

Beyond the Book: Metadata Practices at the University of British Columbia Press

by

Alison Joanne Strobel

B.A. (Hons.), English Literature, University of British Columbia, 2014

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Publishing

in the
Publishing Program
Faculty of Communication, Art and Technology

© Alison Joanne Strobel 2016

SIMON FRASER UNIVERSITY

Fall 2016

This work is licensed under the Creative Commons
Attribution-NonCommercial-NoDerivatives 4.0 International
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Approval

Name: Alison Joanne Strobel
Degree: Master of Publishing
Title of Project Report: *Beyond the Book: Metadata Practices
at the University of British Columbia Press*

Supervisory Committee:

John Maxwell
Senior Supervisor
Associate Professor,
Publishing Program
Simon Fraser University

R. Scott Steedman
Supervisor
Lecturer, Publishing Program
Simon Fraser University

Laraine Coates
Industry Supervisor
Marketing Manager,
University of British Columbia Press
Vancouver, BC

Date Approved: December 5, 2016

Abstract

This project report focuses on book metadata practices at the University of British Columbia Press. Metadata management has become essential for publishers in recent decades, as book buying has moved online. This report details the significance of metadata, how publishers use it, how customers (both institutional and individual) benefit from it, and how (good) metadata increases sales. Metadata has become increasingly complex, with varying deadlines, standards, levels, and granularity putting immense pressure on publishers to keep current. This project report analyzes the University of British Columbia Press' metadata operations to identify its challenges and successes. The report also draws on the current literature of metadata “best practices” for publishers. In tandem, these resources clarify optimal future directions and recommendations for the Press.

Keywords: metadata; scholarly publishing; University of British Columbia Press; discoverability; online marketing; digital publishing

Acknowledgments

Thank you to everyone who has supported me during my time in graduate school, in my internship, and in the writing of this report. Thanks go to the entire faculty at SFU in the Publishing program, and most importantly to John Maxwell for his supervision of this project. Thanks also to Scott Steedman for giving this report a very close reading. To the countless industry guests from all across the publishing landscape, all of you have convinced me that books are even more exciting than I had ever imagined.

To the staff at UBC Press, thank you all for being so kind and welcoming during my internship. Many of you took time out of your busy schedules to chat with me about metadata, and your insider perspectives helped shape this report. Special thanks go to the production department at the Press, Holly Keller, Lesley Erickson, Megan Brand, Ann Macklem, and Katrina Petrik; all of you answered my thousands of questions with a smile. Also at the Press, Laraine Coates, thank you for so graciously agreeing to supervise my report, your helpful comments and suggestions throughout this project were invaluable.

To Michael Zeitlin, English professor at the University of British Columbia, thank you for always believing in me, and for inspiring me to become a better student.

I could never thank my family and friends enough for the love and encouragement given to me throughout my education. Thank you to my partner, Reuben; you are my best friend and my most ardent supporter. To you especially, mom and dad; your work ethic has, and always will, inspire me throughout my life.

Table of Contents

Approval	ii
Abstract	iii
Acknowledgments	iv
Table of Contents	v
Chapter 1. Introduction.....	1
A Brief History of Metadata	3
Metadata Today	4
Scholarly Publishing and Metadata.....	5
Chapter 2. Metadata at UBC Press	9
Current Practices	10
Book Information Form and Approval To Publish	10
List Schedule and Transmittal	11
Copy Form	13
Author Marketing Questionnaire	14
Impact of Changing Metadata Requirements	15
Chapter 3. Current “Best Practices”	19
Update, Update, Update	19
ONIX 3.0	21
Get Certified	22
Thema	23
Chapter 4. Recommendations	25
Technological	25
Staffing	29
Organization and Workflow	30
Chapter 5. Conclusion	34
References	38
Unpublished Documents Consulted.....	40



Chapter 1. Introduction

The University of British Columbia Press, best known for its main imprint, UBC Press, is a mature, mid-sized university press. Founded in 1971, UBC Press publishes between sixty to seventy books per year, with an active backlist of over seven hundred titles, and is a leading social sciences publisher. In the Press' 2014 review, one of the challenges identified was the task of producing and disseminating metadata for all of these titles. Though “digital technology offers some relief from certain types of labour-intensive work” there remains the “need for elaborate and accurate metadata. Intellectual labour is still the most expensive and most essential asset for all university presses” (Press review 2014, 48).

