus ensure that the development of Europeana is truly user-driven.

Europeana sees all groups of people as potential users. It will not be sufficient to have information on the actual user behaviour; we also need to move from the concept of the user as "anonymous" to an understanding of who is using what. To support this and to generally ensure user-oriented service development and marketing of Europeana, EuropeanaConnect will define "personas", i.e. categories with profiles. We need to understand the link between the results of the deep log analysis and these personas and we need to understand how Europeana can add value to a selected number of these personas. This work will be accompanied by actual user studies either through questionnaires or via different forms of usability tests. From a preliminary screening we aim for a rough number of 500 people to be involved throughout the lifetime of the project.

### **New Access Channels to Europeana**

#### *Spatio-temporal user interface*

EuropeanaConnect will develop a spatio-temporal access service that will support access to Europeana via a number of space- and time-related selection mechanisms as part of the query- and result-presentation process. It will implement a user-friendly access mechanism, which exploits the Semantic Layer of Europeana. The objective is to create an interface that makes use of both time-based and geographical metadata of the Semantic Layer, providing a new visual access to the digital collections in Europeana. The project will build an interactive and generic map combined with a timeline, acting as a "histogram" for tagged events, places and characters in Europeana. The interactive map will show the visual representation of a search-result set in Europeana, e.g. concerning the search term "migration". The histogram on the timeline below the map will indicate how many items are registered in Europeana within a specific timeframe.

#### *Mobile Access Channel*

EuropeanaConnect will address the user demand to access information through mobile devices. The project will develop both a middleware layer, which will allow different mobile devices to access Europeana services

created and a rich mobile client for real end user evaluation. The middleware will allow for access to Europeana from different mobile devices, channels and applications. It will be located between mobile devices on the user side and the Europeana semantic layer and database on the server side. It is intended to overcome the heterogeneity and limitations of mobile devices by selecting and adapting the content for queries to make them displayable on the devices. Therefore, queries from mobile devices are processed by the middleware, which uses the Europeana semantic layer to retrieve the desired data. The query results are collections of media which have to be adapted for mobile usage and a presentation of the results based on the metadata will be prepared. Finally, this presentation will be transferred to the mobile device. Advanced queries and interfaces require more powerful mobile devices, such as Smartphones. Therefore EuropeanaConnect will also implement a rich mobile client application which runs on Windows Mobile-based devices.

### **Value-Added Services for Europeana**

EuropeanaConnect will integrate three value-added services with the Europeana portal that will significantly enhance the functionality and the usability of Europeana:

#### *Multimedia Annotation Service*

This service will allow all Europeana users to make their own contributions to Europeana in the form of tagging, comments, discussions, and linking. Annotations are an important means by which users can contribute knowledge and exchange information about content within a group of experts. Cultural institutions and museums that have already exposed their content on the Web are discovering that annotations can change the traditional role of visitors from passive consumers to actively contributing and collaborating users. Annotations allow users to exploit and add value to existing content, while enhancing accessibility through the user-provided searchable metadata and links to other media. Annotations further enable communication and collaboration with other users interested in the same content, which will support the establishment of social networks built around Europeana. The Annotation Service will allow the tagging, linking, and annotating of Europeana media resources of various formats (HTML, image, audio, and video).

#### *GIS Service Suite*

The GIS Service will allow users to query and display Europeana content based on spatial information, and to discover new relationships between content items, based on location. As only very few objects in Europeana already contain explicit geographical metadata, this service will also provide metadata enrichment for the spatio-temporal visual user interface described above. EuropeanaConnect will integrate a Geoparser and Gazetteer developed within the context of the DIGMAP and TELplus projects with the Europeana portal. By processing Europeana metadata with the Geoparser it will be possible to detect geographic terms and time periods. This will allow to improve the Europeana metadata. Pre-processed data will be stored in the semantic layer and will later be used to populate the spatio-temporal user interface. Finally, we can apply the same processing in real time for the users' queries. When the Geoparser finds a relevant geographic or time reference (for example, "churches in the Scotland", "the battle of Lepanto", etc.) in a query, it can trigger specialised services that can take that in consideration.

#### *eBooks-On Demand*

Although Europeana provides access to some millions of digitised items from European cultural heritage institutions, these figures represent only a very small fraction of the analogue collections available in these institutions. In the case of books we know that the portion of already digitised material is around one percent of all books, in the case of archival material it will be even lower. This situation has been the starting point for the eBooks on Demand (EOD) Service that will be integrated into Europeana: The EOD Network (<http://books2ebooks.eu/>) is a growing consortium of 18 libraries from 10 European countries. They all provide the EOD service which allows users to order on-demand digitisation of public-domain books of the participating libraries and to receive the requested items electronically in the PDF format. EuropeanaConnect will adapt the EOD service allowing for an automated and generic transfer of the eBooks created within EOD to Europeana according to the specification of the Europeana Semantic Elements.

### **Infrastructure components for Europeana**

EuropeanaConnect will implement and deploy key infrastructure components that will enable Europeana to manage the harvesting of thousands of digital heritage content sources across Europe, to promote metadata and content interoperability across content providers as well as interoperability with independent value-added services, and to provide persistent and uniform identification of digital resources:

- The Europeana OAI-PMH Management Infrastructure will be a prerequisite for enabling Europeana to meet the demands of the expected leap in data harvesting scale, managing the harvesting of huge volumes of content from thousands of content providers in Europe.



- The Europeana Metadata Registry will be a precondition for metadata interoperability in Europeana, as it will manage the range of metadata schemata and terms used by the various European cultural heritage institutions
- The Europeana Service Registry will enable Europeana to integrate practically any external service offering an HTTP interface into the portal. It will allow users to search and select value-added services and will provide the functionalities for the Europeana portal to invoke appropriate services depending on the context.
- The Europeana Resolution Discovery Service for Persistent Identifiers will allow for uniform and persistent identification of resources between Europeana and service and content providers. It will provide a crucial component for the portal's interoperability that supports the integration of a critical mass of European digital heritage content for the end user.

### **Europeana Licensing Framework**

Even though content of Europeana is likely to be in the public domain or with the rights cleared, it will be essential to provide the user with an accurate description of the rights attached to the content. Appropriate licensing policies for objects and metadata as well as suitable implementation methods are thus needed in order to create trust for all parties. EuropeanaConnect will create the Europeana Licensing Framework and will build the necessary tools for Europeana to establish and declare the rights status associated with Europeana content. In close collaboration with the Europeana Office and the Europeana v1.0 project, EuropeanaConnect will establish a core set of interoperable licenses that cover the rights information for objects in Europeana. The proposed Europeana Licensing Framework will be validated through a broad stakeholder consultation process.

### **Aggregation of Audio Content**

Apart from building critical components for Europeana, EuropeanaConnect will also act as content aggregator: One of the shortcomings of the current Europeana prototype is its lack of a critical mass of non-textual content. EuropeanaConnect will add the music dimension to Europeana. As well as harvesting massive amounts of audio, the project will create, implement and evaluate access to right-free or wholly-owned music for all users of Europeana. EuropeanaConnect will integrate the framework created by the DISMARC eContentplus project (<http://www.dismarc.org/>) in Europeana and extend the technical, editorial and legal solutions of this platform. The project will provide the necessary infrastructure for harvesting, analysing and storing of audio metadata as well as music audio provision to Europeana. There will be a continuous approach to and involvement of new audio archives for the duration of the entire project, using the established infrastructure and harvesting process. It is anticipated that over the period of the project approximately 150 audio archives will participate as content providers, leading to an estimated 200.000 music tracks to be provided to Europeana by month 30.

### **Conclusion**

The launch of Europeana has yielded a high public interest. Within the next two years the current prototype has to be developed into a fully functional and scalable service that meets the high expectations of the user community. The project EuropeanaConnect which has been outlined in this paper will contribute to this goal on several levels. It will deliver key technical components which are essential for the realisation of Europeana as an interoperable and multilingual service. It will establish methodologies for analysis of user-interaction that will inform the implementation the Europeana portal is driven by user needs. Moreover it will deliver the Licensing Framework for Europeana and aggregate a critical mass of audio content.

### **Acknowledgement**

EuropeanaConnect is co-funded by the European Commission in the eContentplus programme.



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## **Fred STIELOW**

### **Perspectives from an online university community - Community Building in the Web Revolution**

#### **Abstract**

This essay provides digital Cultural Heritage actors with perspectives about Online education—especially, the for-profit universities. Discussions cast these newest players on the educational scene into longer-range historical context—from the birth of the University, appearance of the printing press, and rise of the Mass Press to the ongoing Web revolution. The treatment provides a technical focus on their nature and operational concerns as uniquely digital offspring of the Web in contrast to similar operations in Cultural Heritage. Capitalistic and entrepreneurial viewpoints are also featured in examples from the American Public University System. As depicted, the APUS model calls for re-engineering library, bookstore, and press operations toward direct support of cost-effective, electronic classroom materials. The results suggest proactively adding similar course-specific engineering to Cultural Heritage’s portfolio. The vision is a cooperative one of advanced Web-based applications, multi-institution combinations, and partnerships with the Online educators—all of which could contribute to the emergence of powerful Knowledge Communities on the frontier of learning.

**Keywords:** Online For-Profit Education, Electronic Course Materials, Knowledge Communities

Aldus [Manutius] is building up a library which has no other limits than the world itself (Erasmus, 1508)

#### **Introduction**

The presentation comes with cautions, but also a professed personal desire to explore future options. Online education and burgeoning for-profit universities are prime representatives of the new Information economy—but largely outside the orbit of digital Cultural Heritage. We will need some perspective. For-profits, for example, are uniquely of the Web. Internet applications may be similar, but you should understand that the viewpoints can differ fundamentally from Cultural Heritage. Realize too that these are capitalistic and highly entrepreneurial enterprises. Their cost/benefit analysis may be somewhat tempered for the sake of educational tradition, but assumptions and values that often subconsciously underlie Cultural Heritage may be challenged. Hence, these discussions will add a joint layer of historical background as necessary prelude to virtual community building.

#### **A Fully Virtual Viewpoint**

For-profit online universities are born of the Web. Such origins presuppose technological engagement - but are exclusively in the virtual. They do not include Cultural Heritage’s primary allegiance and steward over physical documents and material culture artifacts. As indicated in the following brief potpourri, even overlapping Web concerns can thus differ markedly. Note: I approach the technical side from early training in Systems Analysis and running data processing installations before encountering work in Special Collections - as well as a couple of more recent, yet now out-of-date books on the Web. My operational bias is KIS (keep-it-simple) planning and letting the machines do the work whenever possible.

*Deep Web v. Open Web:* The Information Highway divides between free information on the Open Web and controlled, often pay-for access through the larger Deep Web. Given the financial implications and personal commitment to the Open Access Movement, the former is much preferred for the Onlines. Our programmatic focus is largely validating for “trusted” resources. Recognized Cultural Heritage sites make this easier and are much appreciated. Some in the Cultural Heritage side may be more interested in Deep Web operations. Outsourcing access is a desired goal, but a good deal of time is spent in pricing negotiations and contracts in what is a chaotic marketplace. In addition, we become involved in the details of Digital Rights Management (DRM) and Authentication routines - e.g., password controls, IP recognition, and proxy servers. One also expects heavy reliance on measurement and metrics, which come at unprecedented levels for the Cultural Heritage community.

*Digital Preservation:* Although I have taught the subject and have ready recommendations, preservation does not enter my current virtual consciousness. The absence of artifactual values, limitless exact reproductions, and inability to identify the original makes the term a misnomer in digital realms. In a key distinction, these institutions are not stewards. They generally do not own print or material culture treasures. Rather materials are



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

merely leased or electronically borrowed for teaching and research purposes. Our management issues are ensuring access and basic readability on normal computer screens - scanning levels and DPI standards are not relevant for our type of studies.

*Finding Aids v. Full Text:* This is not really a question. Modern student clients could care less about Finding Aids and catalog entries. Students are simply spoiled by Google. They increasingly expect direct access to anything ever created. With the effective absence of the Series concept in favor of storage locations, even often touted EAD seems pleasantly naïve with a lot of work and of questionable value for my university's instructional purposes.

*Metadata:* Beyond basic identification information, human cataloguing or input of extensive metadata ala METS is an overblown consideration from a practitioner perspective. Anything that cannot be automatically ingested by the machine implies extra costs and is suspect.

*Search Engines Considerations:* All Web site producers should consider search engines as an audience and the major determinant in driving traffic. Despite the necessity of great care in design for navigation, search engines frequently tunnel around entry portals and directly to your contents. It thus helps educational institutions if Cultural Heritage materials are properly labelled for citation purposes. In lieu of advanced Bayesian mathematics and tuning relevancy engines - recommended design for native discovery normally features registration with DMOS, well constructed Title field, use of HTML hierarchical headings, and exchanging links with fellow repositories and institutions.

*Section 508/Handicapped Student access:* American universities taking Federal Student Aid are required to make accommodations for handicapped students. Thanks to the World Wide Web Consortium's (W3C) commitment to Universal Access Principles, compliance is typically rather simple for electronic textual materials. Cultural Heritage products, however, may need to be interpreted through optical character recognition (OCR) or word-processed transcripts (See JSTOR/Ithaka's new Decapod project for small institutions). Images call for parallel preparation - normally implemented through extensive use of the "Alt" descriptive tag. Media, especially the interactive variety, is more challenging, but we are exploring advances in voice recognition to simplify transcriptions, where those are not already provided.

*Standards:* Long-term solutions or standards from Cultural Heritage or traditional communities are of little interest. We look instead to the marketplace for answers. In Online Higher Education, SCORM compliance is expected for Learning Management System (LMS) operations. Some of us are also active in the ongoing development of Common Cartridge applications, which should draw your attention.

*Transparency/Community Building:* We strive to reduce the learning overhead and raise the comfort level on the Web for the students and faculty with the crucial subtext of building a lasting campus community. Overly jargoned or unnecessarily erudite sites and materials, which require extra hoops to retrieve, are not favourably received. Ideally, we would hope for continuity and seamlessness among Cultural Heritage sites. As will be reinforced later, technical goals extend with the Web's movement toward interactive customer services - ala the all-in-one, personalized "my portal" carrels. In our case, tailored virtual study arenas should be available including scheduling, classrooms, course materials, and social networking options. Amazon-like profiling and RSS feeds add to the future picture and could include Cultural Heritage materials - and communities.

*Web & a Hyper-Incunabulum:* Ultimately, it took Aldus Manutius in Venice - the inventor of portable books, italics, modern diacritics - and other printer/publishers the half-century or "Incunabulum" after Gutenberg to mature the book. The Web is engaged in a similar evolution. Permanence cannot enter our present calculations-planning logically extends no more than five years. At the moment, my staff is engaging mobile and other possible delivery platforms - like the Kindle and Web PC. We also proactively engage communities wherever our students are congregating on the Web - e.g., Facebook, Twitter and on our island in Second Life. I even lecture on the rapid onset of Web 2.0 as a Third Wave of development in online education and stretching the boundaries beyond the LMS. Yet, those talks must also postulate the rapid onset of a Fourth Wave. Potentially transformational applications with voice recognition and, especially, touch screens are already in the queue. More importantly, a very different thinking and technically savvy Millennial generation is entering our colleges; it will likely have a Manutius or two.



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## Entrepreneurial Context/Example

The first publishers are often viewed as the first modern capitalists - they commoditized information and helped launch a Renaissance in the process. That role is being carried forward in the Web Age by online universities. Since the start of the 21<sup>st</sup> Century, online programs have been by far the fastest growing element in Higher Education. The sector has quickly grown to command some 10% of the market and is continuing to grow in excess of 12% a year. The prime example for this presentation, the American Public University System (APUS), is part of a handful of elite onlines that have earned regional accreditation - the *sine qua non* for American colleges. Since earning that status in 2006, APUS has been in hyper-growth. It now delivers Web-based courses to some 60,000 students in 150 countries at the bachelor and master's level. Moreover, the business has gone public with listing on the NASDAQ stock market (APEI symbol); hence, accountable to Wall Street and needing to speak differently than to archivists, curators, and librarians.

Accreditation standards aside, education in online universities differs in a number of ways from what the reader likely experienced. Ironically, Onlines contain throwbacks to the very origins of the university in 13<sup>th</sup>-Century Bologna. The dialectic again reigns, and every student must "speak" every week. Classes are small and not framed for the professor's schedule, but oriented to the convenience of the student client. Instead of two or three semesters a year, APUS launches 8 and 16 week sessions every month. Rather than having to go to a physical classroom and listen at set hours, the Web is the venue and asynchronous education the mode. Discussion groups can be picked up at any time of the day and any place with an Internet connection. Web-based instruction within LMS software brings unprecedented tracking and metrics on students and faculty alike. Our faculty are not physical campus assets, but scattered around the country and the globe. Moreover, the School's business model rigidly controls for pricing. APUS has not raised tuition in a decade and pioneered underwriting the entire costs for its undergraduate course materials.

APUS looks favorably on Cultural Heritage. Unlike many for-profits, the University has not abandoned the humanities or the arts; History remains among its largest programs. Rather than outsourcing or minimizing, the School is also distinguished by significant and ongoing commitment to an academic library - a traditional seat for Cultural Heritage. Regionally or even national-distance accredited schools can have community service as an evaluation factor. In APUS's case, this factor includes outreach to the Cultural Heritage of its home base in West Virginia. Moreover, regionally accredited universities are inexorably linked to the established idea of the university - a concept that arose from the same romantic and nationalistic forces that effectively gave birth to the separate "cultural" sphere.

### The Idea of the University

John Henry Newman's famed *Idea of the University* marked an awkward transition. The Germanic-inspired 19<sup>th</sup>-Century New University Movement redefined Higher Education in a dramatic fashion away from his religious and classical ideals. In the United States, a general rise in literacy from mandatory grammar schools and the revolutionary appearance of a Mass Press set the stage for reforms that followed from Civil War-era Morrill Act. Centuries-long adherence to the liberal arts gave way to practical curriculum. Now familiar, albeit then new departmental structures appeared from agriculture and engineering to history and political science—along with budding professional presence for business and even librarianship.

