Library Linked Data in the Cloud

OCLC's Experiments with New Models of Resource Description

Synthesis Lectures on the Semantic Web: Theory and Technology

Editors

Ying Ding, ***Indiana University***
Paul Groth, ***Elsevier Labs***

Founding Editor Emeritus

James Hendler, ***Rensselaer Polytechnic Institute***

Synthesis Lectures on the Semantic Web: Theory and Application is edited by Ying Ding of Indiana University and Paul Groth of VU University Amsterdam. Whether you call it the Semantic Web, Linked Data, or Web 3.0, a new generation of Web technologies is offering major advances in the evolution of the World Wide Web. As the first generation of this technology transitions out of the laboratory, new research is exploring how the growing Web of Data will change our world. While topics such as ontology-building and logics remain vital, new areas such as the use of semantics in Web search, the linking and use of open data on the Web, and future applications that will be supported by these technologies are becoming important research areas in their own right. Whether they be scientists, engineers or practitioners, Web users increasingly need to understand not just the new technologies of the Semantic Web, but to understand the principles by which those technologies work, and the best practices for assembling systems that integrate the different languages, resources, and functionalities that will be important in keeping the Web the rapidly expanding, and constantly changing, information space that has changed our lives.
Topics to be included:

- Semantic Web Principles from linked-data to ontology design
- Key Semantic Web technologies and algorithms
- Semantic Search and language technologies
- The Emerging ”Web of Data” and its use in industry, government and university applications
- Trust, Social networking and collaboration technologies for the Semantic Web
- The economics of Semantic Web application adoption and use
- Publishing and Science on the Semantic Web
- Semantic Web in health care and life sciences

Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resource Description
Carol Jean Godby, Shenghui Wang, and Jeffrey K. Mixter
2015

Semantic Breakthrough in Drug Discovery
Bin Chen, Huijun Wang, Ying Ding, and David Wild
2014

Semantics in Mobile Sensing
Zhixian Yan and Dipanjan Chakraborty
2014

Provenance: An Introduction to PROV
Luc Moreau and Paul Groth
2013

Resource-Oriented Architecture Patterns for Webs of Data
Brian Sletten
2013

Aaron Swartz's A Programmable Web: An Unfinished Work
Aaron Swartz
2013

Incentive-Centric Semantic Web Application Engineering
Elena Simperl, Roberta Cuel, and Martin Stein
2013

Publishing and Using Cultural Heritage Linked Data on the Semantic Web
Eero Hyvönen
2012

VIVO: A Semantic Approach to Scholarly Networking and Discovery
Katy Börner, Michael Conlon, Jon Corson-Rikert, and Ying Ding
2012

Linked Data: Evolving the Web into a Global Data Space
Tom Heath and Christian Bizer
2011

© Springer Nature Switzerland AG 2022
Reprint of original edition © Morgan & Claypool 2015

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resource Description

Carol Jean Godby, Shenghui Wang, and Jeffrey K. Mixter

ISBN: 978-3-031-79464-3 paperback
ISBN: 978-3-031-79465-0 ebook

DOI 10.1007/978-3-031-79465-0

A Publication in the Springer series
SYNTHESIS LECTURES ON THE SEMANTIC WEB: THEORY AND TECHNOLOGY

Lecture #9
Series Editors: Ying Ding, *Indiana University*
Paul Groth, *Elsevier Labs*
Founding Editor Emeritus: James Hendler, *Rensselaer Polytechnic Institute*
Series ISSN
Print 2160-4711 Electronic 2160-472X

Library Linked Data in the Cloud

OCLC's Experiments with New Models of Resource Description

Carol Jean Godby, Shenghui Wang, and Jeffrey K. Mixter
OCLC Research

SYNTHESIS LECTURES ON THE SEMANTIC WEB: THEORY AND TECHNOLOGY #9

ABSTRACT

This book describes OCLC's contributions to the transformation of the Internet from a web of documents to a Web of Data. The new Web is a growing 'cloud' of interconnected resources that identify the things people want to know about when they approach the Internet with an information need.

The linked data architecture has achieved critical mass just as it has become clear that library standards for resource description are nearing obsolescence. Working for the world's largest library cooperative, OCLC researchers have been active participants in the development of next-generation standards for library resource description. By engaging with an international community of library and Web standards experts, they have published some of the most widely used RDF datasets representing library collections and librarianship.

This book focuses on the conceptual and technical challenges involved in publishing linked data derived from traditional library metadata. This transformation is a high priority because most searches for information start not in the library, nor even in a Web-accessible library catalog, but elsewhere on the Internet. Modeling data in a form that the broader Web understands will project the value of libraries into the Digital Information Age.

The exposition is aimed at librarians, archivists, computer scientists, and other professionals interested in modeling bibliographic descriptions as linked data. It aims to achieve a balanced treatment of theory, technical detail, and practical application.

