Second Edition

Covers **ONIX 3.0** & **EPUB 3**

The METADATA Handbook

A Book Publisher's Guide to Creating and Distributing Metadata for Print and Ebooks

Renée Register & Thad McIlroy

a DATACURATE publication

The Metadata Handbook

A Book Publisher's Guide to Creating and Distributing Metadata for Print and Ebooks

Renée Register & Thad McIlroy

Second Edition

DATACURATE

Columbus, Ohio
2015

Table of Contents

Introduction x

Part 1:
Book Industry Players and the Metadata Lifecycle

Part 1 explores the path and evolution of metadata through publishing and bookselling activities and channels and describes the players that create, distribute, enhance, and use metadata.

CHAPTER 1 1
The Metadata Lifecycle

CHAPTER 2 4
The Players

- Publishers
  - Trade Publishers
  - Professional and Scholarly Publishers
  - Educational Publishers
  - Small and Independent Publishers
  - Self-Publishing
  - Digital Publishing
- Book Data Aggregators
- Wholesalers, Distributors, and Ebook Aggregators
- Retailers
  - Independent Bookstores
  - Chain Bookstores
  - Online Retailers
- Libraries
- Metadata Management Vendors

Part 2:
Book Industry Metadata Standards

Part 2 outlines the history and development of book identifiers and book metadata standards, the organizations that administer the standards, and the role of standards in facilitating efficient data exchange in support of e-commerce.

CHAPTER 3 17
A Brief History of Metadata Standards and Practices

CHAPTER 4 22
Book Industry Standards for Sharing Metadata

- ISBN and Other Identifiers
  - ISBN, ISNI, ISTC, SAN, LCCN
- Metadata Schemes
  - ONIX for Books
  - Other Schemes
    - MARC, Dublin Core, Schema.org
- Standards for Subjects
  - Book Industry Standards
    - BISAC Subject Headings
    - BIC Standard Subject Categories
    - Thema Category Scheme
  - Library Standards
    - Library of Congress Subject Headings
    - Library of Congress Classification
    - Dewey Decimal Classification

CHAPTER 5 33
National and International Standards Organizations

- ISO and NISO
- The International ISBN Agency and National ISBN Registry Agencies
- EDItEUR and the ONIX Standard

Part 3:
Essential Metadata Elements

Part 3 describes types of book metadata and the essential information about a book needed to fully support description, discovery, selection, and commerce needs. This section also explores the ways in which various metadata elements are used behind the scenes to drive search engines and machine-based decisions on what books to display to which readers.

CHAPTER 6 39
Essential Elements for All Books

- Metadata for Content Description
- Metadata for Product Description
- Metadata for Commerce

CHAPTER 7 43
Enhancing Metadata to Stand Out in the Marketplace

CHAPTER 8 49
Search Engine Optimization, Keywords, and Subjects

- SEO and Keywords
- Keywords and BISAC Subject Headings
- Bookseller Website Search Engines
- General Search Engines

CHAPTER 9 54
Optimizing Product Metadata for Digital Publishing

- The Essential Elements
  - Content Description
  - Product Description
  - Commerce
- Ebooks and ISBNs
- Expanded Options for Digital Content in ONIX 3.0

CHAPTER 10 59
Metadata in Digital Content Files

- Uses of Metadata in Digital Content Files
- Metadata in the EPUB File Format
- Metadata in Proprietary Digital Formats

CHAPTER 11 49
Metadata for Self-Publishers and Small Publishers

- Identifiers: ISBN, Proprietary Identifier Systems, and Industry Databases
  - Obtaining an ISBN through ISBN Registration Agencies
  - Metadata in the ISBN Registration Process
  - ISBNs Provided by Self-Publishing Services
- Product Metadata Options in Self-Publishing Services
- Digital File Metadata Options in Self-Publishing Services
- Ebook Distribution Options in Self-Publishing Services

Part 4:
Metadata Best Practices and Certification Programs

Part 4 discusses certification and accreditation programs administered by national publishing industry organizations. These programs evaluate and rate publisher metadata for compliance with industry standards, recommended core elements, and best practices.