As epitomized by the University of Illinois, monumental library structures appeared to define the heart of new campuses and as part-and-parcel of an overriding scientific research agenda. The library provided the primary market and storage place for outputs from new Ph.D. degrees with their publish-or-perish spate of scholarly journals and monographs. The Academic library shared the stage with newly invented public libraries, "the people's university" and rapid onset of museums and national libraries in the era—institutions characterized by a theoretical commitment research and development of taxonomies to control the rapid expansion in knowledge and print.

While linked to the historical lineage, APUS's specific version of the academic library could neither luxuriate in romantic assumptions, nor bask in a huge building with millions of volumes. It had to build anew within the context of the Web Revolution and competitive nature of a for-profit institution. APUS does not own, for



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

example, but must follow a new mode by licensing e-books and electronic runs of scholarly journals. The new Online Library also defaulted any thought of holding all the world's knowledge to the Information Highway.

Our analysis revealed an inherent contradiction and opportunity from the past. With the exception of course reserves, classroom support was largely missing from the traditional scenario. American academic libraries would even eschew a role in the post-WWII democratization of education. They deferred to that era's textbook solution. Much to the confusion of generations of students, libraries would neither collect nor, if they had them, loan textbooks.

In a major paradigm shift, APUS's virtual facility and Collection Development were deliberately re-engineered for direct course support. Librarians were recruited as subject specialist to proactively create course-specific electronic bibliographies and support portals. The business goal was a 15 to 1 ROI (return on investment) through projected savings from electronic course materials. Entrepreneurship did not stop with the Library segment. The second leg of a tripartite strategy featured the Library absorbing and revamping bookstore operations along related lines. Rather than a pass-through, they became active bargaining agent in a drive to control rampant inflation in textbooks. The third deals similarly with contradictions in the university press. Instead of monographic lists, the APUS model refocuses a university press toward classroom support by commissioning select online textbooks and course packets. In sum, we present a new type of client for Cultural Heritage.

### **Opportunities - Engineering for Education**

In addition to a fascination with the unfolding Web Revolution, let me admit to personal biases for Cultural Heritage efforts writ large. I come from a small island off Southern Louisiana. In my day, children were not allowed to speak French on the school yard. In the 1970s and 1980s, chefs, folk artists, musicians, government officials, scholars, and the people got together to instead celebrate their culture. Today, the term "Cajun" is internationally known, and I can find Louisiana-style restaurants around the world.

From a business orientation, however, Cultural Heritage products are a typical—albeit often technologically basic parts of the collection development spectrum. Providers are no different than the publishers and vendors that we deal with on a daily basis. Yet, their informational products are not like buying a car or computer—course materials help define the nature of our educational ventures. That factor along the uncertain nature of the Web and its "long tail" alter the simple buyer/seller relationship. Not unlike Erasmus working with Manutius in projects that helped launch personal authorship and humanistic scholarship, my office regularly engages in partnerships with publishers and vendors. The likes of Pearson, Wiley, Ebrary, and even JSTOR's Ithaka Sustainability Project seek our input and to collaborate with us in these challenging times.

Herein lies opportunity, but also the need for Cultural Heritage institutions to engineer a portion of their portfolio for classroom related services and Web product development. Standing still clearly will not work—especially as you digitize and publish results on the Web. Scholars, aficionados, and other institutions are already mining your resources. Yet, opportunities do exist for levels of engagement beyond particular collection, genre or provenance approaches driven from the Cultural Heritage institution. Although individual institutional efforts along these lines are laudable, especially those in the K-12 arena, they do not fully serve Higher Education purposes. A university focus turns on disciplinary requirements and teaching packets for specific courses from multiple resources and with multiple forms of Web delivery. My school would look for partnerships on our specific curriculum. In terms of applications on the immediate horizon, could we engage you in educational product development using budding Voice Recognition systems - or how could the new dynamics of touch screen applications add depth and different nuance to your materials?

The vision for Cultural Heritage pushes even higher. The Web's long tail beckons the evolution of complex Knowledge Communities as worldwide networked universities of the future. Individual and multi-institutional resources are marshalled around a subject, geographic area, or a people. To that base are added scholarly interpretative resources, but also schools of thought and opinion. Academics would be joined by curators, archivists and librarians with their own set of complimentary skills - as well as students and interested amateur. Discussion space is supplemented by wikis, simulations, and still unforeseen applications. Individually tailored and group spaces are enhanced by dynamic updating with new relevant monographs, journal articles, and discussion topics.

### **Conclusions**

On one hand, the vision of Knowledge Communities is an unfolding challenge and an offer of cooperation. On the other hand, my conclusions do contain existential dangers. Unless carefully orchestrated, digitization and allied publication on the Web can make much of Cultural Heritage obsolete. This is the logical culmination of Manutius' drives and Trithemius' related lament *In Praise of Scribes*. With the transcendent appearance of the



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

printing press 5,000 years of scribal cultural was effectively eliminated in short order. Intelligent educational packaging can help obviate that danger for Cultural Heritage in the Web Age.

That said, allow me to close on a very different note from my own time in Special Collections and Cultural Heritage before being recruited to the world of online education. In the late 1990s, I was visiting Jac Treanor, archivist of the Archdiocese of Chicago, along with Monsignor Charles Burns - then head of the Vatican Secret Archives. Discussions turned to labor negotiations that I was having when Burns piped in with his Scottish lilt. Just before leaving Rome, he had been similarly engaged - albeit through an older work-related document. St. Peter's Cathedral was under construction, but the Cardinals were in conclave and the project running out of money. The foreman passed a note in. "For the sake of St. Peter" (purportedly the origins of the aphorism - "for Pete's sake), if funds are not made available, I will need to shut down this job. Signed: *Michelangelo*. Please keep in mind that there is no way that my beloved Web and digitization can duplicate the resulting "Michelangelo Effect" - with its visceral, associational and tactile pleasures.

## **Silvia GSTREIN, Günter MÜHLBERGER**

### **User-driven content selection for digitisation - the eBooks on Demand Network**

#### **Abstract**

Following EC's i2010 strategy European libraries are currently systematically digitising and making available their cultural heritage. Very often, however, researchers, readers and users demand access to historical books „here and now“. This is exactly where the eBooks-on-Demand (EOD) network starts from, providing a trans-European digital document delivery service for end-users from all over the world. Currently the EOD network comprises over 20 libraries from 10 European countries. Since 2007 several thousand PDF eBooks have been produced, delivered to users from over 30 countries and made available to the public through participating digital libraries. User reactions are very positive and more and more libraries are interested in the service. This paper is about the following topics: 1) the user-driven approach of content selection for digitisation, 2) financing such a service, 3) results from a user survey, 4) future perspectives within supplying web 2.0 platforms.

**Keywords:** Digitisation, eBooks on Demand, user selection

#### **Introduction**

In line with EC's i2010 strategy European libraries and other cultural institutions are currently systematically digitising and making available their cultural heritage to a wider public. Due to the enormous amount of material, it will take some decades until all books, journals and other library material will be available in digital form. Ideally, all the works in a collection or holding should be considered for digitisation. In practice, however, this is rarely feasible and choices must be made. According to the MINERVA Good Practices Handbook (Clissmann 2004) the following criteria may be suggested for setting preferences: 1) Access to material which would otherwise be unavailable or of limited availability, 2) wider and easier access to very popular material, 3) condition of the originals, 4) preservation of delicate originals by making digital versions available as an alternative, 5) project theme, 6) copyright and IPR, 7) availability of existing digital versions, 8) cost of digitisation and 9) appropriateness of the source material for online viewing.

In many cases the chosen approach is determined by the source of funding, preferably in combination with the institution's digitisation policy. This way of processing digitisation only very rarely acknowledges interests of users apart from point 2) above which highlights popularity as a criterion of selection. This raises the following questions: 1) how can individual users' needs be considered in digitisation?, 2) what happens to and who will take care of those materials not covered by all those projects?, 3) more precisely, who will digitise books of minor languages or those from special or smaller collections or institutions?

All those considerations formed the starting point for a service called "eBooks on demand": an electronic document delivery service instigated and co-funded by the user. What was needed was a model for ongoing, permanent digitisation, a coordinated European initiative where selection of material necessarily works bottom-up rather than top-down and which is instigated by the individual reader.

Experience has shown that such a project needs co-financing by the reader requesting a certain book to be digitised. By saying so I only refer to the co-financing of scanning itself. The important but often neglected issue of subsequent costs such as storage, access, long term preservation and migration costs will not be touched here.

#### **EOD network and service**

Of course, such an additional library service can't be maintained within the context of usual (mass) digitisation only, but needs extra resources for order management, customer communication, payment procedures and so on. Therefore, a structure allowing for efficient processing of orders was designed. At the end a central-decentral service network was implemented, where some processes such as OCR or online payment are hosted centrally. From October 2006 to June 2008 a pilot project and evaluation was carried out within the framework of "Digitisation on Demand", co-funded by the eTEN programme, consisting of 13 libraries from 8 European countries. By July 2008, a self-sustained network was up and running with these 13 libraries as founding members.

In the following I will briefly describe how EOD works. The starting point is the online catalogue of any participating library. There, the EOD button is placed with all items available for digitisation. At the moment, these are books which fulfil at least one of the following criteria: 1) items not yet digitised, and 2) public domain books. Any user interested in a certain book tagged with the EOD button simply needs to click on this button.

The library then receives the order, scans the requested book and transfers the replica via FTP to the central server located at the network coordinator Innsbruck University. Each library manages and processes the orders in a central database accessible via a normal web browser. Here, also the digital object is created. After the completion of the payment process, the user downloads the PDF from his personal tracking page. After some time the library eventually integrates the images into its digital library or repository and thus makes them available to the general public.

From a library's point-of-view the EOD service is a mixture of central and peripheral services sharing two core components: The order data manager and the digital object generator. For its own orders only, each library has access to the Order Data Manager, a central database with web-interface. According to its needs and objectives any participant is able to adapt and customize the website texts, automatically generated email texts, etc. Core part of the Order Data Manager is the Digital Object Generator, a central service for generating eBooks. It allows OCR recognition (antiqua and gothic), automated cover generation, PDF & RTF delivery, creation of Abbyy XML and the generation of the streaming link for downloading. Online payment in the form of credit cards also forms part of the central services – so the individual library doesn't need to grapple with business like this.

Currently, more than 20 libraries from 10 European countries are offering the service, ranging from Portugal in the far west to Estonia in the East. The University of Innsbruck library is both, the coordinator and central service provider of this network at the same time. Concerning the types of libraries involved - there are 6 national libraries, several university and state libraries as well as research and academy of sciences' libraries.

<p><b>Austria</b></p> <ul style="list-style-type: none"> <li>• University and Regional Library of Tyrol (co-ordinator)</li> <li>• University Libraries of Graz and Vienna, Library of the Medical University of Vienna</li> <li>• Vienna City Library</li> </ul> <p><b>Germany</b></p> <ul style="list-style-type: none"> <li>• University Libraries of Regensburg, Greifswald and Humboldt-Universität zu Berlin</li> <li>• Bavarian State Library</li> <li>• Saxon State and University Library (Dresden)</li> </ul> <p><b>Denmark</b></p> <ul style="list-style-type: none"> <li>• The Royal Library</li> </ul> <p><b>Estonia</b></p> <ul style="list-style-type: none"> <li>• National Library</li> <li>• University Library of Tartu</li> </ul> <p><b>France</b></p> <ul style="list-style-type: none"> <li>• Medical and Dental Academic Library of Paris</li> </ul>	<p><b>Czech Republic</b></p> <ul style="list-style-type: none"> <li>• Moravian Library in Brno</li> <li>• Research Library in Olomouc</li> <li>• Library of the Academy of Sciences in Prague</li> <li>• National Technical Library</li> </ul> <p><b>Hungary</b></p> <ul style="list-style-type: none"> <li>• National Széchényi Library of Hungary</li> <li>• Library of the Hungarian Academy of Sciences</li> </ul> <p><b>Portugal</b></p> <ul style="list-style-type: none"> <li>• National Library</li> </ul> <p><b>Slovakia</b></p> <ul style="list-style-type: none"> <li>• University Library in Bratislava</li> </ul> <p><b>Slovenia</b></p> <ul style="list-style-type: none"> <li>• National and University Library</li> </ul>
--	---

Table 1: Overview over EOD libraries

### Pricing and user experience

Up to now some 3200 books have been digitised and delivered to customers - books that wouldn't have been otherwise digitised. This makes up some 840.000 scanned pages for nearly 2000 users worldwide. The top 3 libraries receive about one request per working day and deliver 250 to 350 books per year. The average delivery time is one week.

The average price of an order is about 50 EUR. The price of an eBook is set by the respective library and is calculated from a base price plus the number of the pages. There is still a quite heterogeneous pricing system as at the moment each library sets up its own prices. In average, digitisation via EOD includes a basic fee of 10 Euro plus 0.15 to 0.30 Euros per page. In view of this price range, a certain degree of harmonization certainly needs to be agreed upon in the near future.

Experience has shown that the revenue gained from digitisation merely co-funds digitisation – to be more precise: the actual personnel costs for scanning including additional costs for the service as such. Quite clearly then a customer-oriented service leads to higher costs than mass digitisation, which lays huge emphasis on efficiency and economisation. Given, though, that the master files remain with the library and is later made freely accessible to the public, the user by no means can be made to pay the actual total cost of digitisation.

The user must be seen as the instigator and co-funder of digitisation; in turn the library profits by building up its digital library in the long run.

In 2008 we carried out a user survey among EOD customers to receive feedback to the service. We also wanted to find out what users thought about our pricing policy. The findings were that 30 % of our customers thought the price was high or very high, but still the overall price-value relation was found acceptable by the majority.

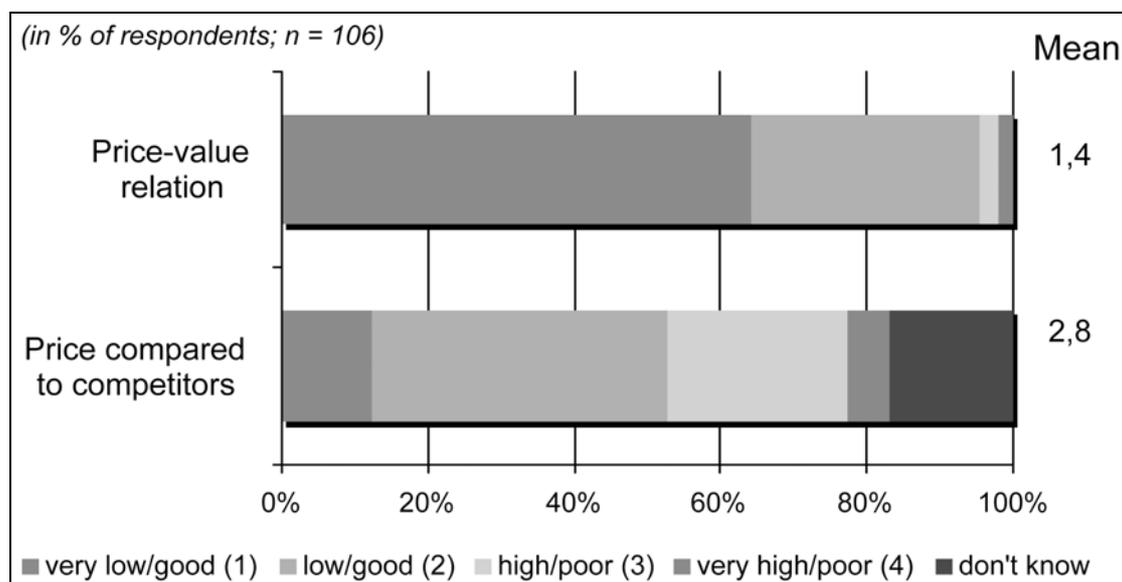


Figure 1: Users' evaluation of price-value relation

Its noteworthy that the overwhelming majority of users are either researchers or people who require eBooks for "professional or scientific use" (over 60 percent). Second place (16 percent) is book collectors and people who could be said to count among special interest groups such as amateur historians, collectors, or ethnographers for example.

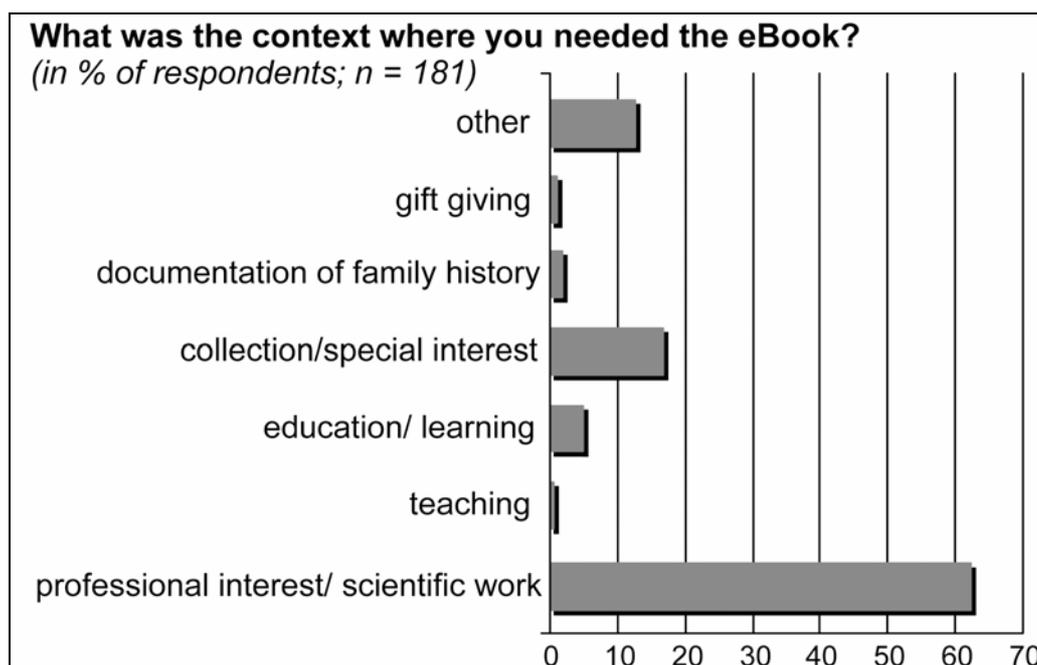


Figure 2: Users' domain of interest

Asked why they chose the EOD service, almost half of the interviewees answered that without EOD the book would have been “impossible or difficult to access”. This shows that EOD achieved one of its main goals of being an additional alternative for accessing books.

