Communications in Computer and Information Science 416

For further volumes:
http://www.springer.com/series/7899

Łukasz Bolikowski · Vittore Casarosa
Paula Goodale · Nikos Houssos
Paolo Manghi · Jochen Schirrwagen (Eds.)

Theory and Practice of Digital Libraries – TPDL 2013 Selected Workshops

LCPD 2013, SUEDL 2013, DataCur 2013
Held in Valletta, Malta, September 22-26, 2013
Revised Selected Papers

Springer

ISSN 1865-0929 ISSN 1865-0937 (electronic)
ISBN 978-3-319-08424-4 ISBN 978-3-319-08425-1 (eBook)
DOI 10.1007/978-3-319-08425-1
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014942569

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This volume contains the papers presented at three different workshops, held in connection with the 17th International Conference on Theory and Practice of Digital Libraries (TPDL 2013). The TPDL Conference Series (formerly known as ECDL, European Conference on Digital Libraries) started in 1997 in conjunction with the activities of the first DELOS Working Group (which later became the DELOS Network of Excellence) and has evolved into the leading scientific forum in Europe, focusing on digital libraries and associated technical, practical, and social issues, meeting the needs of a large and diverse constituency, which includes practitioners, researchers, educators, policy makers, and users.

TPDL 2013 took place during September 22–26 in Valletta, Malta. The conference exhibited the traditional mix of events, which usually include tutorials at the beginning and workshops at the end. TPDL 2013 started with six tutorial sessions, on topics ranging from linked data to preservation to text digitization, to metadata. In parallel with the tutorials, a Doctoral Symposium was held, where PhD students, selected after a call for proposals, had the opportunity to present and discuss their research topics with a panel of experts. Alongside the tutorials, for the first time the Global Workshop of iSchools was held, whose aim was to foster the development of a global iSchool community by inviting delegates from information schools to discuss their academic programs and research strengths/interests so as to allow them to identify areas of possible collaboration.

The conference attracted about 300 delegates from over 40 countries, who presented and discussed challenges and opportunities in digital library architecture, interoperability and information integration, digital library interfaces, user behavior, data re-use and open access, linked data, data visualization, long-term preservation, the Semantic Web in digital libraries and digital curation. The presentations and discussions were continued in a number of workshops held at the end of the conference.

This volume is organized in three chapters, containing the accepted papers of the three workshops LCPD-2013, SUEDL-2013, DataCur2013. A short description of each workshop is given here.

LCPD-2013: Linking and Contextualizing Publications and Datasets

Today's scientific communities are faced with data-driven requirements of e-science and new kinds of research methodologies and approaches inspired by e-research. In particular, these trends led to a new data-centric way of conceptualizing,

organizing, and carrying out research activities and, consequently, revolutionized scientific communication. Scientific communities and funding bodies are eagerly discussing and investigating the need for scientists to publish their raw datasets—e.g., experimental details, analytical methods, and visualizations—alongside scientific publications or using novel types of data publication. Data are becoming a first-class citizen of the modern scientific communication, being published for discovery and re-use together with literature to a greater extent. More specifically, linking and contextualizing publications and datasets in a meaningful way is increasingly becoming a key requirement not only for the scientists, but also for their organizations and ultimately for funding agencies. For example, dataset-publication linking allows researchers to better verify the quality of scientific outcomes, e.g., by reproducing the experiments, and greatly improves dataset availability, discoverability, interpretability, and re-usability. Moreover, contextualizing such information, by adding further interlinking with research funding information and affiliations, would allow funding agencies and organizations to measure their research impact in order to assess the quality of their investments and their activities.