CHAPTER 12 62
Metadata Best Practices and Industry Organizations

- Best Practices and Industry Organizations
  - Standards
  - Best Practices
- The Role of Industry Organizations in Defining Best Practices

CHAPTER 13 65
BISG's Best Practices for Product Metadata: Guide for North American Data Senders and Receivers

- History of BISG Best Practices Documentation
- Intended Audience
- Structure and Format of *Best Practices of Product Metadata*
- A Closer Look at Best Practices for One Metadata Element
- Integrating Best Practices into Organizational Workflows

CHAPTER 14 69
Metadata Certification Programs

- What is Certification?
- The BISG Product Certification Program (PDCP)
- The Evaluation Process
- Fields Reviewed by the PDCP Certification Panel
- Product Data Certification in Other Countries
- Levels of Certification

Part 5:
Metadata and the Future of Publishing

Part 5 discusses how metadata standards and practices are evolving in response to the increased importance of digital publishing, the rise of self-publishing, shifts in the roles of publishing industry players, and changes in reader expectations. In conclusion, the authors share their thoughts on the meaning of these trends for the future of publishing and the role of metadata.

CHAPTER 15 73
Trends in Book Publishing and Metadata

CHAPTER 16 79
Thoughts on the Future of Publishing and Metadata

Glossary 85

Bibliography & References 105

Industry Organizations 111

Vendor Directory 118

About the Authors 127

Introduction

The Metadata Handbook *is for book publishers, including self-publishers and small publishers, who wish to optimize the dissemination of complete, compelling, and accurate information about their books throughout the publishing supply chain. This* ***Handbook*** *describes the metadata that accompanies each new book through the many activities necessary to propel it into the marketplace, to support discovery, and to increase sales potential.*

The ability to describe products completely and accurately and to efficiently share that data with trading partners and consumers is essential for e-commerce in every industry. In this respect book metadata is unexceptional. But every product and every industry has unique aspects that its product metadata attempts to reflect.

It's possible to sell a book online with just a few pieces of information, such as ISBN, title, and price. But providing minimal metadata limits opportunities for a book to be discovered, create a buzz, or make a sale. Success in the virtual marketplace requires the distribution of book metadata that fully supports findability and discovery, provides information that engages the potential reader, and contains all the information needed to support business transactions and business intelligence. Metadata must also meet the needs of booksellers and other trading partners. Many major receivers of publisher metadata require that certain metadata elements are present and correctly formatted and may reject metadata that fails to meet standards, resulting in lost sales.

For digital books, metadata entirely takes the place of all the physical organization, display, and browsing opportunities possible in bricks-and-mortar stores. For readers who prefer to buy books online, this is true for physical books as well.

Metadata that is thoughtfully created and shared using industry standards and best practices supports all aspects of publishing and bookselling. And when creating, controlling, and monitoring book metadata is fully integrated into each publication stage, the result is a powerful asset for effective selling across traditional and evolving sales channels as well as for the collection and analysis of sales results and market trends.

With the growth of digital content and publishing platforms, selling options have expanded from the traditional wholesalers and retailers to include sellers specializing in ebooks and easier options for selling directly from author or publisher websites. Digital formats are also more closely tied to their respective distribution channels (Amazon's proprietary KF8 format for Kindle, EPUB for Nook, etc.) than traditional hardcover, trade, and mass-market paperback formats ever were, so it's crucial that metadata enable users to select a format compatible with their reading device. To maximize discovery and sales potential, metadata should be complete, correctly formatted, and consistent for all formats and across all sales channels.

While metadata is a science of great breadth and is relevant to many industries and endeavors, this book focuses specifically on metadata used to describe, market, and sell books throughout the publishing supply chain. Rather than a how-to book, the *Handbook* is intended to be an industry reference. The reader can turn to it as the authoritative source on best practices in the day-to-day use of metadata in book publishing. The central goal of the *Handbook* is to describe and provide references to information about book metadata standards and best practices that are presently endorsed in the United States, the United Kingdom, and Canada.

