



Proceedings of the 5th International Workshop on
Semantic Digital Archives
(SDA 2015)

held as part of the
19th International Conference on Theory and Practice of Digital Libraries
(TPDL 2015)
September 18, 2015 in Poznan, Poland

<http://sda2015.dke-research.de>

Edited by:

Thomas Risse, L3S Research Center, Hannover, Germany

Livia Predoiu, University of Oxford, United Kingdom

Andreas Nürnberger, Otto-von-Guericke University of Magdeburg, Germany

Seamus Ross, University of Toronto, Canada



Vol-1529
urn:nbn:de:0074-1529-8

Copyright © 2015 for the individual papers by the papers' authors. Copying permitted only for private and academic purposes. This volume is published and copyrighted by its editors.

Preface

The 5th Workshop on Semantic Digital Archives (SDA 2015) has built upon the success of the previous editions in 2011 to 2014 and has been held as part of the 19th International Conference on Theory and Practice of Digital Libraries (TPDL 2015) on September 18, 2015 in Poznan, Poland. Organized as full-day workshop, SDA 2015 has aimed to promote and discuss sophisticated knowledge representation and knowledge management solutions specifically designed for improving Archival Information Systems.

Archival Information Systems are systems that are tailored to preserve digital information and provide access to current and future users. Such systems are becoming increasingly important. For decades, the amount of content created digitally is growing and its complete life cycle nowadays tends to remain digital. A selection of this content is expected to be of value for the future and can thus be considered being part of our cultural heritage. However, digital content poses many challenges for long-term or indefinite preservation, e.g. as digital publications become increasingly complex by the embedding of different kinds of multimedia, data in arbitrary formats and software. As soon as these digital publications become obsolete, but are still deemed to be of value in the future, they have to be transferred smoothly into appropriate Archival Information System where they need to be kept accessible even through changing technologies.

The successful previous Semantic Digital Archives (SDA) workshops showed: Both, the library and the archiving community have made valuable contributions to the management of huge amounts of knowledge and data. However, both are approaching this topic from different views which shall be brought together to cross-fertilize each other. There are promising combinations of pertinence and provenance models since those are traditionally the prevailing knowledge organization principles of the library and archiving community, respectively. Another scientific discipline providing promising technical solutions for knowledge representation and knowledge management is semantic technologies, which is supported by appropriate W3C recommendations and a large user community. At the forefront of making the semantic web a mature and applicable reality is the linked data initiative, which already has started to be adopted by the library community and the digital humanities community showcasing already exciting applications involving end-users. It can be expected that using semantic (web) technologies in general and linked data in particular can mature the area of digital archiving as well as technologically tighten the natural bond between digital libraries and digital archives. Semantic representations of contextual knowledge about cultural heritage objects will enhance organization and access of data and knowledge. In order to achieve a comprehensive investigation, the information seeking and document triage behaviors of users (an area also classified under the field of Human Computer Interaction) needs also to be included in the research.

One of the major challenges of digital archiving is how to deal with changing technologies and changing user communities. On the one hand software, hardware and (multimedia) data formats that become obsolete and are not supported anymore still need to be kept accessible. On the other hand changing user communities necessitate technical means to formalize, detect and measure knowledge evolution. Furthermore, archival records are usually not deleted from classical archives and correspondingly, digital archival records might have to be preserved infinitely in Archival Information Systems as well. Therefore, the amount of digitally

archived (multimedia) content can be expected to grow rapidly. Efficient storage management solutions adequate for big data and geared to the fact that cultural heritage is not as frequently accessed like up-to-date content residing in a digital library are required. Software and hardware needs to be tightly connected based on sophisticated knowledge representation and management models in order to face that challenge.

In line with the above, we invited contributions to the workshop that focus on:

- Architectures and Frameworks for semantic Archival Information Systems (AIS) and Archival Information Infrastructures (AII)
- Semantic (Web) services implementing AIS & AII
- Contextualization of digital archives, museums and digital libraries
- Linked data for AIS, AII, museums and digital libraries
- Ontologies for AIS, AII, museums and digital libraries
- Semantics of complex content (e.g. Social Media, Multimedia)
- Information integration/semantic ingest (e.g. from digital libraries)
- Semantic search & information retrieval in digital archives, digital museums and digital libraries
- User interfaces for (semantic) AIS, AII, digital museums & semantic digital libraries
- Semantics for Preservation Processes and Protocols
- Preservation of work flow processes
- (Semantic) provenance models
- Semantics for the appraisal and selection of content
- Evolving semantics in long-term archives
- Trust for ingest & data security/integrity check for long-term storage of archival records
- User studies focusing on end-user needs and information seeking behavior of end-users
- Implementations & evaluations of (semantic) AIS, AII, semantic digital museums & semantic digital libraries
- Semantic long-term storage & hardware organization for AIS & AII & digital libraries

We received submissions covering a broad range of relevant topics in the area of semantic digital archives. With the help of our program committee all articles were peer-reviewed. These proceedings comprise all accepted submissions which have been carefully revised and enhanced by the authors according to the reviewers' comments.

