

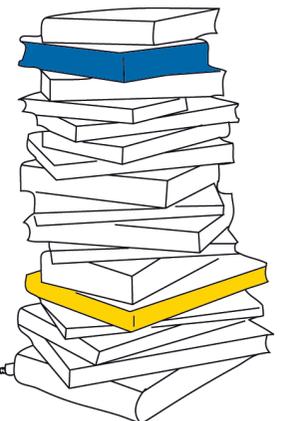
**12 / 13 November 2009**

„Juridicum“ building, University of Vienna



# Perspectives on metadata

**Digital edition & preservation**



# Links



Visit the conference website at <http://metadata.univie.ac.at>

## **The conference is jointly organized and hosted by units of two Austrian Universities:**

Center for Information Modelling in the Humanities (University of Graz)  
<http://www.uni-graz.at/zim/>

Center for Teaching and Learning (University of Vienna)  
<http://ctl.univie.ac.at/>

University Library (University of Vienna)  
<http://bibliothek.univie.ac.at/english>

## **... in relation to the projects:**

GAMS  
<http://gams.uni-graz.at>

Phaidra  
<http://phaidra.univie.ac.at/>

Fedora  
<http://www.fedora-commons.org/>



## **Perspectives on metadata**

Digital edition & preservation



# Programm

## 12. 11. 2009

14:00  
Maria Seissl (Director, Vienna University Library and Archive Service) and Peter Marksteiner (Central Services, Vienna University Computer Center): Welcome and Opening, moderated by Gerhard Budin (University of Vienna)

14:15  
Keynote by Mariella Guercio: Conceptual framework and chain of custody for sustaining the digital trustworthiness (University of Urbino)

15:00  
Andreas Rauber: Home Archiving: Moving Digital Preservation Capabilities from Large Institutions to SMEs and Home Users (Vienna University of Technology)

15:45  
Coffee break

16:00  
Tore Hoel: Metadata in the wild - the laborious creation of a new international standard for describing learning resources (Oslo University College)

16:45  
Sarah Currier: If You Tag it, Will They Come? Metadata Quality and Repository Management (Freelance consultant, JISC CETIS)

17:30  
Johanna Puhl: The PLANETS-Ontology in the context of the PLANETS-Testbed and the XCL-Software (University of Cologne)

18:15  
End of Conference Day 1

18:30  
Evening reception with dinner buffet



## 13. 11. 2009

9:30  
Opening, moderated by Gerhard Budin (University of Vienna)

9:45  
Keynote by Tomaž Erjavec: Metadata as a Strategy of domain-specific Content (Department of Knowledge, Jožef Stefan Institute, Ljubljana)

10:30  
Thomas Burch & Andrea Rapp: Infrastructures for Digital Edition (Competence Center for Electronic Indexing and Publication Procedures in the Humanities, University of Trier)

11:15  
Coffee break

11:30  
Walter Koch: Modelling and Implementation of webservice-based Vocabularies (Department for Analysis und Computational Number Theory, Graz University of Technology)

12:15  
Monika Hagedorn-Saupe: Metadata Harvesting in the European Context (Institute for Museum Research of the State Museums of Berlin – Prussian Heritage)

13:00  
Discussion

13:30  
End of Conference

15:00  
„Digital Library Austria“ Workgroups Meeting



# Abstracts, Day 1

**Mariella Guercio:** (Keynote) Conceptual framework and chain of custody for sustaining the digital trustworthiness

With reference to the subject of my presentation I would like to stress the relevance of a consistent conceptual framework and of the robustness of the methodological approach when the digital curation and preservation will transform the traditional environment. The fragmentation in implementing the organizational and the technological solutions of the investments implies the risk of losing the quality of the investments.

Specifically I would like to focus on the needs for a holistic approach on the whole digital heritage, of course with respect of the specificity of each domain but also with the capacity of planning a system able to provide compatible solutions and integrated for all the digital repositories and digital contents created and acquired by the creators with reference to the policies for accessioning, the procedures and the metadata/attributes/elements for the retrieval and the authenticity, the chain of responsibilities.