The authors bring significant experience in U.S. and Canadian publishing-related activities but do not claim first-hand experience within the UK market. We have relied in part on the resources issued by major publishing industry organizations in each of the three countries.

The *Handbook* does not specifically cover any of the other countries or regions that publish English-language books, including Australia, South Africa, The Republic of Ireland, India, or any other country where English is not the first language. We are interested in hearing from parties who would like to adapt the information in this book for their home market(s).

The Metadata Handbook does not duplicate readily available documentation but presents an overview of metadata's role in the publisher supply chain. It collects and consolidates information about industry guidelines and resources into one volume for use as a reference in the practice of book data management.

New in This Edition

This edition is fully updated and expanded to include the most recent information on metadata standards, practices, and use in the publishing industry.

Part 1, *Book Industry Players and the Metadata Lifecycle* is revised.

Part 2, *Book Industry Metadata Standards* is revised throughout for consistency with the latest versions of industry standards and best practices documentation.

The *Identifiers* section of Part 2, Chapter 4, *Book Industry Standards for Sharing Metadata*, includes expanded and updated information on International Standard Name Identifier (ISNI) and International Standard Text Code (ISTC).

A new section on *Metadata Schemes* is added to Chapter 4. Information on ONIX is incorporated into this section and it also includes new subsections covering *Dublin Core* and *scema.org*.

A new section on *Standards for Subjects* was added and includes information on Thema as well as information about BISAC, BIC, and library standards.

Part 3, *Essential Metadata Elements*, is revised throughout for consistency with industry recommendations included in latest edition of Book Industry Study Group (BISG) best practices documentation, *Best Practices for Product Metadata: Guide for North American Senders and Receivers*, published in 2014.

Three new chapters are added to Part 3.

- A new chapter on *Search Engine Optimization, Keywords, and Subjects* (Chapter 8)
- A new chapter called *Metadata in Digital Content Files* outlines options for metadata embedded within digital content files (Chapter 10)
- A new chapter specifically addressing *Metadata for Self-Publishers and Small Publishers* (Chapter 11)

Part 4, *Metadata Best Practices and Industry Certification Programs*, is fully revised to provide an overview and guide to using Best Practices for Product Metadata: Guide for North American Senders and Receivers, published in 2014 by the Book Industry Study Group (BISG) in coordination with BookNet Canada.

Part 5, *Metadata and the Future of Publishing*, includes updates from industry experts on trends in metadata and publishing.

The ***Bibliography and References*** section is revised and expanded to include recent publications and revisions of existing documentation.

New entries are added to the ***Glossary*** and some entries are revised to reflect the most recent information.

Part 1
Book Industry Players and the Metadata Lifecycle

In Part 1, we outline the role and lifecycle of metadata in major publishing activities and sales channels. "Players" covered include:

- Publishers, including trade, professional, educational, independent, and self-publishers
- Book metadata aggregators, such as Bowker and Nielsen
- Retailers, including independent booksellers, chain bookstores, and online retailers
- Digital-only ebook aggregators such as OverDrive and ebrary
- Library metadata aggregators and library metadata uses
- Vendors specializing in products and services supporting book metadata management and distribution activities

We discuss how and why metadata changes as it flows through the publishing process and then through various reseller systems, platforms, and websites.

Part 2
Book Industry Metadata Standards

Part 2 covers existing and emerging standards important for bookselling in the 21st century. It offers a brief history of metadata standards and practices, introduces standards commonly used within the publishing industry, and discusses the national and international organizations and registries that administer and support them.

As with most major industries, data standards evolved to support the consistent flow of information in electronic form across the supply chain. The need for identifiers and data transmission standards became acute as computer technology became an important part of trade. Early numbering systems in the 1960s grew into the first International Standard Book Number (ISBN) system; the need for machine-to-machine transmission of business data between trading partners led to the development of Electronic Data Interchange (EDI) standards; and the explosive growth of the internet and online retailing required XML-based communication of robust book metadata, resulting in the development of ONIX (Online Information Exchange) for Books as the international standard for sharing book industry product information.

