



Proceedings of the 3rd International Workshop on **Semantic Digital Archives**

held in conjunction with the
17th Int. Conference on Theory and Practice of Digital Libraries (TPDL)
on September 26, 2013 in Valetta, Malta.

<http://sda2013.dke-research.de>

Edited by

Livia Predoiu, University of Oxford, UK

Annett Mitschick, University of Dresden, Germany

Andreas Nürnberger, University of Magdeburg, Germany

Thomas Risse, University of Hannover, Germany

Seamus Ross, University of Toronto, Canada

November, 2013

Copyright © 2013 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.

Acknowledgement: Original photo of the Siege Bell War Memorial in Valetta/Malta taken by Seán Russell, PhD Researcher, Clarity Centre for Sensor Web Technologies, School of Computer Science & Informatics, Science North, University College Dublin, Belfield, Dublin 4 Ireland, Sean.Russell@ucd.ie.

Preface

The 3rd Workshop on Semantic Digital Archives (SDA 2013) has built upon the success of the previous editions in 2011 and 2012, and has been held in conjunction with the International Conference on Theory and Practice of Digital Libraries, TPD (formerly known as European Conference on Digital Libraries, ECDL). Organized as full-day workshop, SDA 2013 has aimed to advance and discuss appropriate knowledge representation and knowledge management solutions specifically designed for improving Archival Information Systems. The main objective was 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 was indeed 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.

The talks given at the workshop were joined by an invited keynote by *Giorgio Maria Di Nunzio* (University of Padua, Italy) on the subject of *Digital Geolinguistics: On the Use of Linked Open Data for Data-Level Interoperability between Geolinguistic Resources*, which gave an interesting overview on how linked data is used within the Open Language Archives community.

In the paper *Towards Preservation of semantically enriched Architectural Knowledge* Stefan Dietze et al. present an overview of early work within the DURAARK project for creating a semantic digital archive for the building and architecture domain.

Nina Tahmasebi and Thomas Risse (*The Role of Language Evolution in Digital Archives*) discuss problems that arise as a result of language evolution in digital archives and review methods to make digital archives semantically accessible and interpretable in the future.

With *BlogNEER: Applying Named Entity Evolution Recognition on the Blogosphere?* Helge Holzmann et al. propose an approach for the identification of named entity changes within blog content using a novel semantic filtering method and semantic resources like DBpedia.

In the paper *Elevating Natural History Museums' Cultural Collections to the Linked Data Cloud* Giannis Skevakis et al. introduce an infrastructure and methodology developed and applied for the transition and conversion of cultural heritage metadata to Linked Data within the Natural Europe project.

Alexandre Rademaker et al. showcase *A linked open data architecture for contemporary historical archives*, being proposed for the Center for Teaching and Research in the Social Sciences and Contemporary History of Brazil (CPDOC).

In *Pundit: Creating, Exploring and Consuming Semantic Annotations* Marco Grassi et al. present new results and aspects of their matured project (a semantic web annotation tool) focusing on demonstrative use cases for sharing, exploring and visualizing semantic annotations.

Stefan Haun and Andreas Nürnberger (*Towards Persistent Identification of Resources in Personal Information Management*) discuss the requirements and critical aspects of persistent identifiers for resources and entities in Personal Information Management, especially in personal file systems, which is also important in digital archives and digital libraries.

The paper *A Storage Ontology for Hierarchical Storage Management Systems* (Sandro Schmidt et al.) introduces an ontology-based concept to extensively describe files, their properties and interrelations as a basis to decide whether they should be stored on fast or cheap storage media.

In *Semantic Retrieval Interface for Statistical Research Data* Daniel Bahls and Klaus Tochtermann propose an intuitive multi-step retrieval prototype which supports researchers in finding, accessing and composing distributed statistical data sets.

Finally, the paper *Semantification of Query Interfaces to Improve Access to Deep Web Content* (Arne Martin Klemenz and Klaus Tochtermann) discusses the limitations of Deep Web Information Retrieval and the significant potential of additional semantic annotations and proposes the idea of 'Semantic Deep Web Services'.

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>.

November 2013

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

Program Committee

Sören Auer	University of Leipzig, Germany
Kai Eckert	University of Mannheim, Germany
Marco Grassi	Università Politecnica delle Marche, Italy
Tudor Groza	University of Queensland, Australia
Armin Haller	CSIRO, Australia
Andreas Harth	KIT, Karlsruhe, Germany
Steffen Henniecke	Humboldt-Universität zu Berlin, Germany
Stijn Heymans	SRI International, USA
Pascal Hitzler	Wright State University, USA
Claus-Peter Klas	FernUniversität in Hagen, Germany
Thomas Lukasiewicz	University of Oxford, UK
Mathias Lux	Klagenfurt University, Austria
Knud Möller	Datalysator, Berlin, Germany
Kai Naumann	State Archive of Baden-Württemberg, Germany
Gillian Oliver	Victoria University of Wellington, New Zealand
Jacco van Ossenbruggen	VU University Amsterdam, Netherlands
Andreas Rauber	Vienna University of Technology, Austria
Sebastian Rudolph	University of Dresden, Germany
Heiko Schuldt	Universität Basel, Switzerland
Kunal Sengupta	Wright State University, USA
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

Invited Talk:

Digital Geolinguistics: On the Use of Linked Open Data for Data-Level Interoperability between Geolinguistic Resources	1
<i>Giorgio Maria Di Nunzio</i>	

Digital Preservation and Language Evolution

Towards Preservation of semantically enriched Architectural Knowledge	4
<i>Stefan Dietze, Jakob Beetz, Ujwal Gadiraju, Georgios Katsimpras, Raoul Wessel and René Berndt</i>	
The Role of Language Evolution in Digital Archives	16
<i>Nina Tahmasebi and Thomas Risse</i>	
BlogNEER: Applying Named Entity Evolution Recognition on the Blogosphere?	28
<i>Helge Holzmann, Nina Tahmasebi and Thomas Risse</i>	

Linked Open Data and Semantic Web Content

Elevating Natural History Museums' Cultural Collections to the Linked Data Cloud	40
<i>Giannis Skevakis, Konstantinos Makris, Polyxeni Arapi and Stavros Christodoulakis</i>	
A Linked Open Data Architecture for Contemporary Historical Archives	52
<i>Alexandre Rademaker, Suemi Higuchi and Dario Augusto Borges Oliveira</i>	
Pundit: Creating, Exploring and Consuming Semantic Annotations.....	65
<i>Marco Grassi, Christian Morbidoni, Michele Nucci, Simone Fonda and Francesca Di Donato</i>	

Aspects of Information Management and Access

Towards Persistent Identification of Resources in Personal Information Management	73
<i>Stefan Haun and Andreas Nürnberger</i>	
A Storage Ontology for Hierarchical Storage Management Systems.....	81
<i>Sandro Schmidt, Torsten Wauer, Ronny Fritzsche and Klaus Meißner</i>	
Semantic Retrieval Interface for Statistical Research Data	93
<i>Daniel Bahls and Klaus Tochtermann</i>	
Semantification of Query Interfaces to Improve Access to Deep Web Content	104
<i>Arne Martin Klemenz and Klaus Tochtermann</i>	