Modern scientific communities and their research infrastructures serve for discipline-specific activities whose input and output affect and are affected by scientific communication. To cope with the increasing speed of such activities and the growing volumes of research outcome, communities need to be facilitated in publishing, interlinking, contextualizing, preserving, discovering, accessing, and reusing their research outcomes. Achieving such objectives would foster multi-disciplinarity, generate novel research opportunities, and endorse quality research. However, researchers rely on different technologies and systems to deposit and preserve their research outcome and their contextual information. Datasets and publications are kept into digital libraries and data centers together with descriptive metadata. Contextual information is scattered into other systems, for example, CRIS systems for funding schemes and affiliations, national and international initiatives and registries, such as ORCID and VIAF for people and authors. The construction of *modern scientific communication systems* capable of collecting and assembling such information in a meaningful way has opened up several research challenges in the fields of digital library, e-science, and e-research.

The goal of this workshop was to provide researchers and practitioners in the fields of digital library, e-science, and e-research with a forum where they could constructively explore these topics. Ten contributions and two invited papers report on theoretical, systemic, and foundational work targeting popular topics of linking and contextualizing datasets and publications. The quality of the contributions was ensured by a rich and qualified Program Committee.

SUEDL-2013: Supporting Users Exploration of Digital Libraries

There is a pressing need to better exploit valuable resources in cultural heritage digital libraries through improved access and support for users in information discovery and use. Current challenges in information access include: raising awareness and

discovery of the availability of collections; helping users to understand the overall content of a digital library and gain an overview of the collection as a starting point for information discovery; and, enabling more exploratory modes of information seeking, such as open-ended browsing and serendipity. In cultural heritage contexts, especially when domain and collection knowledge is limited (e.g., novice users, the general public, and students), the simple search box approach often delivers less than satisfactory results, and there is therefore a need to offer new ways of discovering and engaging with digital libraries in this environment.

Many different approaches to these challenges are being taken, bringing together knowledge and expertise from research areas as diverse as interactive information retrieval, human–computer interaction, data visualization, guided paths and trails, personalization and recommendation, NLP and content enrichment, among others. There is also broad acknowledgment that support for more complex tasks may require a more holistic approach to the design of digital library systems, with the provision of tools that extend user interaction into areas such as sensemaking and information (re)use. Furthermore, as solutions and results begin to emerge, there is a need to consider how to evaluate these novel systems and how to interpret results beyond the typical IR measures of precision and recall, including aspects of user experience such as exploration, task performance, user engagement, and satisfaction.

At this second SUEDL workshop, we once again aimed to bring together academic and practitioner perspectives, and to promote discussion and collaboration in addressing some of the issues outlined above, with a strong focus on users and solutions to support their needs. The workshop comprised five research papers and five demonstration papers, covering topics as diverse as human–computer interaction, digital library and archive evaluation techniques, linked data, image similarity, visualization and recommendations, in a variety of digital library contexts including national libraries, museum, archival and special collections, sound and vision, and oral history. The program of papers and demonstrations was enriched further by a keynote speech from Professor Ann Blandford of University College London, on "Exploring the Information Landscape: The Digital Library in Context", and a lively panel-led discussion on the topic of "Supporting Users' Exploration of Digital Libraries: Priorities and Future Directions."

DataCur2013: Moving Beyond Technology: iSchools and Education in Data Curation. Is Data Curator a New Role?

The increase of digital content in the broad areas of institutional and domain-specific repositories, libraries, archives, and museums and the increased interest in the sharing and preservation of research data have triggered the emergence of some buzzwords that more and more often appear in the literature, such as *convergence, digital curation, and data curator.*

Convergence, in this context, is related to the merging of the education curricula for information professionals in the disciplines of library science, archival, and museum studies, under the assumption that the digitization of the collections is blurring the traditional boundaries between those three professions.

Digital curation is a term generally used to indicate those activities that add value and knowledge to the collections, and the added value is usually given by the curator or manager of the cultural institution. The term digital curation is also used to describe the actions needed to maintain digital material, including digital research data, over their entire life-cycle and over time, for current and future generations of users. Assuming that there is a set of core competencies needed by a digital curator, it is yet to be understood how these competencies could at the same time play in favor of the convergence (given the digital nature of the resources to be curated) and in favor of a professional identity (given the different focus and mission of the three disciplines, where the value adding and the access to the collections would remain different).