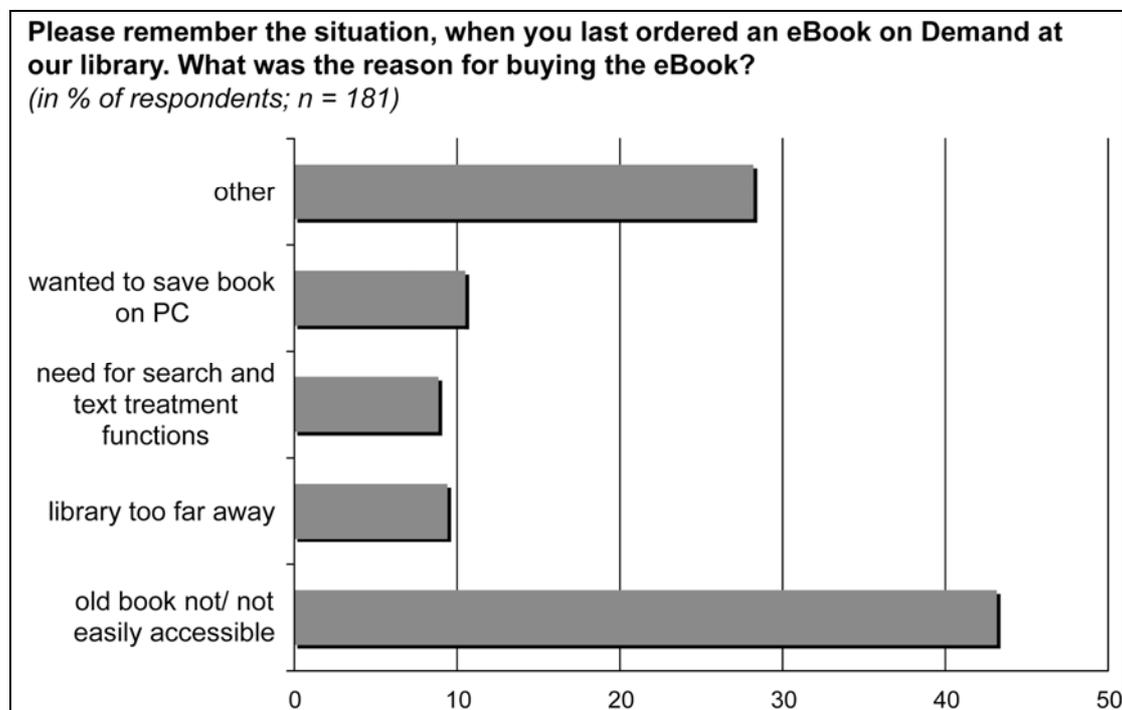


Figure 3: Reasons for ordering

In the EOD customer survey, 60% of our customers told us that they usually print out selected pages or even the whole book on paper. Thus, there was an obvious request for “re-materialization” of digital material. In response, we have lately been offering “reprints on demand”, supplementing the digital file. This service is also carried out in a centralised way. Each library only needs to take care of scanning images and some more metadata. Everything else connected with the service such as image enhancement, the creation of pre-press PDF and related files is carried out by the central coordinator at the University of Innsbruck library.

### Future perspectives

Within the EC Culture programme a new project was initiated in May 2009 with duration of 4 years. The project will focus on larger scale involvement of three main target groups: participating libraries, requesting users and the general public. 20 libraries from 10 European countries take part in this project which mainly concentrates on the following 3 objectives: 1) to enlarge the EOD network by additional European libraries, especially those from countries not yet represented 2) to take EOD as a best practice model for any other European-wide network of this kind; furthermore to train stakeholders (libraries, museums, or other cultural operators) to run such a multinational cultural service based on state-of-the-art information technologies and 3) to support intercultural dialogue among readers and users of historical books with the help of web 2.0 technology.

Within this latter goal web 2.0 based social platforms such as Wikipedia, LibraryThing or Goodreads will be supplied with information on selected historical books. Readers all over the world interested in a specific book thus will be able to easily interact with each other, exchange information and share reading and research experiences – independently from where they are and which background they may have.

During the last years the internet has become more and more of a “social network”. Cultural organisations need not only “offer” their cultural heritage in the internet, but they need to pro-actively “supply” web 2.0 platforms with their content, since these are the sites where the majority of users are actually present. Not only do users create and edit information, but upload texts, images, and videos, they “meet” other users, they “share” content, they “cooperate”, “rate” and generally “take part” in a digitally mediated community. Whereas many sites, such as library catalogues or repositories, are only visited sporadically in the event of retrieving some clearly defined data, people have their personal accounts at web 2.0 applications and use them regularly – comparable to the way they regularly read a newspaper, meet friends at a cafe or visit a museum. First, EOD will therefore



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

explore the technical feasibility to automatically generate Wikipedia source-code for selected books which have been ordered and digitised via the EOD service to then later on add metadata to articles about the respective authors. Other social platforms will be tackled in a second step.

## Conclusions

Experiences made so far indicate that “on demand” services are desirable ways to make books accessible either in digital or in re-printed form. EOD has proven to be an important additional library service, making holdings all over Europe more accessible to customers and providing researchers with material they need “on demand”, here and now.

In fact more “on demand” services could be imagined, such as digitisation on demand for the blind and visually impaired and the creation of „real“ eBooks with a corrected full text approximating 99% accuracy. This would pave the way to transfer eBooks onto mobile-devices, making the written word available every time, everywhere.

## Acknowledgement

This work programme has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

## References

- [1] Clissmann, Ciaran (2004): Good practices handbook. Version 1.3. Minerva Project. Minerva Working Group 6 Identification of good practices and competence centres. Online: [http://www.minervaeurope.org/structure/workinggroups/goodpract/document/goodpractices1\\_3.pdf](http://www.minervaeurope.org/structure/workinggroups/goodpract/document/goodpractices1_3.pdf) [Access: 25/10/2009]
- [2] Mezö, Zoltán; Svoljsak, Sonja; Gstrein, Silvia (2007): EOD - European network of libraries for eBooks on demand. In: Research and advanced technology for digital libraries, p. 570–572.
- [3] Mühlberger, Günter / Gstrein, Silvia (2009): eBooks on Demand (EOD): a European digitisation service. IFLA Journal (35/1), 35–43.
- [4] Online: <http://archive.ifla.org/V/iflaj/IFLA-Journal-1-2009.pdf> [Access: 25/10/2009]
- [5] Svoljsak, Sonja; Gstrein, Silvia (2007): EOD - eBooks on demand. In: Users and use of DL & economics of digital libraries, p. 229.

## **Aly CONTEH, Asaf TZADOK**

### **User collaboration in mass digitisation of textual materials**

#### **Abstract**

By utilising web-based collaboration tools, institutions can engage users in the building of historical printed text resources created by mass digitisation projects. The paper presents the drivers for developing such tools and identifies the benefits that can be derived by both the user community and cultural heritage institutions. The perceived risks, such as errors introduced by the users or whether users will engage with resources in this way, will be set out. The paper will present the lessons that can be learnt from existing activities, such as the National Library of Australia's newspaper website, which supports collaborative correction of Optical Character Recognition (OCR) output.

The user collaboration tools being created by the IMPACT Project (Improving Access to Text, <http://www.impact-project.eu>), a large-scale integrating project funded by the European Commission as part of the Seventh Framework Programme (FP7), will be detailed. A primary aim of IMPACT is to develop tools that help improve OCR results for historical printed texts, specifically those works published before the industrial production of books in the middle of the 19th century.

While technological improvements to image processing and OCR engine technology are key to improving access to historic text, engaging the user community also has an important role to play. Utilising the user community can aid in achieving the levels of accuracy currently found in born digital materials. Improving OCR results to this level is key to producing resources that support better resource discovery and enable greater performance when applying text mining and accessibility tools to the extracted text. The IMPACT project will specifically develop a tool that supports collaborative correction and validation of OCR results and to allow user involvement in building historical dictionaries that can be used to validate word recognition. The technologies use the characteristics of human perception as a basis for error detection.

**Keywords:** Digitisation, OCR, User Collaboration, IMPACT

#### **Introduction**

##### **The Challenge**

The digitisation of historical text resources and use of sophisticated software tools to translate the images of text into machine-readable text has transformed the way researchers engage with these types of resources. CENL (Conference for European National Librarians) surveyed its members in 2008, revealing an expectation of a 350% increase in the digitisation of historical books and newspapers between 2006 and 2012, which would make them the most popular type of material being digitised[1]. The benefits of OCR (Optical Character Recognition) in the digitisation workflow were recognised but the experience at the British Library in their project to digitise 19th Century newspapers indicated that there were issues in the quality of the OCR text with, on average, over 20% of the text on a page not being correctly recognised [2]. Many factors influence poor performance, such as quality of the original material, storage practices, and the fonts and languages used. Current OCR is tuned to processing modern printed text and so there is a need to improve the performance of OCR tools when dealing with historical texts.

Achieving 100% percent or even the 99.9% accuracy that is usually specified through completely automated solutions will therefore be difficult to achieve for historical text. Indeed, when accuracies of this level are required then the preferred solution is to get human operators to re-key the data. While this produces good results, it is not scaleable and when institutions are digitising millions of pages the costs are prohibitively high. The cognitive ability of the humans make them ideally suited to task of recognising text that computers cannot. So there is a need to harness that power in a way that can scale to support projects that are digitising millions of pages of text. The ideal solution is to partner the strengths of the computer and humans to achieve our goals for accuracy.

##### **The Power to be Harnessed**

The advent of technologies which carry the moniker Web 2.0 technologies, the interactive web, has meant users are not just presented with information but play a full part in the creation, enhancement and semantic mark-up of information. Amazon, Wikipedia, Youtube and Facebook all provide evidence that the web community has fully embraced this paradigm, indeed Flickr recently announced that the 4 billionth photo had been uploaded [3].



The ability to harness the millions of users who interact with web-based cultural heritage resources to fill that gaps that OCR software leaves behind is attractive, but there remain questions as to whether users will engage to the same extent in collaborative correction.

The concept of user collaboration in the clean-up of OCR text is not a new one. Distributed Proofreaders have used a volunteer force of over 4,000 people to correct the OCR text of over 16,000 titles in a nine-year period [4]. While this approach demonstrates that users are willing to engage in such activity for the benefit of the community, it is not a mainstream activity routinely deployed in the digitisation workflow of cultural heritage institutions. There is a need to significantly increase user throughput to match the levels of digitisation that is currently under way.

Two recent initiatives provide further demonstrable proof that user collaboration is a key tool to resolving the accuracy gap over purely computational approaches. Indeed a fusion of the computational approach of modern computer software allied with the cognitive power of the human brain can help us achieve the results we seek.

### Case Studies

reCAPTCHA [5], which has a tag line of “Stop spam, Read books”, is the result of a project undertaken by the School of Computer Science based at Carnegie Mellon University. CAPTCHA stands for Completely Automated Turing Test to Tell Computers and Humans Apart, and are text forms used by websites to prevent automated programs from accessing websites for malicious purposes. reCAPTCHA uses computer-unrecognised OCR results for these text forms, and thereby uses human interaction to improve the initial OCR results for historical documents. There are no published statistics on the number of words that reCAPTCHA has corrected but over 200 million CAPTCHAs are solved every day [5] and almost one billion reCAPTCHAs were solved in 2007/2008[6].

In July 2008 the National Library of Australia released a public beta of the Australian Newspaper service [7] which supported public collaborative OCR text correction. It was the first service of its kind. It was a low-key launch with a desire to get feedback from early users to see what they required from such a service and how they would engage with the service. One part of this approach was to leave moderation of user changes to the community.

In the first 6 months of the project the usage of the service was as follows:

- 2,994 registered users
- 2.2 million lines of text corrected
- 104,000 articles corrected

The verdict of both the National Library of Australia and the site’s users on the first 6 months of the service is overwhelmingly positive. The issue of users making malicious changes did not surface and no vandalism of text was detected. Indeed, users validated the community moderation approach by explicitly stating that users should moderate each other, rather than the Library moderating, and have the ability to report and correct issues.

These two examples demonstrate that user communities are a power to be harnessed in improving the quality of OCR text.

The IMPACT approach to Collaborative Correction

IMPACT (Improving Access to Text, <http://www.impact-project.eu>) is a large-scale integrating project funded by the European Commission as part of the Seventh Framework Programme (FP7). The project commenced in January 2008 with the following objectives:

- Develop OCR software and technologies which exceed the accurateness of current software significantly.
- Provide a software system which will allow the realisation of new concepts of collaborative correction (in order to lower the costs for full featured full-text) by taking up and integrating Web2.0 concepts
- Develop language tools and lexica in order to provide access to historical texts independently of historical variants of a given language.
- Support adopters of these tools so that more European historical lexica can be built.
- Develop a number of smaller modules such as image enhancement and segmentation toolkits, functional parsers, etc. in order to support the automated text recognition and/or access to historical text.

It was recognised that while the project could seek to advance the state-of-the-art for language and OCR tools, there was a need to provide advanced solutions in engaging users to improve the word accuracy of digitised historical texts. The context for this is the i2010 vision of a European Digital Library: an ambitious plan for large scale digitisation projects that will transform Europe’s printed heritage into digitally available resources.

Building on the experience of such initiatives as reCAPTCHA and the Australian National Library's Newspaper project, IMPACT will develop an alternative approach to user collaboration in the clean-up of OCR text which will increase the power of these types of tools.

The tools will allow for involvement of the general public in validation and correction of OCR results. These tools will be based on the SmartKey idea described below. This technology uses the characteristics of human perception as a basis for error detection. The result is a data acquisition procedure that is very efficient and virtually error-free.

The OCR engine recognition process concludes with a rating score for each character. These scores are further refined by the spell checker, which can either increase or decrease the probable success of individual characters. Based on such probabilities, all the characters are classified as Sure (characters with a high enough probability), Medium or Unsure (characters with a low enough probability). While Sure characters need no further verification and can be accepted automatically, and Unsure characters are sent for manual data entry, the Medium characters are sent for fast verification via "carpets".

In a "carpets session", all the Medium characters from different sources (possibly different pages, chapters or even books) are sorted in alphabetical order. For example, all the characters that were recognised as an 'A', but with a low score rate, are grouped together in a single "carpet". It has been shown to be very easy to identify the few errors and thus automatically approve the other characters as valid 'A's. Hence, instead of keying in 100 characters, it is sufficient to point the mouse at few errors and the others will be automatically deduced. As a result, the validation process becomes very fast and effective. Figure 1 provides an example of a carpet session where the user is asked to identify all items which are not the letter e.

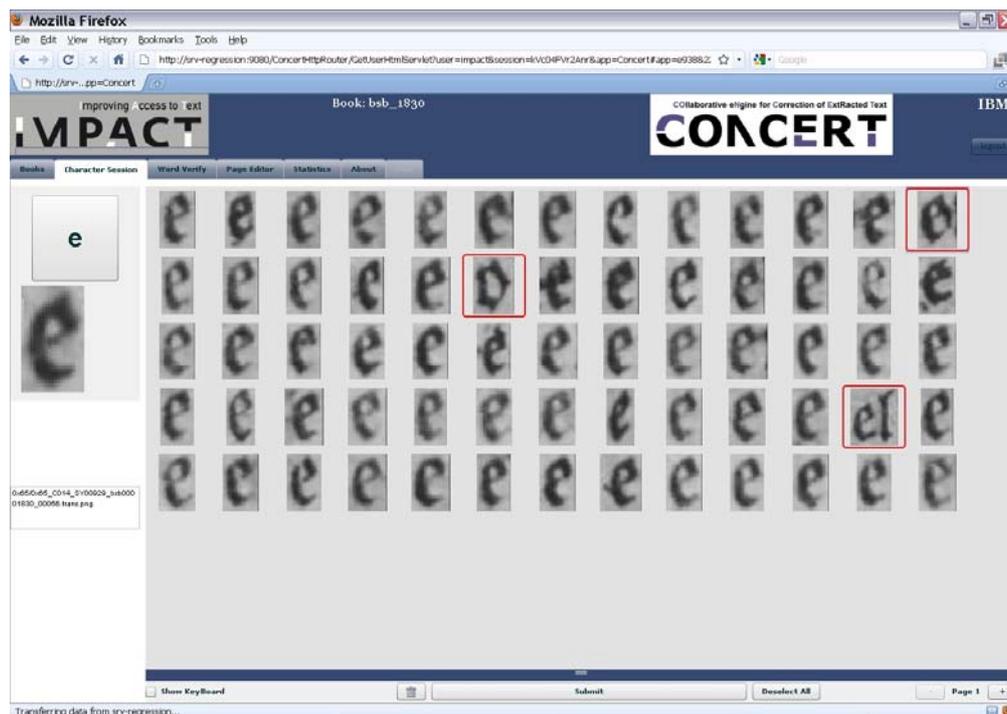


Figure 1 A View of a carpet session

At the end of this process, some words may remain unrecognised. Indeed, if image quality is very poor, it may be impossible to recognise a given character without the context of the entire word or sentence. For these infrequent cases, word-based data entry will be introduced and context information made available as necessary. In this way the user can also add words to the dictionary supporting the OCR process by identifying words which are not currently included.

This dictionary is part of an Adaptive OCR Engine and user collaboration will therefore not only correct words but train and enhance the engine's vocabulary and language analysis features. The system will dynamically decide how much manual intervention is needed to achieve a certain level of accuracy.