Metadata has been defined as “information about information or, equivalently, data about data” (Brand, Daly, and Meyers 2003, 1). This means that the “data” is the book itself, and the data about the book is the “meta” in metadata. Metadata used to be very simple, known as “title-level,” when it first began in the 1960s in libraries. The type of information included was very “limited, in many cases consisting only of ISBN, availability, and price,” because all “transactions were done around a physical object,” the physical book was always there, and thus detailed information became unnecessary (Dawson 2012). In contrast to the bare-bones minimum administrative information about a book, metadata can now be “enhanced,” meaning that it includes the more “descriptive” elements of a book, such as “the short description, long description, review, author biography and table of contents” (Breedt and Walter 2012, 5). Indeed, in today’s digital world, metadata actually moves beyond description or “data about data,” and in fact “becom[es] a part of the object itself” (Jackson 2008, 166).

Instead of just title-level metadata, three distinct types of metadata have been identified, serving three distinct purposes. The first is “bibliographic,” which is the old title-level metadata. Its purpose is to ensure that “users can find an item” and includes

“title, author, and ISBN.” The second type of metadata is “descriptive,” which includes “subjects, summary and table of contents”; this kind of data “drives content discovery.” The third type of metadata is “evaluative,” which “contributes to selection decisions and helps the user make buying choices,” and includes information such as “reviews, awards, and author and contributor biographies” (Register 2013, 30). Bibliographic and descriptive metadata have always been available on a physical book, as browsing readers can generally find all of this information from the book’s front and back covers. But the third type, evaluative metadata, is also now essential in online buying environments, and replaces the function of a very knowledgeable bookseller.

In a 1997 *Publisher’s Weekly* article, James Lichtenberg wrote that metadata “holds the key to being able to use the full power of new information technology” but that “very few people in the industry pay much attention to this kind of data, or how it can be used” (Lichtenberg 1997, 44). Today, metadata is such a powerful a tool that it is impossible to ignore, and most “publishers know they need to do it right, but there seems to be a poor industry-wide understanding of what exactly [metadata] is” (Greenfield 2012). According to interviews done by the World Wide Web Consortium’s Metadata Task Force, there are four main “pain points” or fundamental issues for publishers regarding metadata: its granularity, or the specificity that metadata must have in order to be effective; its complexity, or how non-technical people can find the vocabulary needed to understand metadata confusing; its difficulty, in regard to the change in workflow required to accommodate metadata; and its futility, in that metadata can be misinterpreted and displayed incorrectly by supply chain partners (W3C 2015). Metadata has also continued to be a prevalent issue for university presses in particular. According to an Association of American University Presses (AAUP) report in 2015, “metadata” was noted as “a mild concern” for 59 percent of university press respondents, and elicited the response of a “cause for serious concern” for another 20 percent (AAUP 2015, 15).

In the same 1997 article, Lichtenberg also predicted that metadata would be an essential function for the new digital age, and that “those who are currently publishers will not continue to be at the forefront of the business without understanding metadata as intuitively as they understand a printed page” (Lichtenberg 1997, 44). He presciently

states that “consumers of information products have become much more demanding because of instant gratification through the media and the Internet, and today they require prompt responses and delivery,” a statement that has only become more true for today’s digital natives (Lichtenberg 1997, 46).

A Brief History of Metadata

Laura Dawson, former product manager at Bowker and self-professed “metadata queen,” offers an illuminating essay in *Book: A Futurist’s Manifesto*, edited by Hugh McGuire and Brian O’Leary. Dawson reminds us that in internet bookselling, “discoverability via metadata” is critical as “the content is out of the container” (Dawson 2012). As noted earlier, title-level metadata began in libraries. This information was collected and printed in massive catalogues, but in the mid-1960s, “libraries developed computerized systems and moved away from print bibliographies, [and] MARC [MAchine-Readable Cataloguing] became the standard metadata format for catalog records.” For commercial books, however, title-level metadata only began in the 1970s and 1980s when barcodes were first introduced. In 1995 metadata completely changed its audience, as prior to this point, metadata had only been something “the warehouse crew had to worry about” (Reid 2010, 4). The change came as a result of Amazon, the first large online bookseller, and for the first time, metadata turned to face the consumer. Title-level metadata quickly became insufficient, and soon libraries wanted better metadata too (Dawson 2012).