Data curator is used here in the context of the storage, management, and preservation of digital research data, since more and more scientific research, in almost all disciplines, ends up being based on digital sources. There are a number of good reasons for preserving research data, but today it is not (yet) clear which of the existing professional roles are best suited for this activity. Should there be a data librarian, or a data archivist, or a data museum curator? Or is this a new role to be invented from scratch? Or should the responsibility of curating research data be given to the data producers, i.e., the researchers themselves?

In the context described above, the DataCur2013 main goal was to debate ideas and concrete examples of research projects, educational programs and training initiatives in digital curation and research data management, illustrating approaches, methodologies, and success stories addressing the need for an increasingly qualified *information workforce* in the data library, data center, archive, museum, and cultural heritage sectors. Specific objectives were to provide additional insight into the complex interplay between education, research, and curation (including long-term preservation) of digital data, and how these needs could be addressed in curricula for the education and training of the information professionals; and to gain a better understanding of the level at which a convergence of the three traditional professions could be achieved, contributing to a more global view of the access and preservation of research outcomes, which today are very often scattered in libraries, archives, institutional repositories, data bases, etc.

This workshop was the ideal continuation of a series of workshops and events that started in 2005 with a workshop on "Information Technologies Profiles and Curricula for Libraries" (held at the University of Parma), and has continued through five more events up to the last one in February 2013 with the workshop "iSchools Building on the Strengths Found in the Convergence of Librarianship, Archival, and Museum Studies to Improve the Education of Managing Digital Collections" (held in connection with the iConference 2013 at Fort Worth). A detailed descriptions of these events and an analysis of their results can be found in the volume edited by A.M. Tammaro,

V. Casarosa, and D. Castelli: *Closing the Gap: Interdisciplinary Perspectives on Research and Education for Digital Libraries* (IRCDL 2013, Rome, Italy), in the series *Communications in Computer and Information Science*, published by Springer.

September 2013

Łukasz Bolikowski
Vittore Casarosa
Paula Goodale
Nikos Houssos
Paolo Manghi
Jochen Schirrwagen

Organization

Committees

LCPD Organizing Committee

Paolo Manghi
Łukasz Bolikowski
Nikos Houssos
Jochen Schirrwagen

LCPD Program Committee

Eloy Rodrigues
Brigitte Mathiak
Stefan Kramer
Richard Cyganiak
Petr Knoth
Sarantos Kapidakis
Asunción Gómez-Pérez
Philipp Cimiano
Anna Clements
Dragan Ivanović
Jane Greenberg
Brigitte Joerg
Johanna McEntyre
Miguel-Angel Sicilia
Brian Matthews
Cezary Mazurek
Keith Jeffery

SUEDL Organizing Committee

Paula Goodale
Mark Stevenson
Eneko Agirre
Kate Fernie

SUEDL Program Committee

Maristella Agosti
Donatella Castelli
Paul Clough
Jillian Griffiths
Mark Hall
Gareth Jones
Jaap Kamps
Birger Larsen
Séamus Lawless
Johan Oomen
Arantxa Otegi
Juliane Stiller
Tinne Tuytelaars
Max Wilson

DataCur Organizing Committee

Vittore Casarosa
Donatella Castelli
Seamus Ross
Anna Maria Tammaro

DataCur Program Committee

Kevin Ashley
Christoph Becker
Toni Carbo
Paul Conway
Kate Fernie
Mariella Guercio
Carolyn Hank
Cal Lee
Jens Ludwig
John McDonald
Nancy McGovern
Mary Molinaro
Andy Rauber
Michael Seadle
Stefan Strathmann
Manfred Thaller
Terry Weech

Contents

First Workshop on Linking and Contextualizing Publications and Datasets

Moving Beyond Technology: iSchools and Education in Data Curation: Is Data Curator a New Role?