Quality monitoring will be enabled by feeding known errors into user "carpets" and seeing whether the user detects those errors. If the user doesn't, their results will be weighted as less accurate than those of someone



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

who does. In extreme cases, the user may be defined as a malefactor and his/her contribution discarded altogether. Quality monitoring will be done online to facilitate the adaptive utilisation of the results.

To summarise, the IMPACT approach to collaborative correction involves:

- creation of a data validation/correction application that is simple and intuitive enough to be attractive to untrained users and yet effective enough to ensure high productivity
- developing a powerful control system capable of analysing and segmenting books and other documents into individual small jobs, ensuring successful job completion, and reassembling the final result
- creating a web-based application suitable for mass volunteer participation.

## Summary

Large-scale digitisation of historical printed text resources is a reality and is transforming the research landscape for this type of material. With the research benefits there are weaknesses, a key one being the quality of OCR text.

Advances have been made and will continue to be made in advancing the state-of-the-art in the automated translation of printed text into machine-readable text, and IMPACT is active in this area.

Significant benefit can be derived by using humans to bridge the gap between what can be done in an automated way and desired accuracy levels. Examples of this have been deployed with great success and will significantly improve digitised resources. IMPACT is introducing a new paradigm in this area, where not only is the text corrected but the user input is used to improve future OCR and language processing in a seamless manner.

Material that is born digital can have extremely high levels of accuracy, user collaboration will allow us to approach those levels of accuracy for historical texts in a cost effective manner.

## References

- [1] Hans Petschar, et al. "EDL Report on Digitisation in European National Libraries 2006-2012" February 2008. Available at < [http://www.cenl.org/docs/report\\_digitisation\\_nls.pdf](http://www.cenl.org/docs/report_digitisation_nls.pdf) >
- [2] Simon Tanner, Trevor Muñoz, Pich Hemy Ros. "Measuring Mass Text Digitization Quality and Usefulness" D-Lib Magazine. July/August 2009, vol. 15 no 7/8 < <http://www.dlib.org/dlib/july09/munoz/07munoz.html> >
- [3] < <http://blog.flickr.net/2009/10/12/4000000000/> >
- [4] < <http://www.pgdp.net/c/> >
- [5] < <http://recaptcha.net/learnmore.html> >
- [6] < <http://markmail.org/message/orklsq2e4csqr2s4> >
- [7] Holley, Rose (2009) "Many Hands Make Light Work: Public Collaborative Text Correction in Australian Historic Newspapers." ISBN 978-0-642-27694-0. Available at <[http://www.nla.gov.au/ndp/project\\_details/documents/ANDP\\_ManyHands.pdf](http://www.nla.gov.au/ndp/project_details/documents/ANDP_ManyHands.pdf)>

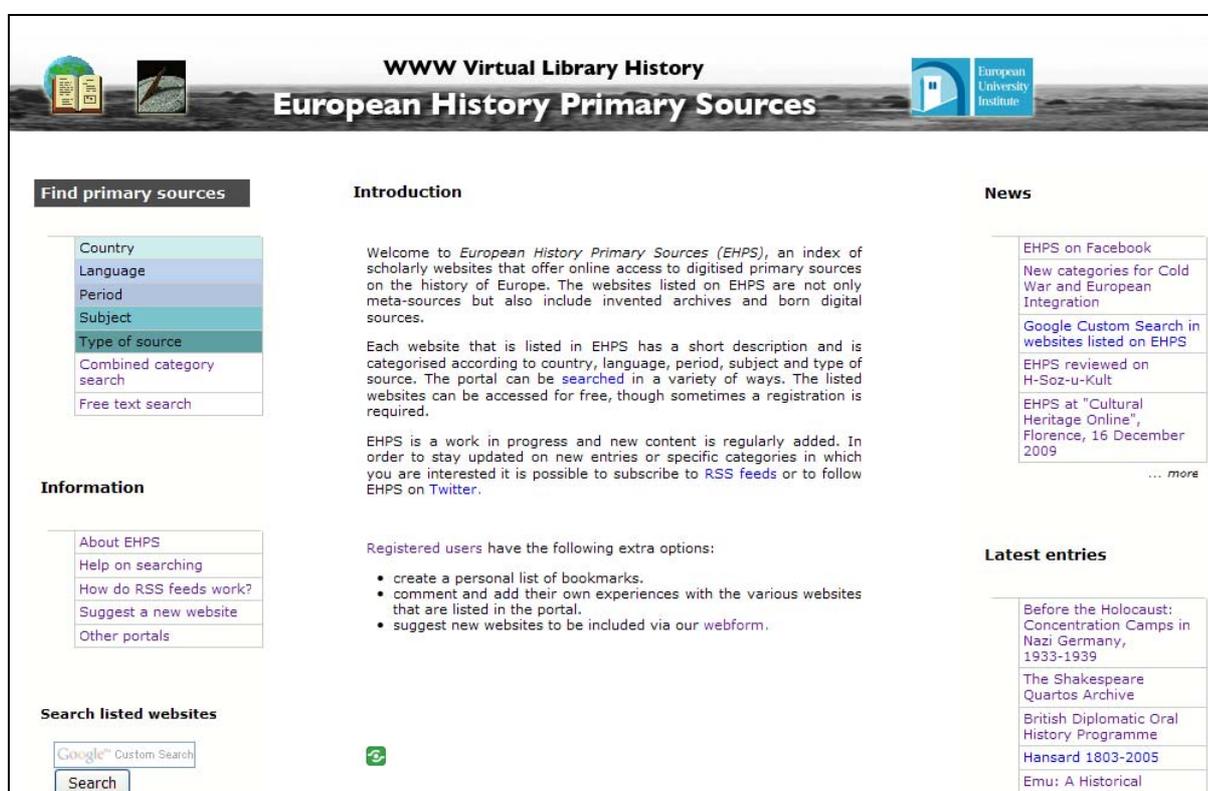
## Serge NOIRET

The European History Primary Sources (EHPS) portal at the European University Institute, Florence.

### Abstract

EHPS, [<http://primary-sources.eui.eu>], is a portal serving a community of Ph.D. researchers and post-doc fellows and professors at the Department of History and Civilisation its users with an easily searchable index of multi-lingual collections of scholarly websites that offer online access to digitised primary sources, invented archives and born digital sources relating to the history of Europe, either as a whole or for individual countries. EHPS offers web 2.0. features to remain connected to the portal and be informed about new entries.

**Keywords:** Portal, Digital History, Primary Source Collections, E-resources, Digital libraries, Meta-sources, Born digital Sources, History of Europe, European University Institute.



The screenshot shows the EHPS Home page with the following layout:

- Header:** "WWW Virtual Library History" and "European History Primary Sources" with the EUI logo.
- Find primary sources:** A sidebar with filters for Country, Language, Period, Subject, Type of source, Combined category search, and Free text search.
- Introduction:** A central text block explaining the portal's purpose and providing instructions on how to use it.
- News:** A list of recent updates, including "EHPS on Facebook", "New categories for Cold War and European Integration", and "EHPS at 'Cultural Heritage Online'".
- Information:** A sidebar with links for "About EHPS", "Help on searching", "How do RSS feeds work?", "Suggest a new website", and "Other portals".
- Search listed websites:** A search bar with a "Search" button.
- Latest entries:** A list of featured historical topics such as "Before the Holocaust: Concentration Camps in Nazi Germany, 1933-1939" and "The Shakespeare Quartos Archive".

European History Primary Sources EHPS Home page

### Introduction

This is the time for a new craft and a new methodology for a digital historian, a more common and day to day «pratique de l'histoire» as Marc Bloch would have said also using socially connected new web 2.0. features [1]. Historians and the communities to which they refer, need to organize better in their computers and through their browsers, the access to digital repositories and archives, manage data's through appropriate software's, archive in durable and secure ways their digital documents and artifacts and diffuse their writings in OA repositories.

Important digital archives and invented archives, research tools, digital libraries and other e-resources are available online. One may say that looking at these is now part of the process of selecting the necessary materials to write an historical essay. But new methodological issues are arising, not only for who's creating these new web-sites, invented archives and primary source libraries from physical materials, but also for who's using them for research and the writing of history both in a traditional way or again for the web.

The library supporting the research activities of historians at the EUI is not old and has no old collections of books and primary sources. It never even had a card catalogue but always had a computerized database of bibliographical records [2]. Such a small academic library couldn't afford to deal with a massive selective



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

project of digitization for primary sources collections even for the early XX century. In the 1980s the library was concentrated on the purchase of primary sources in microforms and, later, when the web developed itself, had to face the financial situation academic libraries are now trying to deal with coping with subscribing and purchasing digital materials, mainly in English for his multi-lingual and multi-national public.

So, accessing digital primary sources at the EUI is accomplished locally, retrieving the microform documents, going on missions to libraries and archives abroad or using the internet and a browser based procedure. EHPS has been created in this context to avoid -when possible- the use of costly and time-consuming missions abroad.

## A Digital Turn in History

Today, after the first important primary sources digitization projects in the 1990s in North America and Europe, we face new scenarios in history because the digital turn is now offering an abundance of digital primary sources worldwide. This is happening even if the American historian, one of the fathers of digital history, Roy Rosenzweig, challenged this idea of abundance: the immediate volatility and disappearance of new digital media formats and web-sites were challenging that assumption he wrote [3].

Kirsten Sword, professor of American women's history at Indiana University, suggest to use these new digital archives which are de facto offering an enormous potential for new historical inquiry: "digital resources are expanding and redefining the archival base for most fields and thereby redefining the fields themselves, she said even if "this is driven more by libraries and the tech industry than by historians" [4] and, during the same round-table about the "promises of digital history", Steven Mintz, past president of H-Net, wrote about what I would personally suggest to call a real digital turn in historical practices, that modified the way to make history: "it has greatly expanded the range of sources —primary and secondary— that I use..." [5]

During an international conference on Contemporary History in the Digital Age organized in Luxembourg in October 2009 [6], Marin Dacos, director of the Cléo, Centre pour l'édition électronique ouverte in the CNRS in France [7] (he founded in 1999 the project Revues.org) [8], insisted in his keynote speech that if a History 2.0 has to be defined, it is because we are building new cyber-networks for accessing our sources. He underlined the fact that our goal was to bring scholarly literature and the primary sources to everybody's own computer. This has to be organized in a open society model through web based 2.0 technologies and bypassing as much as possible, commercially owned technologies, commercial databases and private actors.

## Small scale Primary Source Collections in a fragmented internet

It is a fact that from the end of the 20th century, scholarly digital contents is each day more in the hand of few commercial e-actors which used the intellectual production of scholars to create new revenues from online access to their own scholarly contents. Essays written by academics were sold back to their libraries and their universities.

A Web 2.0. debate "avant-la-lettre" was organized online in 2001 by Noga Arikha and Gloria Origgi for the Centre Georges Pompidou in Paris. Eminent scholars dealing with the online cultural heritage participated [9]. The historian Theodore Zeldin wrote a paper on "The Future of Internet". During the online discussion about the paper, I asked the author of a history of an intimate history of humanity [10], about the growing internal fragmentation of the web and about the parallel existence of different levels of access to web contents. I summarized this closure of the access to web contents in French as "internet et les internet(s)" asking him what could still be done for supporting the free diffusion and access to scholarly materials —primary sources and secondary literature- in the humanities [11].

One of the possible response to such a commercial challenge, is that huge digital projects of primary sources digitization in history are also coupled with small scale projects. These new projects are often using new 2.0 web technologies to attract the user's capacity to add contents or refine existing description of contents. These projects are extremely useful to digital historian activities organizing better the interoperability and integration with their research needs. Digitizing few sources or a single collection of sources is also done today within smaller cultural institutions, to concentrate online on more specialized contents with high scientific standards and high technological added value. This is done allowing good retrieval procedures within the digitized documentation also for remote uses.

This new trend to support digitization of less pharaonic projects, with more precise and delimited contents and freely accessible, is a new recommended policy to support digital history activities. Edward L.Ayers, an American pioneer digital historian, realized between 1991 and 1993 a SGML project called «The Valley of the Shadow, two communities in the American Civil War» which became in 1996 a web project. After more than ten years of technical developments and the addition of an enormous amount of digitized primary sources about the Civil War in the USA, Ayers is now suggesting us to change our agenda and not to launch such large scale



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

projects: «we've tended to build big things in the hopes of capturing as many uses as possible. But maybe now we need to build lighter, smaller things. We might build simpler ways to use our vast collections.» [12]

### **The stages of the historian's workshop**

Historians confronted with the digital age and digital history, even for those not thinking at all of producing digital artifacts within a Ph.D. research, needs to adopt a new digital craft made of new methodologies and new critical paradigms [13]. This new “workshop” is made of different practical and methodological working stages interacting with the internet and the web. If we could divide the process of doing digital history, four different steps would describe a kind of working process with which an historian would be confronted starting from the “production” of the digital documentation up to an individual “consumption” of digital primary sources. This process of research summarizes the history digital turn and defines the practice of digital history. All the four different activities below, are each singularly and together too, defining the field of digital history as a way to produce history using the new media and digital primary sources now available on the web.

- Production of e-sources;
- Information about e-sources;
- Selection and evaluation of e-sources and meta-sources
- Use of e-sources to write history in a traditional way or again for the web.

### **EHPS: monitoring primary sources websites and collections**

But how to monitor, list, collect and organize the research of these scientific primary sources dealing with the history of Europe ? The European History Primary Sources portal tries to answer to this question dealing with the specific needs of a European community of historians based in Florence, Italy, at the European University Institute [14]. EHPS integrates itself within this new “digital turn” when not only digital historians will have to look for their documentation in the web because of the many added values the internet will offer them outside the traditional ways of dealing with primary evidences.

EHPS inaugurated officially in June 2009. While not claiming to be complete, it contains the major national digital libraries and many smaller series of e-sources and smaller digitization projects in Europe in all national and sub-national languages. It thus reflects in a way, the current state of digitization of historical source materials in Europe, as well as those digitized outside Europe pertaining to its history. The portal is aiming at tracing a map of all the different digital libraries and primary sources collections and databases available for the history of Europe from medieval times to nowadays in all European languages. EHPS precisely enters the second stage of the digital historian's journey, because its main goal is about informing of the existence of digitized primary sources and trying to offer its users a way to collect them and access them inside their browser and for further evaluation and uses, steps 3 and 4 of that same journey. EHPS was born as an autonomous part of the galaxy of web-sites belonging to the World Wide Web Virtual Library History Central Catalogue moved in 2004 from the University of Kansas to the European University Institute in Florence [15].

It responds, within its selected contents, to the question of how we could link better the process of searching for primary sources in the digital age thanks to History 2.0. services. In this way, EHPS is a complement of the new ways offered by new semantic OPACs for searching inside library holdings and combining internal search with external potential contents. EHPS follows also the economic necessity of adopting low technological profile solutions because few financial resources were available and no specialized ICT staff was available. EHPS is made of a very light open source CMS cyber-infrastructure, easy to build and develop, easy to understand and use. EHPS is a small scale tool, low technological profile database using Dries Buytaert's Drupal open CMS [16], and the Zen theme system maintained by John Albin Wilkins [17].

EHPS has been conceived as a web 2.0 tool aiming to win the participation of its specialized public in order to complete the single web-sites descriptions and abstracts it offers with personal experiences within these archives. It is hoped that EHPS would be considered not only as a redirecting device to other web-sites, but also as a research tool that needs critical judgments from its users for the benefit of other users. This would be also about collaborative tagging and creating more articulated folksonomies completing EHPS keywords.

EHPS portal is a tool for indirectly «publishing» primary sources in the browser from a request of a fully digitized single newspaper like L'Unità in Italy or the Journal de Genève in Switzerland, to broader research topic like the Cold War or the European Integration, to a request for specific types of primary sources like posters or postcards for the history of the UK during the Victorian period, etc.. For postgraduate programs, the need is to discover original and unexploited primary sources or to consult and precise the use of original sources.

## Searching Primary Sources in EHPS

Within EHPS, primary sources are to be retrieved and accessed in two steps. The first step is made of a retrieval of the portal's contents, the second, viewing the primary sources, have to be performed leaving EHPS itself, for the web-site where the sources are directly viewable. Performing a search in EHPS is done using of four different ways [18].

The simplest one is to browse one of the five categories offered in the left-hand column: Country, Language, Period, Subject, Type of source.

The most selective one is to search the portal's combination of tags [19], the list of meta information divided within five different categories: a chronological one from the medieval times to nowadays; a linguistic one dealing with all the European languages in which the primary sources may have been written; a list of countries on which the sources are telling something as single nation's history or on Europe as a whole; a typology of sources trying to define different kind of available digital documents and, finally, a broad list of subjects under which the primary sources may belong to. A third primary source retrieval possibility is offered with a free text search -also in advanced mode- in the whole indexed content of EHPS [20]. This search is useful when names, places, titles of primary sources are already known by who's using EHPS in order to discover them through the descriptive in nEnglish abstracts created for each single web-site indexed.

A fourth way to search for contents is going also outside EHPS itself, using the Google Custom Search [21] for a wide search inside the listed web-sites indexed.

## Interacting with EHPS contents

Being informed about new contents is easy receiving RSS Feeds. Following EHPS directly from Twitter is also possible. Each time a new web-site is added to the portal, whoever follows EHPS receives a tweet with the title of the indexed site and URL to visit it [22].

There's also the possibility to subscribe to become a fan of EHPS in the Facebook and to be informed on all new entries there too [23].

Registered users [24] have the possibility to "vote" for the qualities of the indexed web-site and comment, annotate and complete the abstract and the description of each single web-site with their own information. A registered user is also able to create ones own list of bookmarks and to suggest via a web-form, new web-sites to be included in the portal. Some News are also directly available on the portal itself [25].

Google Analytics is monitoring accesses to the portal. After few month of activity, we may say that some of the first web-sites indexed were already viewed more than 1.000 times each and the portal has been abstracted in Intute [26] and reviewed in H-Soz-u-Kult [27] and is connected within specific widgets in some library and digital humanities web-sites [28].

