

Communications in Computer and Information Science

988

Commenced Publication in 2007

Founding and Former Series Editors:

Phoebe Chen, Alfredo Cuzzocrea, Xiaoyong Du, Orhun Kara, Ting Liu,
Dominik Ślęzak, and Xiaokang Yang

Editorial Board

Simone Diniz Junqueira Barbosa

*Pontifical Catholic University of Rio de Janeiro (PUC-Rio),
Rio de Janeiro, Brazil*

Joaquim Filipe

Polytechnic Institute of Setúbal, Setúbal, Portugal

Ashish Ghosh

Indian Statistical Institute, Kolkata, India

Igor Kotenko

*St. Petersburg Institute for Informatics and Automation of the Russian
Academy of Sciences, St. Petersburg, Russia*

Krishna M. Sivalingam

Indian Institute of Technology Madras, Chennai, India

Takashi Washio

Osaka University, Osaka, Japan

Junsong Yuan

University at Buffalo, The State University of New York, Buffalo, USA

Lizhu Zhou

Tsinghua University, Beijing, China

More information about this series at <http://www.springer.com/series/7899>

Paolo Manghi · Leonardo Candela ·
Gianmaria Silvello (Eds.)


Digital Libraries: Supporting Open Science

15th Italian Research Conference
on Digital Libraries, IRCDL 2019
Pisa, Italy, January 31 – February 1, 2019
Proceedings

Editors

Paolo Manghi 
Italian National Research Council
Pisa, Italy

Leonardo Candela 
Italian National Research Council
Pisa, Italy

Gianmaria Silvello 
University of Padua
Padua, Italy

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-030-11225-7 ISBN 978-3-030-11226-4 (eBook)
<https://doi.org/10.1007/978-3-030-11226-4>

Library of Congress Control Number: 2018967044

© Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The Italian Research Conference on Digital Libraries (IRCDL) is an annual forum for the Italian research community to discuss the research topics pertaining to digital libraries and related technical, practical, and social issues. Since 2005, it has served as a key meeting for the Italian digital library community. During these years, IRCDL has touched upon many of the facets underlying the term “digital library,” adapting the solicited research topics to the evolution of the issues in this domain and to the evolution of the whole process of scholarly communication. Today, the term digital library is associated with theory and practices that go well beyond its original meaning, *de facto* reflecting the evolution of the role of libraries in the scholarly communication domain, and of late embracing the latest questions and desiderata posed by open science.

The theme of the 2019 edition of the conference was “Digital Libraries: On Supporting Open Science.” This was motivated by several reasons. The results of research are no longer just scientific publications, of which libraries have always been the custodians. Science is increasingly and rapidly becoming digital, in the sense that more and more research is performed using data services and tools available online or on desktop computers, and the products and outcome of science are increasingly encompassing also datasets, software, and experiments. Being digital, such products can be shared and re-used together with the article, thus enabling comprehensive research assessment and various degrees of reproducibility of science. Positive consequences of this shift toward open science are: accelerating science, optimizing cost of research, fraud detection, and fully-fledged scientific reward.

Digital libraries are therefore facing new challenges of supporting the digital scientific process and becoming an important element in the evolution of research outputs. Firstly, by targeting deposition, findability, preservation, interlinking, and re-use of any kind of research product, ranging from publications, to research data, to software and others. Secondly, by becoming a pro-active and integrated component of the scholarly communication process as a whole; for example, by providing support to researchers in the preparation of datasets suitable for subsequent deposition, or by enriching and linking the deposited content with data from other sources, such as other repositories, ORCID, organization IDs, recommender systems, distributed annotations, aggregated statistics, etc. Finally, by integrating with research infrastructures and e-infrastructures to serve specific research community publishing needs or to benefit from economy-of-scale provision of storage and computing capacities.

This volume contains the revised accepted papers from among those presented at the 15th Italian Research Conference on Digital Libraries (IRCDL 2019), which was held at the Italian National Research Council, Research Area of Pisa (Italy) from January 31 to February 1, 2019. Contributions touched a rich array of topics ranging from citation, provenance, and curation of scientific datasets to interlinking of research products and novel peer review practices.