One focus of this year's edition of the workshop was the topic of managed forgetting and contextualizing remembering, which was nicely introduced by the invited talk given by Claudia Niederée. The accepted papers of this year's edition covered topics from preserving research data involving geographical information, an archival infrastructure for a textbook research archive, semantic expert search in textbook research archives, automated annotation for the SciDocAnnot scientific document model supporting faceted search, an investigation of the usability of anchor text as proxy for user queries and, finally, use cases, requirements and models for entity-centric preservation of linked open data. All these topics lie at the heart of the area of Semantic Digital Archives and provided the base for fruitful discussions, also in the final panel discussion of the workshop.

We sincerely thank all members of the program committee for supporting us in the reviewing process. Altogether, the diversity of the papers in these proceedings represent a multitude of

interesting facets about the exciting and promising research field of semantic digital archives and semantic digital archiving infrastructures. During the workshop itself we had many fruitful and inspiring discussions which would not have been possible without the well done presentations and the interested audience. Many thanks to all workshop attendants for a great workshop!

We would also like to thank Sun SITE Central Europe for hosting these proceedings on <http://ceur-ws.org>.

December 2015

T. Risse, L. Predoiu, A. Nürnberger, and S. Ross

Organizing Committee

- Thomas Risse, L3S Research Center, Hannover, Germany
- Livia Predoiu, University of Oxford, Oxford, UK
- Andreas Nürnberger, Otto-von-Guericke University, Magdeburg, Germany
- Seamus Ross, University of Toronto, Toronto, Canada

Program Committee

- Elena Demidova, L3S Research Center, Hannover, Germany
- Tudor Groza, The Garvan Institute of Medical Health, Australia
- Claus-Peter Klas, FernUniversität in Hagen, Germany
- Birger Larsen, Royal School of Library and Information Science, Denmark
- Thomas Lukasiewicz, University of Oxford, UK
- Erik Mannens, iMinds, Ghent University, Belgium
- Annett Mitschick, TU Dresden, Germany
- Andreas Nürnberger, Otto-von-Guericke University, Magdeburg, Germany
- Gillian Oliver, Victoria University of Wellington, New Zealand
- Jacco van Ossenbruggen, VU University Amsterdam, Netherlands
- Livia Predoiu, University of Oxford, Oxford, UK
- Andreas Rauber, Vienna University of Technology, Austria
- Thomas Risse, L3S Research Center, Hannover, Germany
- Seamus Ross, University of Toronto, Toronto, Canada
- Heiko Schuldt, Universität Basel, Switzerland
- Herbert van de Sompel, Los Alamos National Laboratory Research Library, USA
- Marc Spaniol, Max-Planck-Institut Saarbrücken, Germany
- Jun Zhao, University of Oxford, Oxford, UK

Table of contents

Invited Talk

Learning from Human Memory: Managed Forgetting and Contextualized Remembering for Digital Memories	1
<i>Claudia Niederée</i>	

Semantics in Practice

Integrating Research Data Management into Geographical Information Systems	7
<i>Christian T. Jacobs, Alexandros Avdis, Simon L. Mouradian and Matthew D. Piggott</i>	
Semantic-based Expert Search in Textbook Research Archives	18
<i>Marco Pavan and Ernesto William De Luca</i>	
An Automated Annotation Process for the SciDocAnnot Scientific Document Model	30
<i>Hélène De Ribaupierre and Gilles Falquet</i>	
World Views – A Digital Archive Infrastructure for the Georg Eckert Institute for International Textbook Research	42
<i>Lena-Luise Stahn, Steffen Hennicke and Ernesto William De Luca</i>	

Searching, Preserving and Forgetting

Temporal Anchor Text as Proxy for Real User Queries	49
<i>Thaer Samar and Arjen P. de Vries</i>	
Entity-Centric Preservation for Linked Open Data: Use Cases, Requirements and Models .	61
<i>Elena Demidova, Thomas Risse, Giang Binh Tran and Gerhard Gossen</i>	