## Conclusions

What would of course change enormously the importance of the portal [29] would be to become as complete as possible including all primary sources on the history of Europe completing the Multilingual inventory of Cultural Heritage in Europe portal and other European portals which are also included in EHPS [30] so to become the main history portal and reference for multi-lingual contents on the history of Europe.

If the consensus of the community of historians will grow, who knows ? Until now, accesses to the portal have increased exponentially during a year of activity [31].

## Notes

- [1] Marc Bloch: *The historian's craft*, New York, Vintage Books, 1953.
- [2] Serge Noiret: "History in the EUI Library: A Retrospective", in *EUI Review*, Spring 2008, pp.30-31, <[http://www.eui.eu/PUB/EUIReviewPDF/EUIreview\\_spring\\_2008\\_LIGHT.pdf](http://www.eui.eu/PUB/EUIReviewPDF/EUIreview_spring_2008_LIGHT.pdf)>.
- [3] Roy Rosenzweig: "Scarcity or Abundance? Preserving the Past in a Digital Era", in *The American Historical Review*, Vol.108, n.3, June 2003, <<http://www.historycooperative.org/journals/ahr/108.3/rosenzweig.html>>.
- [4] Kirsten Sword in "Interchange: The Promise of Digital History", in *The Journal of American History*, 95.2, September 2008, § 35, <<http://www.journalofamericanhistory.org/issues/952/interchange/index.html>>
- [5] Steven Mintz in *Ibid*, § 195..
- [6] Digital Humanities Luxembourg: Contemporary history in the digital age, [<http://www.digitalhumanities.lu/Pages/default.aspx>].
- [7] CLEO, <<http://cleo.cnrs.fr/>>.
- [8] Revues.org, <<http://www.revues.org/>>
- [9] Noga Arikha and Gloria Origgi: Text-e, <<http://www.text-e.org/index.cfm?switchLang=Eng>>.

- [10] Theodore Zeldin: *An intimate history of humanity.*, London : Vintage, 1998.
- [11] Serge Noiret, "Internet et les internet(s), lettre à Théodore Zeldin, Vendredi 14 Décembre 2001", in *The Future of the Internet: A conversation with Theodore Zeldin*, <[http://www.texte.org/debats/index.cfm?fa=view&type=view&ConfText\\_ID=9&Parent=0&Intervention\\_ID=332&Top\\_ID=332](http://www.texte.org/debats/index.cfm?fa=view&type=view&ConfText_ID=9&Parent=0&Intervention_ID=332&Top_ID=332)>
- [12] Edward L Ayers: «The Academic Culture & The IT Culture: Their Effect on Teaching and Scholarship», in *EDUCAUSE Review*, v.39, n.6, pp.48-62, Nov-Dec 2004, p.55, <<http://net.educause.edu/ir/library/pdf/ERM0462.pdf>>.
- [13] Daniel J. Cohen and Roy Rosenzweig: *Digital history: a guide to gathering, preserving, and presenting the past on the Web.*, Philadelphia: University of Pennsylvania Press, 2005.
- [14] European University Institute, <<http://www.eui.eu>>.
- [15] WWW VL History Central Catalogue, <<http://vlib.iue.it>>.
- [16] Dries Buytaert, <<http://buytaert.net/resume>> and Drupal, <<http://drupal.org/>>.
- [17] Zen, <<http://drupal.org/project/zen>> and John Albin Wilkins, <<http://www.albin.net/>>.
- [18] How to search the portal, <<http://primary-sources.eui.eu/search>>.
- [19] Category browser, <<http://primary-sources.eui.eu/combined-category-search>>.
- [20] Free Text Search, <<http://primary-sources.eui.eu/free-text-search>>.
- [21] Google Custom Search Engine, <<http://www.google.com/cse/>>.
- [22] EHPS Twitter account, <<http://twitter.com/EHPS>>.
- [23] EHPS in Facebook, <<http://www.facebook.com/pages/European-History-Primary-Sources/223099761969>>.
- [24] User account. Create a new account, <<http://primary-sources.eui.eu/user/register>>.
- [25] News Archive, <<http://primary-sources.eui.eu/news/>>.
- [26] Intute, <<http://www.intute.ac.uk/cgi-bin/fullrecord.pl?handle=20090121-13211595#user>>.
- [27] Martin Munke: Web-Rezension zu European History Primary Sources EHPS, in: *H-Soz-u-Kult*, 07.11.2009, <<http://hsozkult.geschichte.hu-berlin.de/rezensionen/id=158&type=rezwww>>.
- [28] Repertorio di risorse web a cura della Biblioteca di Filosofia e storia dell'Università di Pisa, <<http://filosofiaistoria.wordpress.com/>>;
- [29] MICHAEL, <<http://www.michael-culture.org>>.
- [30] Other Portals, <<http://primary-sources.eui.eu/other-portals>>.
- [31] EHPS was officially launched the 8th of June 2009. Alexa Traffic rank for EHPS, <<http://www.alexa.com/siteinfo/http%3A%2F%2Fprimary-sources.eui.eu#trafficstats>>, 8th December 2009.

## **Zinaida MANŽUCH**

### **Digitisation and communication of memory: from theory to practice**

#### **Abstract**

Digitisation provided enormous opportunities for presenting cultural heritage online. Archives, libraries and museums are commonly entitled as memory institutions and digitisation is considered a useful tool for performing their memory function. Currently, there is a tendency of aggregation of heritage content by building international digital portals oriented at the general public. However, many questions about the nature of memory and its communication, benefits gained by general public using the online collections etc. remain unexplored. The aim of this paper is to explain how memory is communicated in archives, libraries and museums basing on the theories of memory and heritage, and to demonstrate by employing the examples of digitisation initiatives how these theories can be applied for developing online heritage services. As a result of theoretical analysis and case studies a conclusion that digital conversion and online access should not be an ultimate goal of the initiative was made. When developing the concept of digitisation project memory institutions should bridge the idea of heritage service with the current context and needs of the user community. Technologies reflecting common memory practices (e.g. web 2.0 photo sharing systems) and providing attractive communication environment have a great potential for developing online heritage services.

**Keywords:** digitisation, memory, memory institutions, online heritage services, web 2.0

#### **Introduction**

The term “memory institution” originates from the recognition of the significance of memory function which is one of the essential pre-requisites for the existence of archives, libraries and museums. Digitisation offered new opportunities to communicate memory; however, it also brought the challenges of discovering meaningful ways of such communication in the changing socio-cultural and technological environment.

The major advantage offered by digitisation is an opportunity for wide access that allows overcoming geographical and time barriers. Inspired by access potential for cultural heritage archives, libraries and museums engaged in digitisation projects that evolved rapidly from small one institution experiments to large-scale initiatives implemented by international consortia. Today there is a time of large-scale online cultural heritage services as the *European Library*, *Europeana*, the *World Digital Library* etc. Majority of these international initiatives pursue the objective of providing rich collections to diverse audiences like general public, scholars, educators and others.

Such concepts as memory, heritage, knowledge are often put together in diverse visionary statements. “Information technologies can enable you to tap into Europe’s collective memory with a click of your mouse”, as stated by the Information Society & Media Commissioner Viviane Redding in her speech [1]. Later on, in the European Commission communication on the progress of digitisation, accessibility and digital preservation of cultural heritage in Europe the metaphor “collective memory with a click of your mouse” transformed into “Europe’s cultural heritage at the click of a mouse” [2].

The visions formulated by politicians and large-scale digitisation consortia raise a lot of questions. How the general public benefit from cultural heritage collections? How digitisation relates to memory? Are cultural heritage and memory communicated in the same way? Is it enough just to put digitised materials online? Current research reveals that many contemporary digitisation initiatives ignore the needs and the nature of establishing links with the past by individuals and communities. Research of online cultural heritage initiatives performed by the American researcher Maria Dalbello has shown that in many cases project ideas were based on cultural heritage collection – its structure and content [3], but not related to the needs and expectations of its potential users. Similarly, the analysis of cultural heritage projects supported by the European Union performed by Zinaida Manžuch revealed that project initiators were more interested in issues of managing cultural heritage resources than constructing meaningful stories of the past [4].

The aim of this paper is to explain how memory is communicated in archives, libraries and museums basing on the theories of memory and heritage, and exhibit how these theories can be used for building cultural heritage services in the digital environment.

#### **Theory: why and how the modern societies remember**

Memory is a way of individuals and societies to deal with the past. The word „re-membering“ means becoming a member again and indicates that memory is a source of social cohesion in human communities [5]. Although by remembering we create the link to the past, it is not re-living it again. It is impossible to re-live again the same emotions or enter the same contexts or events as they were in the past. By remembering the past is



always constructed again according to the present condition, views and needs of those who remember [5]. Individual recollections are influenced by the membership in communities that form the social memory environment. Communities are remarkable for common needs and interests, which become what Maurice Halbwachs called 'les cadres sociaux de la mémoire' [social frameworks of memory]. Individuals "recall" events or experiences that may precede their birth, and these recollections are very similar within the same communities. Social frameworks, in Halbwachs words, are '... precisely the instruments used by the collective memory to construct an image of the past which is in accord, in each epoch, with the predominant thoughts of the society' [6: p. 40].

Cultural heritage may be approached a mnemonic device that connects us to the past. When there are no living witnesses to tell the story of the past, events or experiences transform into remote symbols and rituals that become a part of the identity and history of a particular community. These stories of the past are mediated by cultural heritage [7].

Cultural heritage is a part of the past selected in accordance with contemporary needs of societies [8]. It is loaded with constantly shifting symbolic meanings. The meanings of heritage are different from memory per se, in that they are explicit: they may be disseminated and comprehended by any person. Memory always means "being" or "belonging", while heritage may also mean "knowing about", which is open for rational cognition. Only common life context, values and experiences enable the transformation of heritage into memory. In other cases, heritage may be valued for particular features, used in education, which have nothing in common with memory. Therefore, interpretations of heritage symbolic meanings also allow archives, libraries and museums to communicate knowledge about the past without any reference to the collective memory of communities [4].

In the modern societies remains of the past are transformed into heritage as a result of selection decisions performed by different institutions that usually include research organisations, memory institutions, governmental bodies etc. In archives, libraries and museums, for instance, particular items become cultural heritage as a consequence archival appraisal or well-defined selection processes in libraries and museums. The process relies on expert views on societies, their needs, contemporary context and values. This inevitably creates a gap between heritage experts and citizens. Therefore, cultural heritage does not become a heritage in a full sense until it is acknowledged by the society [9]. When cultural heritage is meaningful to citizens they are able to remember, i.e. construct their images of the past. The major role of memory institutions is to interpret and contextualize cultural heritage for it to become meaningful to people in their present lives.

Communication of memory in archives, libraries and museums is also shaped by the development of new media that introduces novel ways of interaction with cultural heritage collections. The impact of the communication technologies (in a broad sense – language, print, digital media) to the ways a human being understands the surrounding world and shares these meanings with others was widely researched and argued. Contemporary internet technologies (web 2.0 in particular) are remarkable for growing user-centeredness, focus on co-authorship and collaboration, interactivity of services [10]. Although all ideas about computer and web are rooted in older communication technologies, when used they undergo enormous transformations. This argument was illustrated by Christine Borgman reflection on email which was rooted in metaphors of traditional paper post and letters but evolved into an interactive service with newly developed language (i.e. short messages and abbreviations) and elements (e.g. smileys) [11]. In order to meet user requirements memory institutions should adapt to evolving communication technologies, environments and practices.

### **Practice: communicating memory in digitisation initiatives**

The essential feature of memory and heritage is the link between the past and present context of community life. Memory institutions should consider these links while formulating the objectives of digitisation initiatives. In this paper the focus is put on services related to communication of memory; thus excluding all spectrum of services aimed at communication of heritage (e.g. services for scholars, learners, cultural tourists etc.).

Archives, libraries and museums often employ two ways of linking the past with the present. The first way involves orientation at the identity, values and structure of particular community, while the second – particular social issue or phenomenon that creates new communities. The instance of the first way are geographical communities, who mainly associate their identity and personal history with particular place. The instance of geographically oriented digitisation initiative is the project Worthington Memory (USA) undertaken by the Worthington libraries and Worthington Historical Society. The initiative was grounded on the needs of the local community to cultivate social bonds and determine identity of the community [12]. One of the remarkable opportunities offered to the users is so-called Time Machine, the service enabling the user to compare past and present views of particular geographical objects by manually uncovering the 'elderly' image layer placed under the modern photo of the item. The photos showing changes of the city become meaningful for local citizens, even those, who have never seen before how the object looked in the past. Communicating memory

by focusing on a specific social issue/phenomenon is illustrated by the project Moving Here (UK), which was implemented by the consortium of 30 memory institutions and dedicated to exploration of 200 history of immigration to the country [13]. With increased migration flows this issue of the past is very actual to both those who immigrate to UK nowadays and those who face the challenge of multicultural society. Each person, visiting the portal can write a digital story about his/her own migration experiences. Sharing personal migration stories bridges past events with the current experiences and thus makes cultural heritage collections meaningful mnemonic devices for the contemporary migrants.

Examples of adoption of new environments for communicating memory is illustrated by the growing number of heritage initiatives that employ web 2.0 tools. Two collaborative initiatives with Flickr – PictureAustralia (Australia) and Library of Congress pilot project (USA) – are relevant instances of successful web 2.0 application for the development of heritage services. In the first case, the national online historical image service PictureAustralia was enhanced by enabling users to contribute their own photos. The service became available due to the partnership with Flickr that provided attractive and easy-to-use interface for uploading and describing the photos. One of the outcomes of the initiative was the increased participation of user community, who actively engaged in uploading personal photos and re-photographing historical images. The participative service also increased the visibility of PictureAustralia portal and its usage [14]. Another instance of successful web 2.0 uptake is the initiative of the Library of Congress (USA) aimed at representing historical photographs on Flickr [15]. The project exhibited new ways of engaging communities in cultural heritage interpretation, i.e. remembering: “It is particularly gratifying to see Flickr members provide all kinds of connections between the past and the present through discussions of personal histories including memories of farming practices, grandparents’ lives, women’s roles in World War II, and the changing landscape of local neighborhoods” [15: p. 26]. The project evolved into large-scale international consortium Flickr: the commons, which covered cultural heritage institutions from various countries.

## Conclusions

Memory institutions, initiating digitisation projects, should realise that digital conversion and online access are not the ultimate goals of the project, but rather tools for developing heritage services. Understanding of cultural and social mechanisms of remembering should guide the development of service concept. In most cases representative of general public have no definite information need but is willing to reinforce the feeling of belonging to particular community. Examples illustrated that the links between heritage and communities could be established by making references to important geographical coordinates of community life and contemporary social issues that were also faced in the past.

Heritage is not memory; therefore, the definitive feature of a memory institution is not merely holding a cultural heritage collection, but also performing activities that transform heritage into the cultural intermediary of memory. Heritage becomes meaningful mnemonic device when it is related to the present life context of communities and individuals. It is the responsibility of memory institutions to make the links between the past and the present meaningful to the user.

Experiences of current digitisation initiatives have shown that reliance on widespread memory practices and popular communication tools is an effective way of engaging users in recollection and active interaction with cultural heritage collections. Successful experiments with Flickr proved that this web 2.0 tool utilising popular habits to collect photo albums and imitating common actions of making descriptive notes allows developing attractive environment for communicating memory.

## References

- [1] “European Commission steps up efforts to put Europe's memory on the web via a “European Digital Library” (2006), available at: <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/06/253> (accessed at 15 October 2009).
- [2] “Europe's cultural heritage at the click of a mouse: progress on the digitisation and online accessibility of cultural material and digital preservation across the EU”(2008), available at: [http://ec.europa.eu/information\\_society/activities/digital\\_libraries/doc/communications/progress/communication\\_en.pdf](http://ec.europa.eu/information_society/activities/digital_libraries/doc/communications/progress/communication_en.pdf) (accessed at 15 October 2009).
- [3] Dalbello, M. (2004), “Institutional shaping of cultural memory: digital library as environment for textual transmission”*Library Quarterly*, vol. 3, issue 74, p. 265-299.
- [4] Manžuch, Z. (2009), “Archives, libraries and museums as communicators of memory in the European Union projects”, *Information Research*, vol. 14, issue 2, available at: <http://informationr.net/ir/14-2/paper400.html> (accessed 15 October 2009).
- [5] Olick, J. K., Robbins, J. (1998), “Social memory studies: from “collective memory” to the historical sociology of mnemonic practices”, *Annual Review of Sociology*, vol. 24, p. 105-140.



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

- [6] Halbwachs, M. (1992). On collective memory. Chicago, IL: University of Chicago Press.
- [7] Assmann, J. (2004). Kul'turnaya pamiat': pis'mo, pamiat' o proshlom i politicheskaya identichnost' v vysokih kul'turah drevnosti. [Cultural memory: literacy, memory of the past and political identity in the high cultures of antiquity.] Moskva: Yazyki slavianskoy kul'tury.
- [8] Graham, B., Ashworth, G.J., & Tunbridge, J.E. (2004). A geography of heritage: power, culture and economy. London: Arnold.
- [9] Čepaitienė, R. (2005), Laikas ir akmenys: kultūros paveldo sampratos modernioje Lietuvoje. Vilnius: Lietuvos istorijos instituto leidykla.
- [10] Musser, J., O'Reilly, T. (2007), Web 2.0 principles and best practices. Sebastopol, 2007.
- [11] Borgman, C. (2000), From Gutenberg to the global information infrastructure: access to information in the networked world. Cambridge: MIT Press.
- [12] "Worthington memory" (2009), available at: <http://www.worthingtonmemory.org/> (accessed 15 October 2009).
- [13] "Moving Here – 200 years of migration to England" (2009), available at: <http://www.movinghere.org.uk/> (accessed 15 October 2009).
- [14] Hooton, F. (2006), "Democratising history: evaluating PictureAustralia's Flickr pilot project", National Library of Australia Gateways, December, no. 84, available at: <http://www.nla.gov.au/pub/gateways/issues/84/story13.html> (accessed 15 October 2009).
- [15] Springer M. et al. (2008), "For the common good: the Library of Congress Flickr pilot project", available at: [http://www.loc.gov/rr/print/flickr\\_report\\_final.pdf](http://www.loc.gov/rr/print/flickr_report_final.pdf) (accessed 15 October 2009).

