Proceedings of the 2nd International Workshop on **Semantic Digital Archives**

held in conjunction with the
16th Int. Conference on Theory and Practice of Digital Libraries (TPDL)
on September 27, 2012 in Paphos, Cyprus.

http://sda2012.dke-research.de

Edited by

Annett Mitschick,
Technische Universität Dresden, Germany, annett.mitschick@tu-dresden.de

Fernando Loizides,
Cyprus University of Technology, fernando.loizides@cut.ac.cy

Livia Predoiu,
Otto-von-Guericke University Magdeburg, Germany, livia.predoiu@ovgu.de

Andreas Nürnberger,
Otto-von-Guericke University Magdeburg, Germany, andreas.nuernberger@ovgu.de

Seamus Ross,
University of Toronto, Canada, seamus.ross@utoronto.ca

September, 2012

Preface

The 2nd Workshop on Semantic Digital Archives (SDA 2012) builds upon the success of the previous edition in 2011, held in conjunction with the International Conference on Theory and Practice of Digital Libraries, TPDL 2011 (formerly known as European Conference on Digital Libraries, ECDL). Organized as full-day workshop, SDA 2012 aims to advance and discuss appropriate knowledge representation and knowledge management solutions specifically designed for improving Archival Information Systems. The main objective is to have a closer dialogue between the technical oriented communities with people from the (digital) humanities and social sciences, as well as cultural heritage institutions in general in order to approach the topic from all relevant angles and perspectives. This workshop is an exciting opportunity for collaboration and cross-fertilization.

Intending to have an open discussion on topics related to the general subject of Semantic Digital Archives, we invited contributions that focus on one of the following topics:

- Ontologies & linked data for digital archives and digital libraries (incl. multimedia archives)
- Semantic search & semantic information retrieval in digital archives and digital libraries (incl. multimedia archives)
- Implementations and evaluations of semantic digital archives
- Theoretical and practical archiving frameworks using Semantic (Web) technologies
- Semantic or logical provenance models for digital archives or digital libraries
- Visualization and exploration of content in large digital archives
- User interfaces for semantic digital libraries and intelligent information retrieval
- User studies focusing on end-user needs and information seeking behavior of end-users
- Semantic (Web) services implementing the OAIS standard
- Logical theories for digital archives
- Knowledge evolution
- Information integration/semantic ingest (e.g. from digital libraries)
- Trust for ingest & data security/integrity check for long-term storage of archival records
- Semantic extensions of emulation/virtualization methodologies for digital archives
- Semantic long-term storage and hardware organization tailored for digital archives
- Migration strategies based on Semantic (Web) technologies

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.

These papers were joined by an invited keynote by *Andreas Rauber* (Vienna University of Technology, Austria). In *Digital Preservation in Data-Driven Science: On the Importance of Process Capture, Preservation and Validation* he points out the necessity of capturing and documenting processes (in addition to the context of data objects), especially in e-Science and business settings, and presents an approach for process preservation and verification upon later re-execution.

The paper *Entity Extraction and Consolidation for Social Web Content Preservation (S. Dietze et al.)* presents an approach to extract and consolidate information from archived social Web content in order to facilitate semantic search of Web archives. The work was developed in the EC-funded Integrating Project ARCOMEM.

With *Do we need metadata? - An On-line Survey in German Archives* Marcel Ruhl presents the revealing results of a survey among German archives regarding the use of metadata standards for the annotation of audiovisual media.

The paper *Automatic Classification of Scientific Records using the German Subject Heading Authority File (Ch. Wartena & M. Sommer)* introduces an approach to assign subject classifications to records without using machine learning techniques but by the application of the German Subject Heading Authority File (SWD).

In *Pundit: Semantically Structured Annotations for Web Contents and Digital Libraries (M. Grassi et al.)* the authors propose an annotation system, developed in the context of the Semlib project, which provides the user with the ability to annotate distributed resources, i.e. multimedia content published on the Web, using an extension of the Open Annotation Collaboration (OAC) ontology.

The paper *Towards a Recommender System for Statistical Research Data (D. Bahls et al.)* presents the conceptual ideas and the system architecture of a case-based recommender system for statistical data used in scientific research, and discusses possible similarity measures and notification services.

With *A method and guidelines for the cooperation of ontologies and relational databases in Semantic Web applications* L. Bozzato et al. showcase a methodology for mapping relational data to an ontology structure to support SPARQL queries and inference and to take advantage of the representation possibilities offered by both data models.

A critical issue when developing RDF-based semantic archives is the right choice of an appropriate large-scale storage solution for the data. *Yet Another Triple Store Benchmark? Practical Experiences with Real-World Data (M. Voigt et al.)* presents the experimental setting and the results of extensive performance tests of state-of-the-art RDF stores using non-synthetic RDF datasets.

Finally, the paper *Implementing CIDOC CRM Search Based on Fundamental Relations and OWLIM Rules (V. Alexiev)* proposes an approach to provide a higher-level perspective on RDF data by mapping complex sub-graph patterns to simpler, more abstract descriptions using OWLIM rules. The author presents an implementation of the concept regarding search with the CIDOC Conceptual Reference Model.

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.

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

September 2012

A. Mitschick, F. Loizides, L. Predoiu, A. Nürnberger, and S. Ross

Program Committee

Vassilis Christophides	Foundation of Research & Technology - Hellas, Greece
Kai Eckert	University Library of Mannheim, Germany
Armin Haller	CSIRO ICT Centre, Australia
Steffen Hennicke	Humboldt-Universität zu Berlin, Germany
Stijn Heymans	SRI International, USA
Pascal Hitzler	Wright State University, USA
Christian Keitel	State Archive of Baden-Württemberg, Germany
Birger Larsen	Royal School of Library and Information Science, Denmark
Thomas Lukasiewicz	University of Oxford, UK
Mathias Lux	Klagenfurt University, Austria
Knud Möller	Talis, Birmingham, UK
Kai Naumann	State Archive of Baden-Württemberg, Germany
Jacco van Ossenbruggen	VU University Amsterdam, Netherlands
Andreas Rauber	Vienna University of Technology, Austria
Thomas Risse	L3S Research Center, Hannover, Germany
Sebastian Rudolph	Karlsruher Institut für Technologie, Germany
Mike Salampasis	Alexander Technology Educational Institute of Thessaloniki, Greece
Herbert van de Sompel	Los Alamos National Laboratory Research Library, USA
Marc Spaniol	Max-Planck-Institut Saarbrücken, Germany
Manfred Thaller	University of Cologne, Germany

Table of Contents

Digital Preservation and Metadata

Structuring and Recommendation

Semantic Technologies and Ontologies