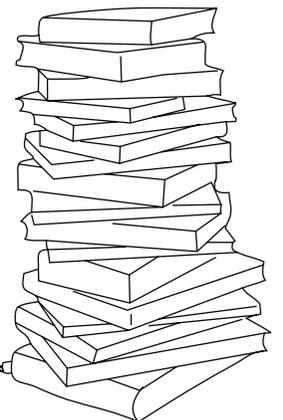
**Tore Hoel:** Metadata in the wild - the laborious creation of a new international standard for describing learning resources

For a number of years the International standards organisation (ISO/IEC JTC1 SC36) has struggled to come up with a new metadata standard for learning resources. The MLR standard was supposed to be compatible with with Dublin Core and IEEE LOM, the first metadata standard in the Learning, Education and Training domain. When the shortcomings of LOM were discovered the long and laborious process of agreeing upon a new and „future proof“ framework for MLR started. The presentation will outline the principles underlying the upcoming MLR standard and reflect on the process of international standardisation.

**Andreas Rauber:** Home Archiving: Moving Digital Preservation Capabilities from Large Institutions to SMEs and Home Users

While large heritage institutions have been addressing the challenges posed by digital preservation need for some time, small institutions are less prepared to handle them, be it due to lacks of funding, expertise or awareness. Yet, tremendous amounts of valuable digital assets are being held by smaller institutions, small and medium enterprises, and - last, but not least - by virtually everybody on their home computers, digital cameras and mobile phones.

This talk will present the HOPPLA system addressing the challenge of automating digital preservation activities to a level where it can be seamlessly and transparently integrated into daily operations, not requiring specific expertise in digital preservation issues. It addresses both bit-level as well as logical preservation, outsourcing preservation planning expertise to external service providers, while integration appropriate migration activities in an automated and transparent manner.



## Perspectives on metadata

Digital edition & preservation

# Abstracts, Day 1

**Sarah Currier:** If You Tag it, Will They Come? Metadata Quality and Repository Management

When we build collections of scholarly works, learning materials, or other educational „stuff“, we want people to be able to find it. This raises a number of problems, including ensuring that resources are tagged with adequate metadata. In 2004 a pioneering paper on this issue noted:

„At its best, “accurate, consistent, sufficient, and thus reliable” (Greenberg & Robertson, 2002) metadata is a powerful tool that enables the user to discover and retrieve relevant materials quickly and easily and to assess whether they may be suitable for reuse. At worst, poor quality metadata can mean that a resource is essentially invisible within the repository and remains unused.” (Currier et al, 2004).

Have the five years since the above-quoted paper was published borne out its prediction: that simply expecting resource authors to create their own metadata at upload would lead to metadata of insufficient quality? Have repository managers been able to persuade funders that including professional metadata augmentation is worth the money? What has been the impact of recent Web developments allowing easier exposure, searching and sharing of resources? How is metadata being treated within the emerging domain of open educational resources? And what does all this mean for repository managers wanting to increase the discoverability of their resources, and to implement workflows for creation of good quality metadata?

**Johanna Puhl:** The PLANETS-Ontology in the context of the PLANETS-Testbed and the XCL-Software

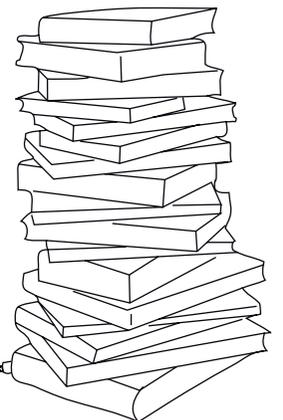
Within the EC-funded project „Planets“: „Preservation and Long-term Access through Networked Services“, the subproject Testbed delivers a web-service containing different experiment-pathways in order to find the best services for long-term preservation of digital objects.

Such services can be characterisation, migration and emulation.

In a scenario, where a librarian/archivist tries to find the best tools and destination-formats for a certain file-collection in an automated procedure, an ontology that constitutes relations between different file-format properties as well as properties of metadata-formats is very desirable and can fundamentally support an evaluation of these tools.

The XCL-Ontology is one part of the XCL-Software developed in cologne for the PLANETS-Subproject „Preservation Characterisation“ and constitutes different experiment-tools in the Planets-Testbed.

In this presentation the XCL-Ontology and it's overall-structure will be explained and it will be described how the Ontology fulfils the different needs of the user-community in the context of digital preservation and PLANETS.



## Perspectives on metadata

Digital edition & preservation

# Abstracts, Day 2

## **Tomaž Erjavec:** (Keynote) Metadata as a Strategy of domain-specific Content Modelling

Intelligent storage and management of digital resources have become a key issue in all scientific disciplines. Concurrently, the term “metadata” itself has evolved beyond the context of library science to describing general aspects of content modelling and semantisation of digital objects.

The Text Encoding Initiative (TEI) Guidelines are a formal specification and accompanying documentation defining a vocabulary of XML elements, meant for the annotation of texts for scholarly purposes. TEI is widely used for the encoding of texts in digital libraries, esp. those with complex annotations. A required part of every TEI-conformant document is its meta-data, i.e. the `teiHeader` element, with a rich and detailed structure.