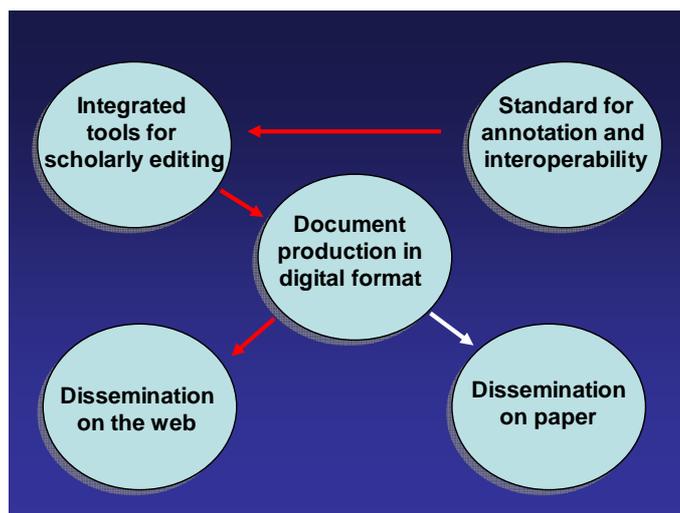
## Andrea BOZZI

### Pinakes Text. A tool to compare, interoperate, distribute and navigate among digital texts

Over the past years text processing systems have become part of the daily language of many scholars working in the field of the Humanities, despite some objections raised against this type of technology which still seems to be distant in terms of simplicity of usage, appropriateness, and flexibility.

Usage requires particular attention as concerns the interface between the information system and the user, while appropriateness and flexibility have not been sufficiently taken into consideration even due to the fact that they almost seem to be in contradiction. Therefore, it is not easy to plan and implement a text processing system which is suitable for specific types of research and at the same time flexible to operate in various sectors of study.

The project Pinakes Text (PKT) that I am presenting here, is aimed at achieving this ambitious target through an architecture based on interconnected modules. In other words, the system works with a nucleus of components for the treatment of both text files and digital image files, which form the core of the system. According to the specific needs, from time to time a number of programs are added both for the management of images (enhancement, segmentation, pattern recognition, etc.) and of text (natural language processing, information extraction, data mining, electronic editing, ecc.). The simplified scheme which lies behind PKT could be represented as follows:



1. The first element is represented by the respect of internationally shared standards, so that the information managed by PKT is interoperable with other data produced in the humanistic field. The standards are also followed when not only primary data (texts, images, etc.), but also secondary information, such as annotations, variants, comments and/or information produced by computational systems (e.g. morphological, syntactic, semantic analyses) are introduced. The software development tools are totally open source in order to avoid any fees for end-user licences.
2. The information system is entirely web-based and the tools for the production or search for information are oriented towards the sector of critical and textual scientific editing. At present, the target of PKT is represented by specialist users. However, the structure of the system also envisions a number of operations, in particular those connected to the phases of search and query, which can be further developed so as to meet the needs of a non-specialist-user.
3. PKT allows to produce on a web server data that have been labelled and annotated in collaborative form, as long as all the members of the same community (e.g. mediaeval philologists, Greek papyrologists, Egyptologists, Latin epigraphists, historians and science philosophers, etc.) agree with the same standards, as evidenced in point 1.
4. Some experiments are in course to check whether PKT meets the needs requested by a community of scholars working on documents of Egyptian archaeology. The documents and annotations are produced in digital format and are classified according to a domain ontology agreed upon by the same members of the community. This semantic-conceptual structure can be replicated not only to classify the



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

documents, but also part of their content. In this way, it is possible to retrieve information both at the level of forms (words, strings of characters, lemmas), and at the level of concepts expressed in the single parts of the texts.

The main form of dissemination envisioned by PKT is the one on-line on a server permanently connected to the internet. The encoding of the data, entirely performed in XML language, also allows distribution in paper format. As concerns this particular aspect, for the next phases of development of the system, the modes of production of the information managed by PKT on e-book will be taken into consideration. The introduction of e-book on the market should provide considerable medium-term increase percentages.

With regard to the current use of PKT:

- it is one of the text and image management components participating in a COST Action of the European Science Foundation (<http://www.interedition.eu/>);
- it has been considered the suitable technological basis for the project ERC (Advanced Grant) “Greek into Arabic”, approved and financed in the first days of November 2009 by the International SH5 Panel of ERC;
- it manages the corpus of the National Edition of the works of Galileo Galilei (<http://pinakes.imss.fi.it:8080/pinakestext/home.jsf>);
- two European project proposals are underway which in case of success could consolidate the position of PKT as infrastructure of research in the sector of sciences of the text.

Implementation of the system is the result of a collaboration between the Institute for Computational Linguistics “A. Zampolli” of the National Research Council of Pisa, the “Fondazione Rinascimento Digitale” and the Institute and Museum of the History of Science in Florence.



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## **Frank AMBROSIO, William GARR, Edward MALONEY, and Theresa SCHLAFLY**

### **MyDante and Ellipsis: defining the user's role in a virtual reading community**

#### **Abstract**

This paper explores the pedagogical concepts and technological framework underlying the MyDante Project (<http://dante.georgetown.edu>), developed by the Center for New Designs in Learning and Scholarship at Georgetown University. We argue that MyDante, essentially a Web-based, interactive hypertext edition of Dante's *Divina Commedia*, represents a distinctive approach to designing spaces in which readers interact with texts. Always acting on the principle that the technological design should serve the pedagogical goals of the project, we have created an environment to facilitate deep engagement with Dante's poem within the context of a particular method of study known as contemplative reading. MyDante encourages the reader to experience the poem in a way that is profoundly personal, while at the same time enabling a collaborative experience of the shared journey by a community of readers, and creates a record of that journey.

After exploring the interplay between the pedagogical aims and technological structure of MyDante, we briefly describe the current development of Ellipsis, a new iteration of the application which will be customizable for a wide range of texts, disciplines, and pedagogical approaches. Finally, we provide some examples of how students experience MyDante, and conclude with further reflection on the distinctiveness of MyDante's approach.

**Keywords:** MyDante, Digital Humanities, pedagogy, technology, contemplation

#### **Introduction**

If we think of any reading experience as a meeting of a text and its context with a reader and his or her context, then we might say that those of us in the Digital Humanities have tended to focus much of our energy on understanding, marking up, and making available the first half of this equation. That is, the field of Digital Humanities writ large has often been defined by textual projects that emphasize the importance of the texts and their contexts (alternative versions, supplemental texts, historical documents, and so on). The past twenty-plus years of research projects within the Digital Humanities have made available to the humanities scholar a wide range of historical, literary, and cultural texts, marked up for quick searching, complex linguistic analysis, and archiving. We believe that, in many respects, the MyDante project (<http://dante.georgetown.edu>) represents an alternative approach to the Digital Humanities, to our focus on the text and its primacy, and to our understanding of the relationship between the reader and the text.

The MyDante project began a little over ten years ago with the primary goal of providing students with a contemplative space in which to engage with Dante Alighieri's *Divina Commedia*, specifically within the context of the undergraduate philosophy class "Dante and the Christian Imagination" at Georgetown University. Many of the previously available online versions of the *Commedia* focus on markup, searching, and archiving the text. From the beginning, the MyDante project was designed to enable students to understand the text through their interaction with it, their reflection on it, and their engagement with their peers around it. Inspired by the metaphor of the medieval illuminated manuscript, we aimed to allow students to see the text of Dante's poem as a palimpsest, as a place where their ideas and their writing share the same space as the poem; where they could engage with and rethink the poem just as a monk in the Middle Ages might have done through marginalia and illuminations. We created and continue to develop a variety of tools, such as an annotation tool, a journaling tool, and a multimedia editor, to encourage students to interact with the poem and share their ideas with others, much in the way the marginalia of a medieval manuscript would influence future readers. MyDante simultaneously encourages deeply personal reflection as well as scholarly collaboration focused on the text.

Dante's poem is particularly well suited to this type of contemplative experience, but we see this process being extended to virtually any text. MyDante gives us a powerful model that can be localized to the classroom or extended to larger communities interested in reading interactively and reflectively. To facilitate this type of reading experience within a wide range of contexts, we have created a digital publishing platform we call Ellipsis (see Section 2). In many ways, the types of tools we attempted to create ten years ago prefigured much of what has become standard practice on the Web as we've moved from a read-only Web to a read/write Web. Social networking tools such as Diigo, YouTube, and Wikipedia encourage all Web users to see the Internet as a shared text to be viewed reflectively and even at times contemplatively.

## The Pedagogy Underlying MyDante

The goal of the MyDante project is to make available a Web-based, interactive hypertext edition of Dante's epic poem that allows each reader to develop an illustrated, annotated, personalized copy of the text over the course of a lifetime. More than just a technology, MyDante offers a pedagogy of reading and reflection designed to illuminate and document the reader's experience of the poem. Dante wrote the *Commedia* as an invitation to undertake a journey of self-discovery. MyDante serves as a permanent record of accepting that invitation and sharing Dante's journey. It makes the poem profoundly personal, while at the same time enabling a collaborative experience of the shared journey by a community of readers.

A fundamental principle of the project from the beginning has been that technology by itself, no matter how engaging, is only an instrument and must be given a humane purpose. The pedagogy that informs the MyDante project, inspired in part by Dante's "Letter to Can Grande della Scala," is essentially the method of allegorical interpretation developed by the Jewish and Christian traditions to read the Scripture faithfully and which was perfected by the contemplative culture of medieval monasticism. All aspects of the medieval monastic culture were pedagogically integrated around its central purpose and activity: contemplation. The monks developed practices of active engagement with Scriptural texts, from meticulous manual copying to artistic illustration and scholarly commentary. From the manuscripts the monks produced, we can clearly discern how closely connected technology, action, and contemplation are in the fabric of a human way of life.

What must a reader do to experience Dante's poem contemplatively? How does reading as a contemplative practice work? The first requirement is to recognize that, just as Dante told his patron Can Grande della Scala, there are multiple levels of meaning simultaneously at work in the poem, each of which necessitates a different kind of understanding; the second requirement is to learn how to move progressively from:

- The literal, narrative level of the story of Dante the pilgrim's journey from the Dark Wood to the Final Vision. The goal of reading at this first level is clear comprehension of the characters and plot. From here to
- The ironic and metaphoric level communicated by the artistic choices Dante the poet makes regarding characters, episodes, images, and themes. The goal is to arrive at an interpretation of the poet's message to the reader; that is, to explore the questions, Who is Dante the poet? What is he trying so hard to tell me? Then, from this level to go to
- The reflective level, contained in the reader's personal responses to the poet's confession, witness, and testimony, in the form of a dialogue between poet and reader. The goal of this level of understanding is personal reflection. The meaning of the poem is not finally understood until reader and poet find themselves standing face to face, in the presence of all others who confront the same questions of personal identity, freedom, and responsibility. To do this requires the reader to imagine how one's own journey is the same as the poet's, how both are "universal," the same for all persons, despite every difference of time, place, and culture. To understand the poem, then, is to become part of the poem by recognizing oneself in it and by making it genuinely one's own by responding to the question "Who am I?" - not simply as an individual, but as a person who is both the same as and singularly different from every other person in such a way that, as the poem tells us from the beginning, the story it recounts is truly the story of "our life."

Readers can move among these three levels of understanding by using MyDante's tools, which include:

- Side by side Italian and English texts and an Italian audio recording of the poem
- General introductions to each of the *Cantiche*, including some in video format
- An Image Gallery, containing illustrations of the *Commedia* and a wide range of images with thematic connections; users can also upload their own images.
- An annotation tool, mirroring the function of manuscript marginalia, allowing the reader to comment on specific lines and read others' comments
- A journaling tool that enables readers to embed extended reflections into MyDante, creating personal records of their relationship with the poem
- A Biblioteca, which houses digital copies of other works by Dante and relevant texts by other authors, as well as chronologies, maps, and other resources
- A Chapter Room that includes a space for interactive discussion

Given the broad pedagogical goals outlined in the three levels of reading above, in what follows we shall focus on a specific question we faced in developing MyDante: to what extent can digital media effectively guide readers in comprehending, interpreting, and reflectively appropriating the significance of texts, either as an enhancement of the role of human teacher or, in certain cases, as an alternative to direct contact with a personal guide?

First, one clarification: this question is not properly subsumed under the heading of "distance learning," or delivering content and evaluating student mastery of materials or skills online. In the context of



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

Ellipsis/MyDante, the scenarios we envision are either the enhancement of live teacher effectiveness through the transfer of certain components from the classroom to students' independent work outside class - so as to free classroom time for different and arguably more advanced learning activities - or, on the other hand, the creation of a community of readers who are not enrolled in any formal academic program and who are not receiving direct personalized guidance or evaluation from an instructor.

In other words, we asked ourselves how much learning at the level of direct engagement with literary texts could be achieved via a pedagogy that subsists solely within the structure of a digital platform, and to what extent this platform could open up new possibilities for further learning. The challenge we faced was how to deliver digitally not just a broad array of diverse content material, however well organized and intuitively accessible, but to deliver such content in the context of an artfully designed pedagogy that provides the student/reader with skilled guidance that would otherwise be either absent or less efficiently and effectively delivered in person.

One of the most challenging questions we addressed was how to activate and deliver the pedagogical design that would guide students through the threefold dynamic of comprehension, interpretation, and reflective appropriation outlined above. The solution was to develop a Guide function that could overlay the text and be turned on or off by the reader. More than a commentary or scholarly investigation, the Guide would directly engage students in a series of activities, such as listening to the poem in Italian, dwelling on the details of an illustrative work of art, reading commentary on the metaphoric range of a particular symbol or theme, writing a journal entry reflecting on the larger human significance, for example, of Dante's placement of Ulysses in Hell—activities that would both model learning behaviors for students and stimulate them to repeat and develop those behaviors for themselves.

Both technically and pedagogically, developing an effective Guide format proved even more of a challenge than the team had anticipated and required various stages of experimentation, implementation, evaluation, and revision over the course of two years and two iterations of the class "Dante and the Christian Imagination" at Georgetown. The highest hurdle was how to get students started using the site and to become quickly at home with MyDante's pedagogical tools and practices. The solution we are now testing is an introductory "Getting Started" Guide, designed to supplement class lectures and demonstrations. By combining video segments, audio commentary, screen-capture tutorials, thematic commentary, and a reflection on contemplative reading, we believe we have taken a substantial step toward giving students access to not only a rich body of content but also a progressive, self-regulating pedagogy with which to approach this vast literary monument of human culture and even to draw on for future study of other texts.

## Designing Technology to Support Pedagogy

In order to serve the pedagogy outlined above, the MyDante application needed to enable and encourage certain types of activities by the users. First, the site had to direct users inward, toward a deepening contemplation of the primary text, even as it facilitated comparisons among a broadening scope of related texts (in various media - documents, images, sounds, videos, and even animations). Second, MyDante needed to engage readers with one another as a community of scholars, encouraging them to explore one another's emerging perspectives as a way of continually rediscovering their own interpretations of the poem's significance.

Consideration of the philosophy behind the pedagogy played no small role in developing the application's functional design. Rather than focusing on creating a sophisticated markup to encode a particular domain of inquiry around the poem, or on providing an entirely open-ended discussion forum, we chose to integrate several simple actions, which we felt best supported the invitation to students to deepen their personal involvement with the text, within the context of a coherent pedagogical framework. This choice reflects a conviction that readers are ultimately the repository of the deepest meanings in a text. It is the interpersonal connections among a community of readers, moved by a text in similar and different ways, that ensure its survival as a work of art.

The Guide function described above, although by no means the only tool designed to invite students into a deeper reading of the text, allows us to trace how the application supported each level of reading. The Guides address the first level of comprehension by providing contextual historical and literary information in the form of brief annotations similar to footnotes. At the second (metaphoric) level of reading, Guide materials interspersed throughout the poem encourage specific readings of particular themes and symbols, and point out resonances among different sections of the text. Finally, at the third (reflective) level, the Guide asks open-ended questions which prompt directed reflection by the students on how the poem might relate to their own lives.

Our intention was to provide a self-contained digital cloister within the chaos of choices available on the Web. We discovered that by creating ways to link parts of documents with one another, documents that comprise a system of meanings according to a clear pedagogy, we were able to channel the energy of browsing into a



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

directed activity of probing. We realized that this capacity for focusing a reading community's activity around and concentrically toward the meanings of a text would be very powerful in a wide variety of educational contexts. To that end, we have abstracted the functionality developed for MyDante into an application framework we call Ellipsis. Ellipsis promises to provide, for a wide variety of texts in a wide variety of media, a sharable space for study and reflection. The basic tasks of making a statement about a text or some selection within it, and relating it to some other text (or selection of text), whether from primary materials or those generated within the community, are abstracted to apply similarly to images, sounds, video and other media. In addition, collaboration between faculty is encouraged by allowing multiple overlapping spaces, where community members can easily move into and out of collaborative exercises with members of other communities centered around the same system of related texts.

### **Evidence of Student Learning**

Throughout the project, we have collected feedback and suggestions from students through surveys and interviews, and we have gathered evidence of student learning from their work on the site. Student response to MyDante has generally been extremely positive. For example, one student explained that "it added a different dimension to the class that there was so much that you could experience outside of it. It was a very holistic learning experience, and it added a whole new level of discussion - it was like having twice as much class time." Another student finds himself drawing on MyDante's approach long after completing the course: "MyDante's comprehensive resources and scholarship will always be something I turn to, and the site is an ideal model for both how to read such a vast work and how to contemplate its richness. . . . My experience with MyDante is invaluable in continuing to interpret the poem's meaning and in forming strategies for evaluating other texts."

