

*Commenced Publication in 1973*

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

## Editorial Board

David Hutchison

*Lancaster University, Lancaster, UK*

Takeo Kanade

*Carnegie Mellon University, Pittsburgh, PA, USA*

Josef Kittler

*University of Surrey, Guildford, UK*

Jon M. Kleinberg

*Cornell University, Ithaca, NY, USA*

Alfred Kobsa

*University of California, Irvine, CA, USA*

Friedemann Mattern

*ETH Zurich, Zürich, Switzerland*

John C. Mitchell

*Stanford University, Stanford, CA, USA*

Moni Naor

*Weizmann Institute of Science, Rehovot, Israel*

Oscar Nierstrasz

*University of Bern, Bern, Switzerland*

C. Pandu Rangan

*Indian Institute of Technology, Madras, India*

Bernhard Steffen

*TU Dortmund University, Dortmund, Germany*

Demetri Terzopoulos

*University of California, Los Angeles, CA, USA*

Doug Tygar

*University of California, Berkeley, CA, USA*

Gerhard Weikum

*Max Planck Institute for Informatics, Saarbruecken, Germany*

More information about this series at <http://www.springer.com/series/7409>

Sören Auer · Volha Bryl  
Sebastian Tramp (Eds.)

# Linked Open Data – Creating Knowledge Out of Interlinked Data

Results of the LOD2 Project

*Editors*  
Sören Auer  
Institut für Informatik III  
Rheinische Friedrich-Wilhelms-Universität  
Bonn  
Germany

Sebastian Tramp  
University of Leipzig  
Leipzig  
Germany

Volha Bryl  
University of Mannheim  
Mannheim  
Germany

ISSN 0302-9743  
ISSN 978-3-319-09845-6  
DOI 10.1007/978-3-319-09846-3

ISSN 1611-3349 (electronic)  
ISSN 978-3-319-09846-3 (eBook)

Library of Congress Control Number: 2014945220

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

Springer Cham Heidelberg New York Dordrecht London

© The Editor(s) (if applicable) and the Author(s) 2014. The book is published with open access at [Springer-Link.com](http://Springer-Link.com)

**Open Access.** This book is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

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. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

This book presents an overview of the results of the research project ‘LOD2 – Creating Knowledge out of Interlinked Data’. LOD2 is a large-scale integrating project co-funded by the European Commission within the FP7 Information and Communication Technologies Work Programme (Grant Agreement No. 257943). Commencing in September 2010, this 4-year project comprised leading Linked Open Data research groups, companies, and service providers from across 11 European countries and South Korea.

Linked Open Data (LOD) is a pragmatic approach for realizing the Semantic Web vision of making the Web a global, distributed, semantics-based information system. The aim of the LOD2 project was to advance the state of the art in research and development in four key areas relevant for Linked Data, namely 1. RDF data management; 2. the extraction, creation, and enrichment of structured RDF data; 3. the interlinking and fusion of Linked Data from different sources; and 4. the authoring, exploration, and visualization of Linked Data. The results the project has attained in these areas are discussed in the technology part of this volume, i.e., chapters 2–6. The project also targeted use cases in the publishing, linked enterprise data, and open government data realms, which are discussed in chapters 7–10 in the second part. The book gives an overview of a diverse number of research, technology, and application advances and refers the reader to further detailed technical information in the project deliverables and original publications. In that regard, the book is targeted at IT professionals, practitioners, and researchers aiming to gain an overview of some key aspects of the emerging field of Linked Data.

During the lifetime of the LOD2 project, Linked Data technology matured significantly. With regard to RDF and Linked Data management, the performance gap compared with relational data management was almost closed. Automatic linking, extraction, mapping, and visualization of RDF data became mainstream technology provided by mature open-source software components. Standards such as the R2RML RDB2RDF mapping language were defined and a vast number of small and large Linked Data resources (including DBpedia, LinkedGeoData, or the 10.000 publicdata.eu datasets) amounting to over 50 Billion triples are now available. The LOD2 project has driven and actively contributed to many of these activities. As a result, Linked Data is now ready to enter the commercial and large-scale application stage, as many commercial products and services (including the ones offered by the industrial LOD2 project partners) demonstrate.

In addition to the LOD2 project partners, who are authors and contributors of the individual chapters of this book, the project was critically accompanied and supported by a number of independent advisers and mentors including Stefano Bertolo (European Commission), Stefano Mazzocchi (Google), Jarred McGinnis (Logomachy), Atanas Kiryakov (Ontotext), SteveHarris (Aistemos), and Márta Nagy-Rothengass (European Commission). Furthermore, a large number of stakeholders engaged with the LOD2

project, for example, through the LOD2 PUBLINK initiatives, the regular LOD2 technology webinars, or the various events organized by the project. We are grateful for their support and feedback, without which the project as well as this book would not have been possible.

July 2014

Sören Auer  
Volha Bryl  
Sebastian Tramp

# Contents

1	Introduction to LOD2 . . . . .	1
	<i>Sören Auer</i>	
<b>Technology</b>		
2	Advances in Large-Scale RDF Data Management. . . . .	21
	<i>Peter Boncz, Orri Erling, and Minh-Duc Pham</i>	
3	Knowledge Base Creation, Enrichment and Repair . . . . .	45
	<i>Sebastian Hellmann, Volha Bryl, Lorenz Bühmann, Milan Dojchinovski, Dimitris Kontokostas, Jens Lehmann, Uroš Milošević, Petar Petrovski, Vojtěch Svátek, Mladen Stanojević, and Ondřej Zamazal</i>	
4	Interlinking and Knowledge Fusion . . . . .	70
	<i>Volha Bryl, Christian Bizer, Robert Isele, Mateja Verlic, Soon Gill Hong, Sammy Jang, Mun Yong Yi, and Key-Sun Choi</i>	
5	Facilitating the Exploration and Visualization of Linked Data . . . . .	90
	<i>Christian Mader, Michael Martin, and Claus Stadler</i>	
6	Supporting the Linked Data Life Cycle Using an Integrated Tool Stack . . .	108
	<i>Bert Van Nuffelen, Valentina Janev, Michael Martin, Vuk Mijovic, and Sebastian Tramp</i>	
<b>Use Cases</b>		
7	LOD2 for Media and Publishing . . . . .	133
	<i>Christian Dirschl, Tassilo Pellegrini, Helmut Nagy, Katja Eck, Bert Van Nuffelen, and Ivan Ermilov</i>	
8	Building Enterprise Ready Applications Using Linked Open Data . . . . .	155
	<i>Amar-Djalil Mezaour, Bert Van Nuffelen, and Christian Blaschke</i>	
9	Lifting Open Data Portals to the Data Web . . . . .	175
	<i>Sander van der Waal, Krzysztof Węcel, Ivan Ermilov, Valentina Janev, Uroš Milošević, and Mark Wainwright</i>	
10	Linked Open Data for Public Procurement. . . . .	196
	<i>Vojtěch Svátek, Jindřich Mynarz, Krzysztof Węcel, Jakub Klimek, Tomáš Knap, and Martin Nečaský</i>	
	<b>Author Index</b> . . . . .	215