To conclude, we would like to express our gratitude to all our colleagues for submitting papers to the IRCDL and for helping to finalize this volume in due time. We would like to thank those institutions and individuals who made this conference possible: the Program Committee members and the Steering Committee members who took part in the evaluation of manuscripts and provided insights helping to shape a valuable conference program; the OpenAIRE-Advance project (European Union's Horizon 2020 research and innovation program, grant agreement No. 777541) for sponsoring the event; the Organizing Committee for professionally organizing the event.

November 2018

Paolo Manghi
Leonardo Candela
Gianmaria Silvello

Organization

Program Chairs

Paolo Manghi	Italian National Research Council - ISTI, Pisa, Italy
Leonardo Candela	Italian National Research Council - ISTI, Pisa, Italy
Gianmaria Silvello	University of Padua, Italy

IRCDL Steering Committee

Maristella Agosti	University of Padua, Italy
Tiziana Catarci	University of Rome La Sapienza, Italy
Alberto Del Bimbo	University of Florence, Italy
Floriana Esposito	University of Bari, Italy
Carlo Tasso	University of Udine, Italy
Costantino Thanos	Italian National Research Council - ISTI, Pisa, Italy

Organizing Committee

Catherine Bosio	Italian National Research Council - ISTI, Pisa, Italy
Miriam Baglioni	Italian National Research Council - ISTI, Pisa, Italy
Alessia Bardi	Italian National Research Council - ISTI, Pisa, Italy
Vittore Casarosa	Italian National Research Council - ISTI, Pisa, Italy
Emma Lazzeri	Italian National Research Council - ISTI, Pisa, Italy

Program Committee

Giovanni Adorni	University of Genoa, Italy
Miriam Baglioni	Italian National Research Council - ISTI, Pisa, Italy
Lamberto Ballan	University of Padua, Italy
Lorenzo Baraldi	University of Modena and Reggio Emilia, Italy
Alessia Bardi	Italian National Research Council - ISTI, Pisa, Italy
Valentina Bartalesi	Italian National Research Council - ISTI, Pisa, Italy
Andrea Bollini	4Science, Italy
Paolo Budroni	University of Vienna, Austria
Vittore Casarosa	Italian National Research Council - ISTI, Pisa, Italy
Michelangelo Ceci	University of Bari, Italy
Giorgio Maria Di Nunzio	University of Padua, Italy
Achille Felicetti	VASTLAB – PIN S.c.R.L., Italy
Stefano Ferilli	University of Bari, Italy
Nicola Ferro	University of Padua, Italy
Elena Giglia	University of Turin, Italy
Costantino Grana	University of Modena and Reggio Emilia, Italy

Petr Knoth	The Open University, UK
Claudio Lucchese	University of Venice, Italy
Donato Malerba	University of Bari, Italy
Stefano Mizzaro	University of Udine, Italy
Nicola Orio	University of Padua, Italy
Silvio Peroni	University of Bologna, Italy
Antonella Poggi	University of Rome La Sapienza, Italy
Eloy Rodrigues	University of Minho, Portugal
Alessandro Sarretta	Italian National Research Council - ISMAR, Venice, Italy
Marco Schaerf	University of Rome La Sapienza, Italy
Lorenzo Seidenari	University of Florence, Italy
Giuseppe Serra	University of Udine, Italy
Luigi Siciliano	Free University of Bozen-Bolzano, Italy
Annamaria Tammaro	University of Parma, Italy
Francesca Tomasi	University of Bologna, Italy
Fabio Vitali	University of Bologna, Italy

Additional Reviewers

D. Dosso	G. Pio
E. Fabris	D. Pride
D. Herrmannova	M. Soprano
P. Mignone	

Contents

Information Retrieval

- On Synergies Between Information Retrieval and Digital Libraries 3
Maristella Agosti, Erika Fabris, and Gianmaria Silvello
- Making Large Collections of Handwritten Material Easily Accessible
and Searchable 18
Anders Hast, Per Cullhed, Ekta Vats, and Matteo Abrate
- Transparency in Keyword Faceted Search: An Investigation
on Google Shopping 29
*Vittoria Cozza, Van Tien Hoang, Marinella Petrocchi,
and Rocco De Nicola*
- Predicting the Usability of the Dice CAPTCHA via Artificial
Neural Network 44
*Alessia Amelio, Radmila Janković, Dejan Tanikić,
and Ivo Rumenov Draganov*