After we implemented the Getting Started Guide in the fall of 2009, it became clear that the students were demonstrating a better understanding of the material than students had at analogous stages of earlier semesters. They had digested and integrated the various components of the Guide and were relatively comfortable with both the technological and theoretical practices involved in using the site effectively. This deftness was evidenced both by their comments in class and their contributions to the site, in the form of discussion board posts, annotations, and journal entries, in which they applied the concept of the three levels of reading to their interpretations of the text. The students also showed a high degree of engagement with one another on the site and in class. Many of them spoke in class in addition to contributing to the site, and many of them responded to their classmates' posts on the site. For example, after only four class sessions, a student's 1250-word post entitled "Why the Catholic Cosmos?" elicited more than 1700 words of comments by five classmates.

As the site evolves, we are continuing to evaluate student responses to various features of the site and to gather evidence of student learning. For examples of student work demonstrating particular pedagogical goals, and for more testimonials from students about their experiences with the site, please see [http://dante.georgetown.edu/student\\_learning](http://dante.georgetown.edu/student_learning) .

### **Conclusion**

We see the MyDante project as representing a genuinely new approach to designing a space for interaction with texts. This approach enables a distinctively individual reading experience characterized by depth, richness, and intensity, an experience which is further enhanced by the collaborative dimension of reading within a virtual community. MyDante memorializes the reader's journey in a way that is simultaneously imaginative and technological in character, so that each reader's record acts as a personalized archive of meaning. As we continue to develop MyDante and Ellipsis, we remain committed to the principle that technology is subordinate to pedagogy in our aim to transform and enrich the ways in which readers engage with texts of all kinds.

106 Love led the two of us unto one death.  
Caina waits for him who took our life."  
These words were borne across from them to us.

Amor condusse noi ad una morte:  
Caina attende chi a vita ci spense».  
Queste parole da lor ci fuor porte.

Notice the triple repetition of the word "Amor" (the god Amor of the Roman tradition) is a direct reference to the concept of the captive soul and loving heart).

109 When I had listened to the high, clear voice of the poet asked of me:  
I bent my head and hid my face and wept.

112 When I replied, my words were low and sad:  
how many gentle thoughts and longings  
had led them to the a

Clearly Dante is recalling the story of Cain from the Vita Nuova. But he makes a subtle reference to the story of the blind men and an elephant. Perhaps this exposes a difference between the two guides or to himself.

115 Then I addressed my words to the other, and I began: "Francesca,  
how can you be so calm, when you see  
me move me to tears of sorrow and of pity.

Ma dimmi: al tempo d'i dolci sospiri,  
a che e come concedette Amore  
che conosceste i dubbiosi disiri?».

118 But tell me, in the time of gentle sighs,  
with what and in what way did Love allow you  
to recognize your still uncertain longings?"



La Porte de l'Enfer (The Gates of Hell), by August Rodin

CLOSE X

Screenshot of MyDante (<http://dante.georgetown.edu>) showing the image viewer tool. Note the guide sections interspersed within the poem text.



## **Wendy M. DUFF, Jennifer CARTER, Costis DALLAS, Lynne HOWARTH, Seamus ROSS, Rebecka SHEFFIELD, and Cassandra TILSON**

### **The museum environment in transition: the impact of technology on museum work**

#### **Abstract**

This paper reports on a study of the impact of, and challenges posed by information on the museum, and the changing nature of museum work. The study involved semi-structured telephone interviews with sixteen senior North American museum professionals and academics teaching in museum studies programs. Our findings suggest that the ways museums interact with their publics and the areas of dissemination and collections management are profoundly changing. It found that the three most common challenges that museums face include: the cost of designing, implementing, and maintaining technology; a lack of in-house expertise; and information management. The study also indicates that the museum profession is facing a generational shift and that younger professionals perceive technology as a ubiquitous part of their environment.

Keywords: Museums, information technology, museum workers

#### **Introduction**

Many suggest that museums have been transformed by their societal context and the proliferation of information technology (IT) in our contemporary moment. Parry argues that digitality “helped to support a realignment of museography that was taking place, from object-centred to experience-centred design.” [1] Bearman and Gerber posit that new technologies have fundamentally changed the ways museums communicate. They state that, “since the late 1980s, computer-based interactive programs have delivered more varied and exciting information on the museum floor than traditional mechanical interactives or static signage. Today, a museum without a collections database and a Web presence is hardly considered professional.” [2] However, they go on to note “not all institutions are using online access equally well.” [2] The will of museums to participate may not be sufficient impetus and Loran found in a study of British National Museums that a favourable ‘political context’ energized museum uses of web services to reach broader audiences. [5] The 2002 DigiCULT Report analyzed the current state of technological deployment in the cultural heritage and indicated how a range of technologies could be used to unlock the potential of such cultural heritage institutions as museums [3].

There is a large and growing literature evaluating the intersection of new technologies and museums by writers such as Paul Marty, Ross Parry, Sarah Kenderdine, Fiona Cameron, and Katherine Burton Jones. This literature addresses issues ranging from museum informatics [as Paul Marty defines the field, “the study of the sociotechnical interactions that take place at the intersection of people, information, and technology in museums” [6]], to resource development (such as imaging, digitization, and integrated information systems), from new media technologies and critical digital theory (pertaining to object morphologies, virtual systems, digital objects, communication technologies) to visitor interaction and online technology. The diversity and profusion of this literature suggest museum professionals and funding agencies need a better understanding of the challenges information technologies pose for museums, and the ways museums are changes to meet these challenges. The technology observatory research of The DigiCULT Forum is indicative of the needs of the community for access to knowledge about emerging technologies [8] and how these needs might be met.

#### **Methodology**

To gain this understanding, we undertook a research project, funded by the Canadian Heritage Information Network (CHIN), involving semi-structured telephone interviews. We devised a series of questions, grouped into 4 thematic sections: i. new technologies used in the participant’s own institution; ii. new technologies and the museum in general; iii. IT skills and training in museums; and iv. information technology and museums in the future. Telephone interviews took place between March 24 and April 9, 2009 and were conducted in either English or French, as appropriate. In consultation with CHIN, we identified twenty potential senior North American museum professionals and academics, whom CHIN contacted and invited to take part in the study. In the end due to time constraints on the part of potential interviewees we interviewed only sixteen individuals. The participants included three directors of collection management departments, three museum directors, three directors/chiefs of museum technology departments, three academics teaching in museums programs, one director of a museum education department, one project manager, one curator and one head of a museum standards program. We sent documentation to each participant in advance of their interview including a description of the project, a consent form, and an interview script. The researchers conducted the 30 to 75 minutes interviews, and digitally recorded all but one of the interviews. Each researcher took notes during her

interviews, and subsequently replayed the recordings to gather additional comments and ideas. After completing all interviews, the research group met and reviewed the interview notes question-by-question in order to share findings. The research team identified recurring themes represented in the interview data and noted differing points of views. The research team then identified examples that illustrated these themes.

## The Findings

### The Impact of New Technologies on Museums

Interviews revealed that there is no consensus on the extent to which new technologies have impacted museums. One interviewee argued that the museum has not been transformed by new technologies any more than society in general. That is, technology has not altered the traditional core missions of museums to collect, preserve, interpret, and make available cultural heritage. Nevertheless, it has allowed museums to express their missions to a broader audience, and to fulfill them in a variety of new ways. Another interviewee suggested that, although technology has not changed museums' rapport with the public, it has changed the way museums work. Others suggest that technology has changed the way museums think about themselves. For example, one interviewee suggested that information technology has changed the way her museum views itself. She noted that the days of circulating collections are over but the museum can now provide web seminars to teachers in remote communities or other parts of the province. Though some may view these changes as threatening, it is also liberating. Museums now help people make meaning from art and objects in entirely new ways. According to this participant, the museum has become more of a hub and less of a physical resource on which people draw. The development of Web 2.0 technology has also provided museum professionals with new ways of thinking about connecting with their publics online and involving these publics in the museum environment. An interviewee explained that it is "no longer about the visitor in the life of the museum but the museum in the life of the visitor," meaning that, with the interactive Web and new levels of connectivity, museum visitors access the museum before they visit it physically, and continue to visit the museum website after they return home.

Some interviewees urged new museum professionals to rethink the very role of museums, and to consider new technological projects, not just as tools, but also as museums in themselves. Online exhibits, for example, serve as museum spaces; websites are no longer an extension of the museum, but represent the museum as a whole, particularly to those visitors who cannot physically visit the institution.

Most interviewees agreed that museums have been transformed by the proliferation of information technology in our contemporary moment. Within exhibition halls, multi-media installations have provided the most visible manifestation of technology within contemporary museum practices. This aspect has profound implications for the manner that museums communicate with members of the public, and has fundamentally changed the ways that exhibitions are undertaken. The use of cell phones and digital cameras in the exhibition space, for example, is becoming increasingly prolific. Some museums have responded to these changes by easing up on limitations placed on these devices and even encourage the public to use their cell phones in order to access interpretive guides. One interviewee also noted changes in the use of kiosks. He suggested museums used to relegate kiosks to the corner of most museum spaces, but now technology is considered more holistically at the inception of exhibition planning. However, one interviewee pointed out that exhibition teams should include members of the IT department so these teams have adequate IT expertise at all stages of the development of new exhibitions. Museum professionals need to know how to work with kiosks and build new applications for their use; as well as when to use them and when more traditional approaches will be more effective.

An interviewee working in a collection management department suggested the greatest impact of technology on the contemporary museum environment has been in the areas of dissemination and collections management. Another interviewee indicated that almost all museum departments use a collection management system, and a variety of museum professionals use the system and contribute information about objects. The collection management system is the backbone of the museum according to this interviewee. Today far more museum personnel (curators and project managers among them) are familiar with, and have access to, institutional collections because museums have begun digitizing their collections and making them available on-line since the 1990s. Today, researchers may use multiple collections databases and other information technologies to "go shopping" for artifacts. Browsing the collection has been made a feasible reality. According to one interviewee, expertise in creating and working with 3D imaging is becoming increasingly important for some museum professionals. He went on to explain that 3D authoring tools, such as a 3D camera, allow museum professionals to create a three-dimensional image of specimens based on data models. Not only does this permit researchers to access collections at a distance, but it also reduces the handling of the physical specimens by providing visitors with highly flexible digital representations. Imaging technology can increase access to the collection, while helping to preserve the original object.

One interviewee emphasized that museums are about information. In his opinion, the collection remains at the core of information creation and knowledge production, and technology should be seen as a tool or interface to facilitate access to the collection for these purposes. This particular view of technology as a tool was also reinforced by another prominent interviewee who stressed the necessity for museum professionals to both remain curious and creative in their uses of technology, and to develop a solid understanding of the subject specializations of their respective museums, be it in the fine arts, material culture, history, or other. Thus museum professionals must understand and appreciate the capacity of technological tools and the value of these tools for the museum sector. An interviewee pointed out that a collection is compromised if the link between specimens and their information is broken. Therefore, collection managers must have fundamental knowledge of information management techniques and the management and use of databases. He also suggested that the management of legacy data whether still in analogue form or in antiquated collection management systems presents special challenges.

### Challenges Imposed on Museums by New Technologies

The interviewees also discussed the challenges posted by new technologies and identified the three most common challenges: (1) the cost of designing, implementing, and maintaining technology; (2) a lack of in-house expertise; and (3) information management. Many interviewees noted that the collection resides at the core of any museum's work; hence, many museums have committed to building databases or adopting off-the-shelf collection management systems, and/or strengthening metadata to facilitate access to their collections amongst researchers, the public, and other museums. Data-sharing also requires that metadata meets a specific standard of interoperability so that it can be easily migrated from one IT infrastructure to the next. Cleaning data and adding to metadata are, nevertheless, tedious and time-consuming tasks. There is no uniformly adopted shared vocabulary among museums for describing collections, nor is there a consistently deployed standard form of metadata used across all collections even within a single institution. Migrating legacy data to new formats takes time and digital asset production continues at a prodigious rate. Unfortunately, museums are not able to cope with the amount of work necessary, and its expense. Without capturing information about the physical collection, however, museums will not be able to fully participate in the possibilities offered by Web 2.0 technologies such as interoperability between diverse cross-institutional data resources.

Museums are undergoing a conceptual shift that many believe has been triggered by the advent of networked computer technology. Traditionally, collections have fallen under the exclusive domain of the curator; however, as previously noted, information technology allows each collection to be shared, at least as parcels of metadata and digital images travelling across the Internet. While the mission of the museum will continue to be the preservation of its collection, opening up access has changed the focus of the museum from its preservation function to its interaction and engagement with the public. This shift in attention has a number of ramifications for museums themselves. First, the public is not always comprised of savvy technology users. From young people to the elderly, much of the general population suffers from poor computer and information technology literacy skills. For example, few people make a distinction between the deep Web and the information retrieved by a Google search. In addition, museum educators often work with primary and secondary schools, which are chronically under-funded, and, to date, often do not have access to high-speed Internet connections or software packages needed to operate Web 2.0 technologies. Museums must, therefore, strike a balance between low-tech and high-tech services to the public. Second, museums are perceived as authorities on the collections that they preserve. Interviewees point out that bad data or poor information made available on the Web reflects poorly on the subject expertise of museum professionals. At the same time, other interviewees noted that the public demands access to data that museums simply have not had the time or money to properly review or types of data that they do not have the resources to capture. One interviewee notes that most users do not ask for 100% accuracy, and if we wait until all information is 100% accurate and available museums will never upload their material to the web. Finally, the pressure to serve the public has led many museums to pursue IT trends without understanding the costs, benefits, opportunities, and risks associated with these new technologies.

The proliferation of IT has resulted in increased rates of format obsolescence and, as a result, museums now face considerable challenges when attempting to preserve digital assets. One interviewee commented that digital photographs are sometimes mistreated by museums because they are perceived as ephemeral objects, saved en masse onto CD-ROMs and other media storage devices, which are sitting haphazardly on shelves. This is quite unlike "tangible" film negatives and analogue photographs that are being carefully preserved in the vaults. The transition from tangible to ephemeral objects is also forcing museum professionals to re-consider basic concepts of traditional museology. For example, if a digital image can be copied easily and effectively, and disseminated widely via the World Wide Web, which version of this image is the authentic record? Furthermore, how does this scenario impact the "authentic museum experience" for the public?



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

New media art is one area requiring particular attention, especially with respect to how the cycles of life of ephemeral objects are recorded. Organizations like the Fondation Daniel Langlois pour les arts, la science et la technologie in Montréal, whose fundamental mission is research and preservation, has partnered with many museums to further this developing area. Programs such as the Variable Media Network, undertaken by an alliance of universities and museums from 2004-2009, are the products of such work.

One interviewee stressed the importance of preserving the pertinence and authenticity of the museum experience, and this in light of a changing landscape of new technologies. Museum professionals must think carefully about their use of technology, recalling at all times the mandate of the institution and the needs of the institution's visitors. This, she observed, requires a great deal of maturity on the part of the institution. She cautioned that the content of the museum should always be considered the foundation, while the technological tools are always only the gateway.

A further interviewee spoke of yet another challenge to the museum posed by the use of technology when evaluating the validity of user-added content to museum information. Museums, he argued, place a seal of quality on information that they produce, however when members of the public are invited to add images/tags/texts to this information in on-line environments, it becomes increasingly important to differentiate between voices, and to identify who has added what. With the growing developments of cybermuseumology and exhibitions developed for the web, he argues that we must distinguish between institutional and public additions to content so as to avoid the pervasive dilemma facing the wider web: large content, but questionable validity.

Perhaps the most important challenge, however, is convincing management to understand both the benefits and limitations of technology within a dynamic working environment with conflicting priorities. Information technology is not a core function of the museum, and therefore it does not always receive adequate funding. Few museums can exist today without a level of technology and technological support. Acquiring expertise that understands both the museum context and new technologies, whether in-house or outsourced, is key to building a solid information management program and developing innovative and interesting content for the public.

### **Museum Professionals**

As one interviewee noted, the museum profession is facing a generational shift. As older professionals retire and new professionals take their places, the museum environment is undergoing significant changes. He suggested that within the culture of museums, older generations do not tend to prioritize technology to the same extent as newer museum professionals. This he expresses this not as criticism, but merely as a reality of contemporary practices. An outcome of such practices is evident in the vision that younger generations have of the museum: their approach is broader ("transversale"), revealing a greater understanding of the many functions of the museum as a collective. This is especially true in small institutions, where professionals are encouraged to undertake many different roles.

Interviewees suggested that Generation Y tends to perceive technology as a ubiquitous part of the professional environment, not as an add-on, but rather, as inherent in day-to-day work. The challenge is for new Generation Y professionals to stress the importance of new technologies and, more specifically, social networking and other collaborative technologies, to older management. The displacement of older professionals by a younger generation is also evident to others; one interviewee mentioned that there has been a noticeable decline in the number of museum professionals seeking training in specific technologies, indicating that professionals are either entering the field already trained, or have become accustomed to self-directed learning. Several interviewees, however, warned that it is dangerous to assume that all Generation Y professionals will come to the field well versed in new technologies. While it might be true that this cohort of professionals has been more exposed to digital technologies during their schooling both formal and informal (e.g. gaming), this should not be equated with having the ability to critically evaluate, select or use new technologies.

Another change that was noted by interviewees was the shift to project-based business and collaborative work. According to one interviewee, museum professionals are entering the field more accustomed to working in groups, drawing on the expertise of many to perform complicated tasks. As a result, museums must respond to this conceptual shift in work organization to ensure that new professionals can perform in ways with which they are familiar. Institutions, for example, might introduce Intranets to facilitate information-sharing, or chat programs to allow for geographically dispersed real-time communication. Social networking tools are becoming more commonplace within the museum workplace, as more professionals are accepting these technologies as solutions for collaborative project-based work.

