Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resource Description. Series: Synthesis Lectures on the Semantic Web: Theory and Technology (Book 9) By Carol Jean Godby, Shenghui Wang, Jeffrey K. Mixter**, Morgan & Claypool Publishers ©2015, 240p. ISBN: 1627052194, 9781627052191

This book is the ninth book of the series Synthesis Lectures on the Semantic Web: Theory and Technology, published by Morgan & Claypool. It addresses the challenge all libraries as a whole have been facing: exposing library materials to multiple search engines and having the presence of library metadata on the Web.

This book describes OCLC's experiments in linked data technology. OCLC's efforts are enormous and time-consuming, involving various processes such as concept designing, data modeling, metadata mapping, computer programming, and format conversion. Although the series of the Synthesis Lectures cover more topics surrounding semantic web and library linked data, this book provides a relatively comprehensive overview of how to publish traditional library metadata to the cloud. It also gives conceptual and technical details with concrete examples to explain these experiments. As transforming libraries' data into linked data goes beyond the library field, some chapters, terms, and explanation may present a challenge to readers. Without the basic understanding on concepts and topics such as Schema markup, FRBR (Functional Requirements for Bibliographic Records), and library authority files, some readers may have a certain degree of difficulty weaving different pieces of information together to achieve a full understanding of these experiments. The long list of the resources contained in the bibliography section is very useful for those new to library linked data to grasp basic knowledge and for those who have a strong interest in this topic to pursue further understanding. Majority of the bibliographies come with a URL, which is convenient for access.

This book is organized into five chapters. The first chapter begins with an introduction of the difference between the Web of documents and the Semantic Web, which illustrates the fundamental reason for the worldwide linked data movement. Several sections provide a broad view of the milestones related to library linked data from 1995 to 2014. They briefly cover these core library linked data achievements (e.g., FRBR, RDA-Resource Description and Access, and WorldCat Linked Data) that were made through the collaboration among Library of Congress, OCLC, and other library communities. Meanwhile, this chapter concisely discusses the procedure of OCLC's experiments on converting raw bibliographic records in WorldCat to RDF markups. This procedure involves mapping MARC records to the Schema.org standard, creating corresponding RDF turtles, applying FAST (Faceted Application of Subject Terminology) and VIAF (Virtual International Authority File) IDs, as well as assigning cool URIs (Uniform Resource Identifier). The figures and tables in this chapter are particularly helpful for readers to understand activities involved in OCLC's linked data experiments. This chapter does a good job of providing necessary

background information and basic knowledge that can help readers delve into the following chapters.

Chapter two describes modeling library authority files (e.g., Library Congress Subject Headings, FAST, Dewey Decimal Classification, and VIAF) as linked data. It discusses the process of mapping these files in SKOS (Simple Knowledge Organization System), FOAF (Friend of a Friend), and FAST as well as designing/creating the URIs to be put in a set of RDF datasets for these authority files. The authors also point out the strength and limitation of modeling using different standards. For example, SKOS is not rich enough to model persons while FOAF is a better fit.   Additionally, this chapter introduces the VIAF project and describes how to model these authority files as RDF turtles using Schema.org, SKOS, etc. Examining the development of modeling these files calls for the need to evolve VIAF into an authoritative hub of data.

Chapter three provides how to model and discover the creative works library communities have generated with MARC for years in depth. Starting from the review of treating creative works as the entities of objects and explaining the relationship with other three entities (i.e., expressions, manifestations, and items) defined by FRBR Group 1 Conceptual Model, several sections detail modeling FRBR concepts with Schema.org in conjunction with a set of bibliography vocabulary, designing the syntax and semantics of URLs in RDF descriptions, and other modeling efforts of creative works. Furthermore, the chapter succinctly reviews the data mining algorithms that are used to discover creative works in WorldCat. This chapter demonstrates the complexity of modeling and discovering creative works.

Text mining plays an important role in processing unstructured (free-text) information to generate meaningful information and knowledge. Chapter four first analyzes the need to apply text mining algorithms to library linked data projects. Then it discusses the algorithms OCLC used to promote strings in MARC records to entities. These algorithms can identify whether strings are names with identifiers (e.g., appearing in a controlled MARC field), labeled names (not subject to authority control), names in semi-structured text (e.g., an illustrator name in MARC 600 filed), or names in unstructured text (e.g., information contained in MARC 520 field). The chapter also examines subject/concept matching and explains the methods (e.g., thesaurus-matching and instance-based mapping) that can algorithmically map these concepts to authority files and vocabularies in the domain of library resource description. Lastly, the chapter introduces the document clustering which is important to handle a large amount of data. It also analyzes the pros and cons of different clustering algorithms and points out the future need to handle digital objects and cultural heritage materials. This chapter may be particularly interesting for library professionals with some computer programming background.

Chapter five briefly summarizes the experiments OCLC has taken and indicates future directions. Some directions OCLC is looking ahead include closely collaborating with the Schema.org group as it provides a set of robust Semantic Web standards and with other library communities such as Library of Congress as it is also experimenting with BIBFRAME (Bibliographic Framework) and RDA. This chapter shares the lessons OCLC has learned along the way and the action it takes such as participating in community standard initiatives and forming joint forces. In addition, this chapter discusses in details OCLC's challenges in three regards (i.e.,

conceptual, technical, and environmental).

Generally speaking, the detail-oriented description throughout the book is instrumental for those pursuing or will pursue a library data project. This book is a good reference book for those implementing a library linked data project. The challenge of comprehending this book can be overcome by reading more than once and performing the relevant reading. This book clearly demonstrates the efforts and the contributions OCLC has made through its various experiments to transform legacy library metadata and increase the visibility and usability of libraries' creative works.

Overall, this book is a valuable addition to the literature of library linked data. It can spark an interest and provide an eye-opening experience for beginners while offering intriguing thinking and processing specifics for some readers to understand details involved in the library linked data projects.

--- Xiaocan (Lucy) Wang, Emerging Technologies Librarian/Associate Professor, Spiva Library, Missouri Southern State University.