Part 3
Essential Metadata Elements

The ONIX for Books Product Information Message defines and supports hundreds of data fields carrying information about a publication. Publishers are not required to use all the fields, and not all the fields are applicable to every book. Part 3 lists and describes metadata elements considered essential for bookselling and explains why they're important.

Some of the ways metadata elements function in an online selling environment are not immediately apparent by viewing a title record display on a website. Part 3 explores how these essential elements are used behind the scenes in addition to their more obvious use in direct title search.

Part 3 also includes chapters providing information specific to digital publishing and information about metadata for self-publishers and small publishers.

Part 4

Metadata Best Practices and Certification Programs

Part 4 includes an overview of industry-recommended metadata elements, guidelines for applying these elements, and rules for using the ONIX standard. The *Handbook* collects basic information about the organizations and their recommended best practices into one reference work.

Book industry organizations administer programs designed to help publishers ensure compliance with industry standards and practices. Part 4 describes how these programs work and why publishers may wish to seek accreditation.

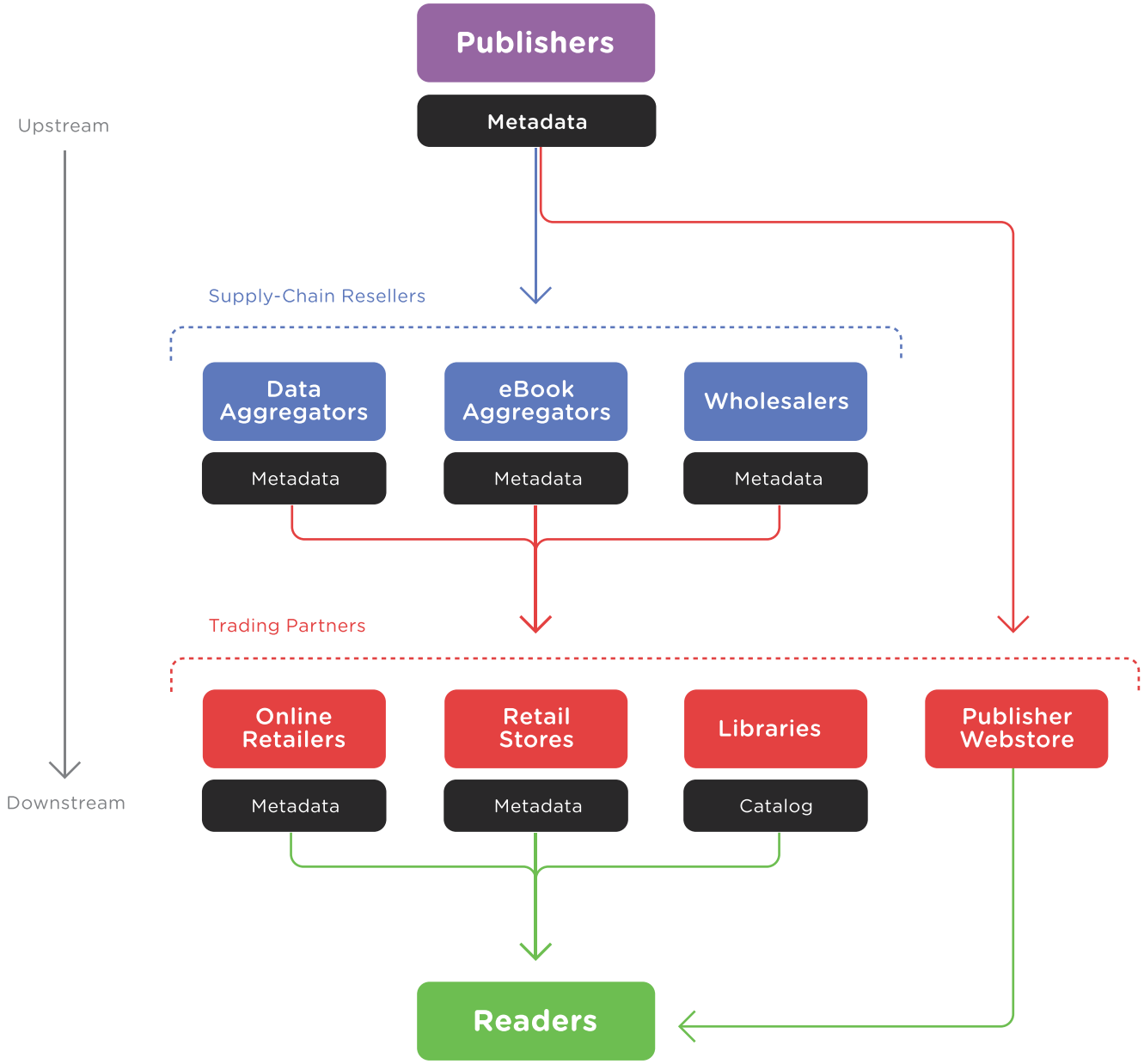
Part 5

Metadata and the Future of Publishing

In Part 5, we conclude with observations from industry leaders on the importance of metadata and how metadata standards and practices continue to evolve along with the rapidly changing nature of publishing itself.

The authors hope *The Metadata Handbook* will serve as an essential reference for publishers. Ultimately, the goal of this *Handbook* is to help connect readers to books by explaining how to best use metadata as an essential part of this endeavor.

Metadata Life Cycle



About the Authors

Renée Register
www.datacurate.com

Renée Register is also author of the *DataCurate Quick Guide* series, including *The Essential Guide to Metadata for Books*, *Metadata Best Practices and Industry Certification*, and *Optimizing Metadata for Digital Publishing*. She developed and instructs a series of classes on metadata for Digital Book World University and conducts webinars and workshops on book metadata.