Digital Libraries and Archives

- Water to the Thirsty Reflections on the Ethical Mission of Libraries
and Open Access 61
Matilde Fontanin and Paola Castellucci
- Computational Terminology in eHealth 72
Federica Vezzani and Giorgio Maria Di Nunzio
- Connecting Researchers to Data Repositories in the Earth, Space,
and Environmental Sciences 86
*Michael Witt, Shelley Stall, Ruth Duerr, Raymond Plante,
Martin Fenner, Robin Dasler, Patricia Cruse, Sophie Hou,
Robert Ulrich, and Danie Kinkade*
- Learning to Cite: Transfer Learning for Digital Archives 97
Dennis Dosso, Guido Setti, and Gianmaria Silvello
- Exploring Semantic Archival Collections: The Case of Piłsudski
Institute of America 107
Laura Pandolfo, Luca Pulina, and Marek Zieliński

Digital Libraries for Open Science: Using a Socio-Technical Interaction Network Approach 122
Jennifer E. Beamer

Information Integration

OpenAIRE’s DOIBoost - Boosting Crossref for Research. 133
Sandro La Bruzzo, Paolo Manghi, and Andrea Mannocci

Enriching Digital Libraries with Crowdsensed Data: Twitter Monitor and the SoBigData Ecosystem 144
Stefano Cresci, Salvatore Minutoli, Leonardo Nizzoli, Serena Tardelli, and Maurizio Tesconi

Populating Narratives Using Wikidata Events: An Initial Experiment. 159
Daniele Metilli, Valentina Bartalesi, Carlo Meghini, and Nicola Aloia

Metadata as Semantic Palimpsests: The Case of PHAIDRA@unipd. 167
Anna Bellotto and Cristiana Bettella

In Codice Ratio: Machine Transcription of Medieval Manuscripts 185
Serena Ammirati, Donatella Firmani, Marco Maiorino, Paolo Merialdo, and Elena Nieddu

Open Science

Foundations of a Framework for Peer-Reviewing the Research Flow 195
Alessia Bardi, Vittore Casarosa, and Paolo Manghi

A Practical Workflow for an Open Scientific Lifecycle Project: EcoNAOS. . . 209
Annalisa Minelli, Alessandro Sarretta, Alessandro Oggioni, Caterina Bergami, and Alessandra Pugnetti

Data Deposit in a CKAN Repository: A Dublin Core-Based Simplified Workflow 222
Yulia Karimova, João Aguiar Castro, and Cristina Ribeiro

Information Literacy Needs Open Access or: Open Access is not Only for Researchers 236
Maurizio Lana

The OpenUP Pilot on Research Data Sharing, Validation and Dissemination in Social Sciences 248
Daniela Luzi, Roberta Ruggieri, and Lucio Pisacane

Crowdsourcing Peer Review: As We May Do 259
Michael Soprano and Stefano Mizzaro

Hands-On Data Publishing with Researchers: Five Experiments
with Metadata in Multiple Domains. 274
*Joana Rodrigues, João Aguiar Castro, João Rocha da Silva,
and Cristina Ribeiro*

Data Mining

Towards a Process Mining Approach to Grammar Induction for Digital
Libraries: Syntax Checking and Style Analysis 291
Stefano Ferilli and Sergio Angelastro

Keyphrase Extraction via an Attentive Model 304
Marco Passon, Massimo Comuzzo, Giuseppe Serra, and Carlo Tasso

Semantically Aware Text Categorisation for Metadata Annotation. 315
*Giulio Carducci, Marco Leontino, Daniele P. Radicioni, Guido Bonino,
Enrico Pasini, and Paolo Tripodi*

Collecting and Controlling Distributed Research Information by Linking
to External Authority Data - A Case Study. 331
Atif Latif, Timo Borst, and Klaus Tochtermann

Interactive Text Analysis and Information Extraction 340
*Tasos Giannakopoulos, Yannis Foufoulas, Harry Dimitropoulos,
and Natalia Manola*

Author Index 351