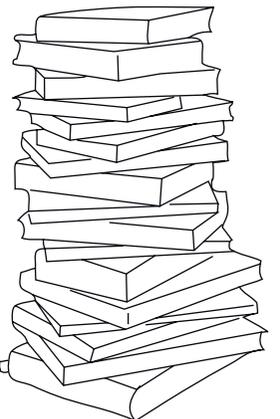
The talk illustrates our use of TEI, esp. the TEI header in the modelling of varied text types: digital manuscript editions, dictionaries and language corpora. The use of XML and TEI makes for a flexible and well-documented modelling framework, which can be easily integrated into Web based applications, with free text or complex metadata searches.

## **Thomas Burch & Andrea Rapp:** Infrastructures for Digital Edition

The infrastructural projects introduced in this talk pursue different goals and methods, but can profit from one another to provide their users with optimal solutions.

The virtual research platform FuD, developed for the collaborative research centre 600 “Foreignness and Poverty”, facilitates the cooperation of projects from diverse humanistic and social science disciplines by providing a framework for work-, publication- and information processes based on the specific needs of its users.

The virtual research environment TextGrid is based on Grid-Technology and develops central, generic services and tools for various humanistic and cultural science communities. The modular structure of both systems and their consideration for existing standards and interfaces allows for an exchange of data and the common development of certain services.



## **Perspectives on metadata**

Digital edition & preservation

## Abstracts, Day 2

**Walter Koch:** Modelling and Implementation of webservice-based Vocabularies

The use of controlled vocabularies is a standard approach within documentation and information activities. Since the advent of computer usage of printed material has been complemented by computerized access especially via the Internet. Global and immediate access to the Web have caused new means of use and integration of computerized and digitized vocabularies.

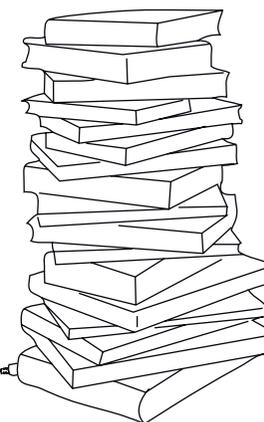
Based on the structural elements outlined in ANSI Z39.19 or related ISO standards the author together with AIT – Angewandte Informationstechnik Forschungsgesellschaft mbH has developed a SOAP/WSDL based Webservice which supports different structures as lists, taxonomies or thesauri. This service supports the retrieval of topics or concepts (identified by a term) using metadata connected with them using boolean logic which also supports faceted searches. As the result of such a service not only metadata related to a term is available, but the semantic structure in which a concept is embedded will be returned, too.

In the presentation examples will be given how authority files or thesauri have been implemented, among them TGN - the Thesaurus of Geographic Names (integrated in museum collection management systems) - or the IconClass classification system. Other aspects include the use of vocabularies in multilingual applications (demonstrated by projects of Europeana, the European Digital Library), visualization of semantic networks based on XTM-TopicMaps and the access to vocabulary Webservices in a modern Cloud Storage and Computing environment.

**Monika Hagedorn-Saupe:** Metadata Harvesting in the European context

In the field of cultural heritage there is a wide range of databases in use to document the books, museum objects, maps, films a.s.o. In order to make information on cultural heritage available throughout Europe and to allow easy access for scientists as well as the general public, a joint initiative on a European level has been put into place: the portal [www.europeana.eu](http://www.europeana.eu). Experts from the fields of libraries, museums, archives and media archives are working together in several European projects to build up Europeana. Currently, a beta-version of the portal is online - already comprising about five million objects. The metadata-format currently used is a basic one: ESE (an extended Dublin Core), based on a more librarian approach. In the museum field working groups in the US, CDWA-lite has been developed, mainly for art museums. In Germany a working group has developed the harvesting museumdat, which is based on CDWALite, but extended in order to serve also other type of museum collections. Within the ATHENA project, which aims especially to support museums in delivering their data to Europeana, the museumdat standard has been crosschecked and further enriched by the UK standard SPECTRUM.

By applying those standards it is much easier for museums to provide their data not only to Europeana, but also to national, regional and also thematic portals, thus allowing to link knowledge much more than was possible in the past.