KEYWORDS

Library, Semantic Web, Library metadata, Resource description, Ontology development, Schema.org

Contents

Preface

OCLC is a nonprofit library cooperative providing research, programs, and services that help libraries share the world's knowledge. OCLC manages WorldCat, the largest and most comprehensive catalog of library resources from around the world. In the time period covered by this book, WorldCat contained more than 300 million bibliographic records that represented more than 2 billion items held by participating libraries. OCLC is also the custodian of the Dewey Decimal Classification, which has been used by libraries for over a hundred years to organize their collections. In addition, OCLC hosts the Virtual International Authority File, or VIAF, the largest aggregation of authoritative information collected by libraries about people and organizations and the creative works they have published. Resources such as these make it easier for libraries to fulfill their public mission of connecting library patrons to the works that satisfy their quest for information.

OCLC Research was founded in 1978 and has made significant contributions to the development of the library and Web standards that form the conceptual underpinning of WorldCat and other library resources. In the mid-1990s, OCLC researchers began making the case that libraries need to be integrated into the Web, because that's where the information seekers are. As Semantic Web technologies have matured, this argument has only become more urgent.

OCLC researchers have participated in the Library Linked Data Incubator Group sponsored by the World Wide Web Consortium. They have been vocal members of the Schema Bib Extend Community Group, which recommends extensions to Schema.org, the indexing vocabulary recommended by Google, Yahoo, Bing, and Yandex. OCLC researchers have also worked with the Wikipedia community to facilitate the cross-directional linking between library resources and Wikipedia articles. The results guide human readers from Wikipedia to libraries and enable machine processes to consume richer linked data through Wikipedia's association with the Wikidata project. In addition, OCLC researchers have served as advisers to the BIBFRAME standard sponsored by the Library of Congress, whose goal is to replace MARC, the legacy standard for bibliographic description, with a linked data model. And in the past five years, OCLC has become a significant publisher of linked data, producing models and publicly accessible datasets containing billions of RDF triples describing the objects and concepts referenced in VIAF; the Dewey Decimal Classification, or DDC; Faceted Application of Subject Headings, or FAST; and the WorldCat catalog. It is from this rich experience that this volume emerges.

This book is about OCLC's experiments in the redesign of traditional library resource descriptions as linked data. In practical terms, the goal of the work reported in this book is to define the first draft of an entity-relationship model of creative works and the events in the library community that impact them. The model is realized by mining the data stores maintained at OCLC and republishing them as large RDF datasets. Though the work is necessarily anchored to a particular point in time, we hope that readers will gain insight into the collective thinking of the world's largest library cooperative, whose solutions will spur development by others who might benefit from our trials and errors as well as our successes. In a program whose goal is to express library metadata as linked data, we are doing work that is consistent with the core values of our profession, which places a premium on collaboration and openness. In return, we are confident that the Web of Data will be enriched by the collective expertise of over a hundred years of librarianship.

The impetus for this book arose from a 2012 conversation between Lorcan Dempsey, OCLC Vice President of Research and Chief Strategist, and Ying Ding, Associate Professor of Information Science at Indiana University. The outcome was an invitation to propose a monograph for the series *Synthesis Lectures on the Semantic Web: Theory and Technology* published by Morgan and Claypool. Once the proposal was accepted, Jean Godby, Senior Research Scientist with OCLC Research, was tasked with organizing contributions from colleagues and contributing to the volume herself. Beginning in 2013 and stretching into the first half of 2014, material was contributed, refined, and in some cases re-written as work in progress at OCLC as researchers and their technical allies moved forward. In describing OCLC's projects, the aim is to tell a story about a large collection of interconnected projects. Each chapter is designed as a lecture on a problem that must be addressed if the enterprise of transforming library data to a format that is more effective at fulfilling the needs of the information-seeking public is to succeed.

Many OCLC colleagues have contributed intellectual content in addition to the three authors of this book: Lorcan Dempsey, Jonathan Fausey, Ted Fons, Janifer Gatenby, Thom Hickey, Maximilian Klein, Michael Panzer, Tod Matola, Ed O'Neill, Stephan Schindehette, Jenny Toves, Diane Vizine-Goetz, Richard Wallis, and Jeff Young. The authors are especially indebted to Karen Smith-Yoshimura, whose own important contributions to research on library metadata and whose thoughtful comments on the entire manuscript produced so many improvements that we realize, in hindsight, that she should have been a co-author. We are also grateful to OCLC colleagues who helped us with editorial and production tasks, including Eric Childress, Chris Galvin, Brad Gauder, Jenny Johnson, Jeanette McNichol, and JD Shipengrover.

In addition, we have benefited from engagement with colleagues in the library community, who mentored us, commented on the manuscript, tested some of our ideas at their own institutions, and joined with us in lengthy and often passionate discussion that produced many photos of whiteboards, some of which have found their way into the illustrations in this book. In particular, we are grateful to Kenning Arlitsch, Montana State University; Ray Denenberg, Library of Congress; Ying Ding, Indiana University; Kevin Ford, formerly of the Library of Congress; Paul Groth, Vrije Universiteit Amsterdam; Antoine Isaac,Vrije Universiteit Amsterdam; Nannette Naught, IMT Associates; Patrick OBrien, Montana State University; Philip Schreur, Stanford University; and Marcia Zeng, Kent State University. And of course, we are deeply indebted to our former OCLC colleagues Eric Miller and Stu Weibel, without whose groundbreaking work this enterprise might never have taken shape.

Carol Jean Godby, Shenghui Wang, and Jeffrey K. Mixter
March 2015