### **Conclusions**

The interviews with sixteen senior North American museum professionals underscored the vital role that information technologies have played in the transformation of their institutions over the past two decades

studies in Europe [e.g. 3, 7] have produced similar conclusions. While the core mission of museums may have changed very little, the activities and operational requirements associated with expanding collections from primarily physical, to increasingly digital objects has necessitated the marshaling of customized resources and a new set of knowledge and skills. Whether collecting, curating, educating, programming, marketing, communicating, or fund-raising, all require, not only some basic engagement with associated hardware and software applications, but also a more critical understanding of the inherent strengths, opportunities, and deficiencies of information technologies to support the present and future direction of the institution. Addressing this level of required IT literacy, as well as a perceived generation gap in engagement with new technologies and tools, such as mobile devices, imaging systems, and Web 2.0 and social networking applications, demands timely, focused, ongoing training at all stages of an individual's career within museums. This is a conclusion that chimes with the results of other international studies [7]. The threshold for participation in the digital environment continues to rise and this has implications for educational needs of museum professionals. As the 2004 *The Future Digital Heritage Space: An Expedition Report*, made evident the diversity of mechanisms from the technologies that support the intelligent ambient landscape to those that underpin multimodal interaction such as virtual and augmented reality will increasingly change the relationship between the museum and its audiences [4]. For this to happen continuous educational learning opportunities must become a core part of the life of museums and museums professionals.

### Acknowledgement

The authors would like to thank Anne-Marie Millner, Irene van.Bavel, Madeleine Lafaille and the Canadian Heritage Information Network (CHIN) for their help with this research. CHIN provided substantial funding for this research through Research Contract. No. 45247369

### References

- [1] Parry, Ross (2007). *Recoding the Museum: Cultural Heritage and the Technologies of Change*. London: Routledge.
- [2] Bearman, David and Kati Geber (2008). "Transforming Cultural Heritage Institutions through New Media," in *Museum Management and Curatorship*, 23(4): 385-399.
- [3] Geser, Guntram and Mulrenin, Andrea. (2002). *The DigiCULT Report: Technological Landscapes for tomorrow's cultural economy Unlocking the value of cultural heritage*. European Commission: Directorate-General for the Information Society  
[http://www.digicult.info/pages/report2002/dc\\_fullreport\\_230602\\_screen.pdf](http://www.digicult.info/pages/report2002/dc_fullreport_230602_screen.pdf)
- [4] Geser, Guntram and Pereira, John (2004), *The Future Digital Heritage Space An Expedition Report*, (Salzburg: DigiCULT, [http://www.digicult.info/downloads/dc\\_thematic\\_issue7.pdf](http://www.digicult.info/downloads/dc_thematic_issue7.pdf))
- [5] Loran, Margarida (2005). "Use of Websites to Increase Access and Develop Audiences in Museums: Experiences in British National Museums". In: CARRERAS, Cèsar (coord.). "ICT and Heritage" [online dossier]. In: *Digithum*. No. 7 <http://www.uoc.edu/digithum/7/dt/eng/loran.pdf>.
- [6] Marty, Paul (2008). "Information Representation: Representing Museum Knowledge," in *Museum Informatics*. Paul F. Marty and Katherine Burton Jones, eds. New York: Routledge, pp. 29-34.
- [7] Ross, Seamus (2001), "Budgetary Suicide at the Altar of ICT", (London, Heritage Lottery Fund,) [http://www.hlf.org.uk/nr/rdonlyres/50742b80-89d2-4e6e-8a71-2469a3bcda6c/0/needs\\_ict.pdf](http://www.hlf.org.uk/nr/rdonlyres/50742b80-89d2-4e6e-8a71-2469a3bcda6c/0/needs_ict.pdf)
- [8] For example see. Ross, Seamus, Donnelly, Martin, Dobrova, Milena, Abbott, Daisy, McHugh, Andrew and Rusbridge, Adam, (2005), *Core Technologies for the Cultural and Scientific Heritage Sector, DigiCULT Technology Watch Report 3*, European Commission, 296 pages. ISBN 92-894-5277-3 <http://www.digicult.info/downloads/TWR3-highres.pdf>. For other examples see <http://www.digicult.info>



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## ***Tomi KAUPPINEN, Tuukka RUOTSALO, Frédéric WEIS, Sylvain ROCHE, Marco BERNI, Eetu MÄKELÄ, Nima DOKOOHAKI, and Eero HYVÖNEN***

### **SmartMuseum knowledge exchange platform for cross-European cultural content integration and mobile publication**

#### **Abstract**

European museums and other cultural institutions host rich collections that have ability to attract EU citizens and tourists. Cultural objects, e.g. paintings, in these collections are related in many ways and in many cases they refer to same underlying concepts, people and places. The Cultural Heritage Knowledge Exchange Platform, SMARTMUSEUM requires that these collections are interoperable over cultural and language barriers, and provides a mobile publication channel for collections.

Keywords: semantics, mobile phones, tourism, personalization

#### **Introduction**

It has been argued that museums should publish their activities, collections, services, and products in cooperation with cultural tourism agencies (Mulrenin, 2002). The Cultural Heritage Knowledge Exchange Platform, SMARTMUSEUM, is a platform for innovative services enhancing on-site personalised access to digital cultural heritage through adaptive and privacy preserving user profiling. Using knowledge bases, global digital libraries and visitors' experiential knowledge, the platform makes possible the creation of innovative multilingual services for increasing interaction between visitors and cultural heritage objects in a future smart museum environment, taking full benefit of digitized cultural information. In this paper we present components needed to realize a knowledge exchange platform for SMARTMUSEUM needs. The annotation framework enables managing of semantic annotations of museum collections. The SmartMuseum recommendation web services use relations defined by ontologies together with user profile and contextual information to provide search and recommendation for a user of SMARTMUSEUM. These web services can be used via mobile and web interfaces.

#### **The Content Architecture of SMARTMUSEUM**

The vast majority of museums hold a legacy database with cataloguing cards for owned objects. In IMSS case (Institute and Museum of the History of Science ) the database contains also some ontological data. The ending result needed from the SMARTMUSEUM application is a set of RDF triples complying to an ontological schema. A clear vision of the target ontology is strictly mandatory especially in a commercial system. Since content providers can have different items and different objectives, small adjustments to the target ontology can be requested for a single implementation, but in general there should be global target ontologies available that all organizations may use.

Next, we will analyze some representative cases about how to map existing metadata to the SMARTMUSEUM ontology. Since not all the required concepts and instances were available in existing ontologies, an integration and an upgrade methodology had to be established to minimize the manual intervention of mapping. In the IMSS case the preexistent database holds information and links among information. For this reason IMSS added SMARTMUSEUM specific ontological annotations to the existing database. This implied a semi-automatic mapping and a manual addition of external references to Getty vocabularies [3]. In the second phase IMSS realized an automatic extractor that used XML-based configuration about what and how to extract from the dataset. The outcome of this process is directly usable by the SMARTMUSEUM system.

#### **SMARTMUSEUM Recommendation Service**

##### **Profile retrieval**

The first phase in the recommendation is to retrieve a profile that matches the user's current context. Retrieving is done by mapping the user's location, determined by GPS to ontological concepts. This is done by expanding the query of the single coordinate point to cover a circular area within some radius  $r$ , say, 1000 meters. This is done in two steps. First, a simple bounding box is created where each of the edges of the box have a distance  $r$  from the user's location, i.e. the distance to the "sides" of the bounding box from the given location is  $r$ . In the second phase, places outside the radius  $r$  are further pruned away from the results. This results into an ontological resource representing the position information.

We use the likelihood of a context generating a certain triple. It can be observed from the relative frequencies of the profile entries. For example, if a user profile contains tags for triples about Italian paintings in the context of

Helsinki, say 10 times, and triples in Helsinki in total 20 times, the P ( $\langle \text{sm:Painting, sm:manufacturedIn, place:Italy} \rangle \mid \langle \text{rdf:Resource, sm:userLocation, place:Helsinki} \rangle$ ) would be  $10/20 = 0.5$ . Because we have the negative or positive votes for the triple we calculate the average of the votes of the triple in the given context and multiply it with the probability of the triple in the context. The contexts in which the observations are done can be very sparse. Therefore, we use Laplace (i.e. add one) smoothing [4] to shave a share of the probability mass to contexts for which no observations are available. In this way, we can observe some probability for every triple even if it has not been tagged in the specified context.

### Recommendation retrieval

Recommendation retrieval is performed by using the query constructed from user profile and context [2]. Based on the earlier phases we have a set of profile triples that each has weight. Each triple may be expanded using query expansion to multiple triples, that each has the weight of the original triples. This is done by including all triple combinations having a Wu-Palmer value higher than a fixed constant. We have set this value to 0.85 by error and trial. As a result we have a set of triples each having a weight.

We can now define the retrieval as a two step matching procedure that utilizes the spatial constraints and a scoring function used to calculate the cosine similarity [6] in vector space model [5], where vectors are formed by using a triples times documents matrix. Further, we cluster the best 300 objects using independent component analysis (ICA). This makes it possible to reveal different viewpoints to the data and avoid over-specialization. For example, if the user has a very strong interest in Italian paintings and a light interest to scientific instruments and telescopes, a traditional retrieval system would only rank Italian paintings high. By clustering, we are able to build three interest clusters, one for Italian paintings, one for telescopes and one for other scientific instruments. Finally, items from each cluster can be included into the final recommendation list.

### Mobile access to SMARTMUSEUM

Mobile access to SMARTMUSEUM is based on two scenarios: the inside scenario - user visits a museum and the outside scenario - user walks around the city, looking for (outdoor) points of interest (POI). In both scenarios the user is equipped with a PDA or a smartphone as a main device for user's positioning and for presenting recommendations and multimedia (A/V, Text To Speech) information. SMARTMUSEUM mobile access is based on five major features:

- User is requested to enter the expected visit duration and a purpose of visit from a predefined list.
- User can acquire contents describing recommended objects by clicking on URLs displayed by the mobile device. According to user's profile preferences, multimedia files or text-to-speech contents are automatically launched when they are available.
- Each object has a unique URI, a URL is usually stored into a RFID tag for indoor scenario, or GPS coordinates are used for outdoor POIs. When looking at an interesting indoor object, user scans the RFID tag attached to the object to get information on the mobile device. Each user browsing action is logged.
- User has an opportunity to rank each page (physical objects and content pages).
- After the end of visit ranking and log information is automatically sent into the SMARTMUSEUM profile server.

### Mobile user positioning and interest monitoring

Mobile user monitoring is performed (1) for positioning and (2) for discovering user's interest in order to process statistical recommendation. Outdoor user positioning in SMARTMUSEUM context is used for determining location for nearby POI search and for determining location context for semantic recommendations. The SMARTMUSEUM solution supports GPS (WGS84 coordinates) and mobile network cell based outdoor localization. For indoor scenario the objects are equipped with 13,56MHz (ISO14443A) RFID tags used for content triggering and user positioning as well.

#### *RFID reading*

A unique solution was developed for RFID tag access. Existing solutions are mainly based on unique hardware ID of the tag. Using this ID for referencing an object would require changes in the SMARTMUSEUM database when tags are replaced. For this reason we made tag data area usable for content. Two solutions have been studied and evaluated: 1) store html contents directly on tags, 2) store an object URI on tag and to retrieve information through wireless network. We made several measurements with 4K tags. Our experiments showed that retrieval of a small html page (3071 octets, without image and CSS style) requires 5 sec. The reading of a simple URI takes only 150 msec. Here the problem is that the user must keep the device near the tag until the end of reading. With the first approach, there is a risk of retrieving partial information. In the second solution, tag reading is almost immediate so that user behaviour is not constrained. Information are automatically loaded

and displayed after tag reading. We made some performance measurements with the server of IMSS. It takes about 1.5 sec to read tag and to load a simple html page (without image of object). Finally the second solution has been fully integrated in the current release of our software.

#### Monitoring user interest feedback

Receiving pertinent user feedback is a crucial factor to process further statistical recommendations for both objects and content pages. Browser activity logging is a widely used method for web content relevance evaluation, especially from server side. In many studies page access duration is taken into account. However, time counting is quite unsuitable for mobile users, because the content access is highly fragmented. So it is difficult to detect idle periods when user, for example, is moving from one object to another one. The second issue is that since content is fully distributed, the activity monitoring cannot be performed in a centralized manner. But in SMARTMUSEUM context, most of the available information is split into separate html pages. In this way it becomes possible to infer user's interest for each museum object and atomic content piece. Mobile software stores all visited URL, and/or visited object URI. And a score is associated to each URIs and URLs.

After the analysis of user scenarios, a combined scale of manual preference input and implicit monitoring of preference and behaviour have been proposed and implemented (see Figure 1):

- score = -0.01: user receives recommended object. If the user shows no interest, and leaves associated recommended link unused, this initial score remains unchanged.
- score = score + 0.3: user fetches basic information about the recommended object or user reads object tag.
- score = score + 0.4: user requests more information about an object to receive a list of additional content URLs.
- score = 1: strong like, manual input on user interface.
- score = -1: strong dislike, manual input on user interface.

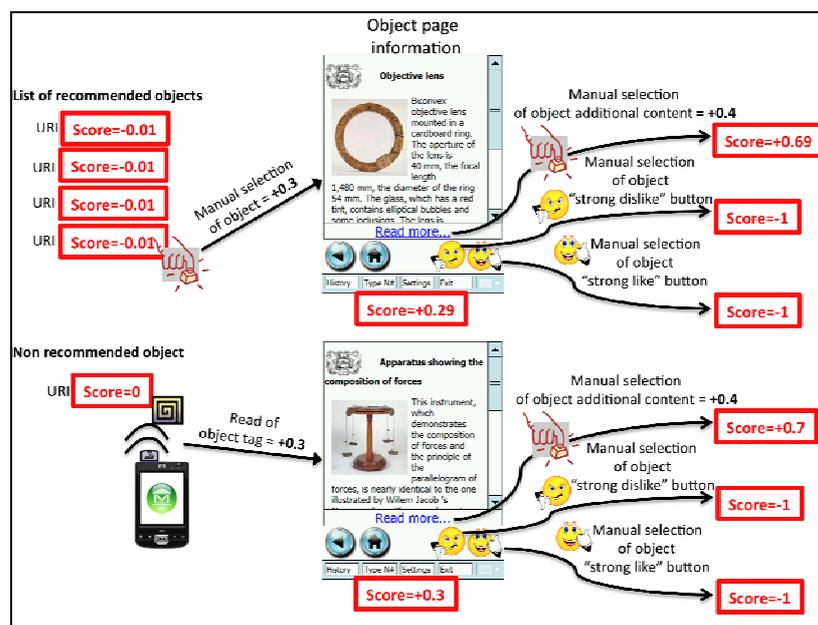


Figure 1: Rating of a SmartMuseum object

Currently the user device client software is implemented for Windows Mobile (WiMo) and Symbian operating system platforms. Two screenshots of WiMo user interface optimized for larger touchscreens are presented in Figure 2 (recommendation list and object information page). Our objective is to minimize user interventions. Upper part of screen is an embedded Web browser window, on bottom part rating buttons can be used. On this screen area multimedia and text-to-speech controls appear automatically when content html page includes respective hidden tags.

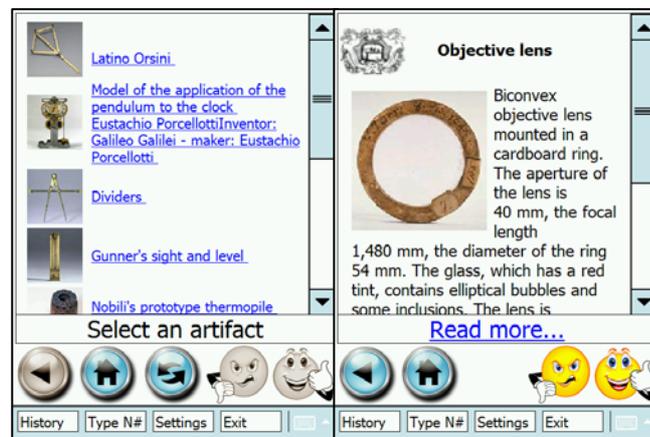


Figure 2: Mobile device main UI (WiMo)

## Conclusions

SMARTMUSEUM is a versatile knowledge exchange platform for hosting and publishing museum collections and POIs for user's of mobile phones and PDAs. In this paper we described different components used for realizing the SMARTMUSEUM.

## Acknowledgements

The research described in this chapter was done in the EU project SMARTMUSEUM supported within the IST priority of the Seventh Framework Programme for Research and Technological Development. Partners of the SMARTMUSEUM include Apprise (Estonia), National Institute for Research in Computer Science and Control INRIA (France), Helsinki University of Technology (Finland), Royal Institute of Technology in Stockholm (Sweden), Webgate JSC (Bulgaria), Heritage Malta (Malta), Institute and Museum of the History of Science (Italy) and Competence Centre of Electronics-, Info and Communication Technologies ELIKO (Estonia).

## References

- [1] Mulrenin, A. (Ed.). (2002). The digicult report. technological landscapes for tomorrow's cultural economy. unlocking the value of cultural heritage. executive summary. Luxembourg: European Commission.
- [2] Tuukka Ruotsalo, Eetu Mäkelä, Tomi Kauppinen, Eero Hyvönen, Krister Haav, Ville Rantala, Matias Frosterus, Nima Dokoohaki and Mihhail Matskin: Smartmuseum: Personalized Context-aware Access to Digital Cultural Heritage. Proceedings of the International Conferences on Digital Libraries and the Semantic Web 2009 (ICSD2009), September, 2009. Trento, Italy.
- [3] [http://www.getty.edu/research/conducting\\_research/vocabularies/](http://www.getty.edu/research/conducting_research/vocabularies/)
- [4] Christopher D. Manning and Hinrich Schuetze. Foundations of Statistical Natural Language Processing. The MIT Press, 1 edition, June 1999.
- [5] G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. Commun. ACM, 18(11):613–620, 1975.
- [6] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. Inf. Process. Manage., 24(5):513–523, 1988.