## Perspectives on metadata

Digital edition & preservation

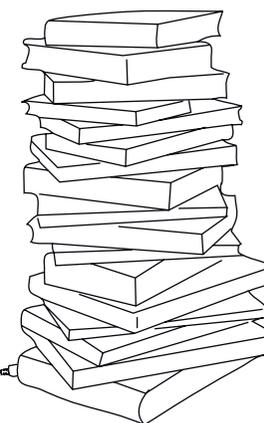
## Profiles

**Mariella Guercio** is a full professor in archival science and ERM at the University of Urbino since 2000. She cooperates with the State Authority for information technology to define the Italian legislation for ERMS. Co-director of the European project ERPANET and partner for the DELOS digital preservation WP, she is partner for the European project CASPAR on digital preservation (2006-2009), member of the European working group for defining a master profile for digital curation and director of the team Italy of the InterPARES project. Since 2002, she is the director of the journal *Archivi & Computer*. She has chaired masters in archival sectors for the University and for Italian Scuola superiore della pubblica amministrazione where now she is responsible of all the education and advanced training programmes for digital curation. She published many articles, a manual on archival science (as co-author with Paola Carucci) and a manual on electronic records.

**Andreas Rauber** is Associate Professor at the Department of Software Technology and Interactive Systems (ifs) at the Vienna University of Technology (TU-Wien). He furthermore is president of AARIT, the Austrian Association for Research in IT and a Honorary Research Fellow in the Department of Humanities Advanced Technology and Information Institute (HATII), University of Glasgow. He received his MSc and PhD in Computer Science from the Vienna University of Technology in 1997 and 2000, respectively. In 2001 he joined the National Research Council of Italy (CNR) in Pisa as an ERCIM Research Fellow, followed by an ERCIM Research position at the French National Institute for Research in Computer Science and Control (INRIA), at Rocquencourt, France, in 2002. From 2004-2008 he was also head of the iSpaces research group at the eCommerce Competence Center (ec3). In 1998 he received the ÖGAI Award of the Austrian Society for Artificial Intelligence (ÖGAI), and the Cor-Baayen Award of the European Research Consortium for Informatics and Mathematics (ERCIM) in 2002. He has published numerous papers in refereed journals and international conferences and served as PC member and reviewer for several major journals, conferences and workshops. He is a member of the Association for Computing Machinery (ACM), The Institute of Electrical and Electronics Engineers (IEEE), the Austrian Society for Artificial Intelligence (ÖGAI) and serves on the board of the IEEE Technical Committee on Digital Libraries (TCDL).

**Sarah Currier** is a librarian who has been involved in research, development and delivery related to metadata, repositories and content in e-learning for the past 10 years. Her experience spans involvement in standards bodies (she is currently Co-Moderator of the Dublin Core Education Community); national-level support for initiatives in UK higher education (via several roles within JISC CETIS); R&D projects (including the seminal SeSDL repository, the Community Dimensions of Learning Object Repositories Project, and, in an advisory capacity, the Jorum national learning materials repository); development and management of strategic repository services (most notably the IRIS Learning Exchange in Scotland); and two years in a key role in customer support and product management with the UK's leading learning materials repository software supplier, Intrallect Ltd. In 2004 Sarah published some of the earliest research into metadata quality for e-learning, and is now a freelance consultant with her own company, Sarah Currier Consultancy. She is most excited by one of her current projects (SHEEN Sharing, with the Scottish Funding Council and Higher Education Academy), looking at the use of free Web 2.0 tools to create a virtual repository for a small community of practice involved in promoting employability in Scottish HE.

**Tore Hoel** is a Norwegian expert to ISO/IEC JTC1 SC36 and co-editor of the MLR Part 5 Educational. He has been working on standards for learning technologies since 2003, and is now the Vice Chair of the European Committee for Standardization Workshop on Learning Technologies. Hoel is currently work package co-leader for dissemination & roadmapping in the European ICOPER project, as well as doing research for a PhD on standards governance and the use of tools to improve the standardisation process.



## Perspectives on metadata

Digital edition & preservation

# Profiles

**Johanna Puhl** is since 07/2008 project-manager at the University of Cologne for the PLANETS-project, participation in the Work-packages Testbed and Preservation Characterisation (Development of the XCL-Ontology and collection of the test-corpus for the Testbed-Service). 2008 Master's degree: Magister Artium; MA thesis on: Evaluation of digital Watermarks for the authenticity of photographs. 10/2001 - 07/2008 Study of Humanities Computer Science (HKI), Sociology and theater-/ film-studies at the University of Cologne. Work Experience: 1998 - 2001: Apprenticeship as mediadesigner for print-products. 10/2001 - 07/2008: Student Assistant in different companies and freelancer mediadesigner.