She specializes in the development of metadata management practices to optimize content discovery, internal operations, and business intelligence. Her company, DataCurate, is focused on products and services to help publishing professionals meet the information demands of the 21st century marketplace.

Renée has over 20 years' experience in building and growing innovative metadata systems, products, and services for materials vendors, metadata providers, and libraries. She holds a master's degree in Library and Information Science and her experience includes ten years with Ingram Book Group and six years with OCLC, a non-profit library cooperative.

Renée Register's profile on LinkedIn.

Thad McIlroy
www.thefutureofpublishing.com

Thad McIlroy is an electronic publishing analyst and author based in San Francisco and Vancouver.

His site, www.thefutureofpublishing.com, is the most thorough exploration of where publishing is headed. McIlroy provides consulting services to publishing and media companies, design and advertising agencies, as well as the full range of vendors serving the publishing industry. Having authored a dozen books and over 300 articles on these subjects, he serves also as an expert witness on patent litigation in the media industries. McIlroy served also for five years as Program Director for Seybold Seminars.

McIlroy is a contributing writer to *PrintAction* magazine and on the editorial board of the journal *Learned Publishing* and the Canadian literary journal, *Geist*. He is a member of the Association for Computing Machinery (ACM), the Electronic Frontier Foundation (EFF), and the Technical Association of the Graphic Arts (TAGA). He co-authored both the first and the second editions of *The Metadata Handbook*. His latest book is *Mobile Strategies for Digital Publishing: A Practical Guide to the Evolving Landscape.*

During the past 25 years he has educated and entertained audiences around the world on every aspect of digital and print publishing including the Intelligent Content Conference, Gilbane Content Management, the Henry Stewart DAM conference, XPLOR, Seybold Conferences, ePublishing (Denmark), the I S & T Annual Conference, and Prepress Asia.

Thad McIlroy's profile on LinkedIn.

About the Designer

Adam Rowe

www.admrwe.com

The layout and graphic design for *The Metadata Handbook* are by Adam Rowe. Adam is a graphic designer based out of Chicago. He received his degree from the College for Creative Studies in Detroit and specializes in identity and information design. Adam also created the graphic design and logo for the DataCurate site.