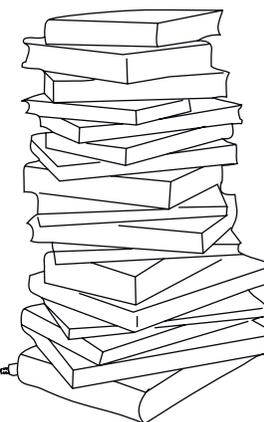
**Tomaž Erjavec** is a research associate at the Dept. for Knowledge Technologies and the Jožef Stefan Institute in Slovenia. He has been working in the field of language technologies for over two decades, and was involved in a number of projects to standardise the encoding of language resources such as corpora, computational lexica and digital editions. He was a member of the Text Encoding Initiative Council and of several Working Groups of ISO TC 37 "Terminology and other language and content resources".

**Andrea Rapp** has been the head of the Goettingen Digitization Center at the State and University Library in Goettingen and is now executive director of the Competence Center for Digital Humanities at Trier University. She has been working in the field of digitization, digital editions, electronic dictionaries, and digital humanities in general for over 15 years. She is in charge of several DFG and BMBF-funded projects and is one of the initiators of the TextGrid project.

**Thomas Burch** worked as freelancer at the Saarbrücker Zeitung Verlag GmbH and the Haufe Verlagsgruppe in Freiburg. 1998 he was one of the founders of the Competence Center for Digital Humanities at Trier University and is one of the executive directors. He is in charge of several DFG-funded projects and is one of the initiators of the FuD project.

**Walter Koch** was born on the 16th of February 1947 in Graz. He received a PhD in Mathematics and Physics from University of Graz in 1970. He is director for AIT Ltd. (Angewandte Informationstechnik Forschungsgesellschaft mbH) and is an ao.Univ.-Professor at the University of Graz and a Guest Professor at the University of Krems. Furthermore he is chairperson of the CSC Europe EEIG and head of the Steinbeis Transferzentrum (IMCHI – Information Management and Cultural Heritage Informatics). Prof. Koch's relevant experience includes: Head of different research institutes at Joanneum Research Ltd. in Graz (1976-1998); Consultant and contractor to different national and international organizations (eg UNESCO, DFG, EC, ESA, Austrian ministries, Graz municipality); Member of several national and international scientific associations (eg. ICOM, ONORM, VÖB).

**Monika Hagedorn-Saupe** studied education, sociology, psychology - with an emphasis on adult education - and mathematics at the Ruhr Universität Bochum, the Kings College London and the Freie Universität Berlin and was appointed honorary professor at the HTW Berlin in 2006. Since 1985, she has been a staff member at the Institute for Museum Research at the State Museums of Berlin, Prussian Heritage, leading the department „Besucherbezogene Museumsforschung und Kulturstatistik“ since July 1994 and acting as the deputy director of the institute since October 1994. Spokesperson of the professional group on Documentation of the German Museums Association; Secretary of ICOM CIDOC (International Council of Museums - Comité internationale pour la documentation).



## Perspectives on metadata

Digital edition & preservation

## Related event

### **„Digital Library Austria“ Workgroups Meeting**

Following the conference proper, participants are cordially invited to join the Digital Library Austria Workgroups Meeting, in relation to the Europeana Local project.

#### **First Workgroup Meeting**

of the Digital Library Austria Workgroups  
Friday, 13 November, 2009 from 15:00 to 17:30  
Top Floor of „Juridicum“ building, Schottenbastei 10-16, Vienna

Agenda:

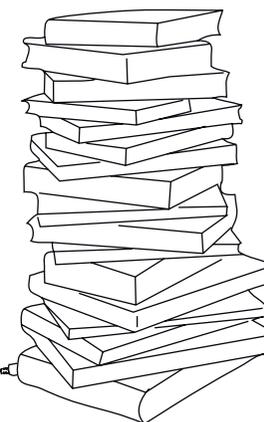
1. The Digital Library Austria & EUROPEANA
2. Constitution of national workgroups
3. Procedural and organisational planning

Further Information on Europeana Local can be found at  
<http://www.europeanalocal.eu> or the  
Austrian mirror site <http://www.europeanalocal.at>



## **Perspectives on metadata**

**Digital edition & preservation**



powered by **CTL** Center for Teaching and Learning **LINIECITÄT BIBLIOTHEK** <ZIM<sup>ig</sup>/>

related projects  **GAMS**  **PHAIDRA**

in cooperation with **ICOPER**

sponsored by  **BUL**  **EMC<sup>2</sup>**  
where information lives<sup>®</sup>