The additional content needed by these supply chain players was fulfilled by many different sources and in many diverse formats. Due to this lack of standardization, in 1998, a standard was created. The solution was ONIX, or ONline Information eXchange, which is “an XML data transmission protocol” which quickly facilitated communication “among retailers, distributors, and publishers” (Dawson 2012). Managed by the international standards body EDItEUR in the UK, ONIX for Books is still in use today, though it has been improved with newer versions since its 1998 release.

On its website, EDItEUR explains that ONIX is not a database but “a way of communicating data between databases” (EDItEUR 2016). ONIX enables the delivery of

rich product information in a standardized way to many supply chain partners. This reduces costs in two ways: the first is that publishers need only supply data in one format, and the second is that because of this shared format, partners can easily interpret the files and there is less need to manually intervene in files that are used for internal or consumer-facing systems (EDItEUR 2016).

Metadata Today

Most of the information a consumer needs to know about a book in order to make a purchase is presented on the book cover. To give the same experience as browsing in a bookstore, this material must be represented in online environments as well. Now used for digital browsing, metadata is needed for both individual and institutional customers, as well as other publishing supply chain partners. In a physical store, if a book has incorrect metadata – if, for instance, the publisher mistakenly communicated the subject code of the book as “cooking” rather than “computer science,” – the book could be shelved improperly and would have low sales. But there would still be a possibility that the book would be found by happenstance and purchased. However, in the online environment, there are doubly bad consequences for having inaccurate metadata. Online retailers have sophisticated algorithms that monitor data and offer recommendations. So if books have bad data, they “will either be dropped from those recommendations (as they aren’t bought) or they will be incorrectly recommended (and therefore never bought)” (Ruffilo 2011).

It is important for metadata to be accurate, timely, fulsome, and distributed and displayed correctly. Once thought of as a once-and-done marketing tool, metadata is now understood to be dynamic throughout the life of a publication, and is therefore touched and changed by many different industry players. Renée Register, owner of DataCurate, a company specializing in metadata management, explains that there are two types of metadata, upstream and downstream, based on the source. Upstream metadata is “information created near the beginning of the publishing cycle, originating with the publisher or content creator and then flowing downstream to aggregators, wholesalers, retailers, libraries, and readers” (Register 2013, 32). Downstream metadata is created when trading partners enhance and adapt metadata for their own ends “in

support of their business needs and for use in their proprietary products and services” or by adding information such as “critical reviews, awards information, contributor biographies, and other evaluative information” (Register 2013, 33). It makes sense that if publishers give the most and best metadata possible, and in the correct format, the result would be less changes to it, and therefore more publisher control of this very important marketing tool. Although it may seem unfair to have to give away metadata for free, especially considering the high cost of intellectual labour required to create it, “in today’s world, it is sharing metadata, not creating a monopoly of data, that makes you more valuable” (Moffat 2006, 25).

Online book buying has increased throughout the world, and has especially accelerated in recent years. According to the 2015 BookNet Canada report on book buying, in the US and the UK, online book sales have actually surpassed in-store sales. In Canada, buying in-store has a faint margin over online sales; however, “online purchases have increased drastically since 2013. Online purchases of books of any format have increased from 29% in 2013 to 45% in 2015” (BNC 2015, 5). This is an increase of 16 percent in just two years. In addition, the chief avenue for discovering a book is browsing, both online and in-store, which makes up 19 percent of discoverability overall (BNC 2015, 8). Having proper metadata displayed online is crucial for sales.

Scholarly Publishing and Metadata

In 2012, Nielsen analyzed the top-selling 100,000 titles in the UK from 2011, and cross-referenced their sales with how complete each metadata record was for each title. BIC, or Book Industry Communication, oversees the UK’s metadata standards, which is analogous to the BISAC standards in the US overseen by the BISG (Book Industry Study Group). BIC considers eleven elements to be “basic”: ISBN, title, product form, main subject category, imprint, publication date, cover image, supplier, availability, retail price, and rights. They compared records for books with complete data (i.e. the basic elements) and a cover image versus incomplete records and no image. The results are clear: “when we look at records with all of the necessary data and image requirements, average sales reach 2,205 [units]. This represents an increase of 473% in comparison to

those records which have neither the complete BIC Basic data elements or an image” which only sold on average 385 copies (Breedt and Walter 2012, 3).

This same study revealed that having a complete or incomplete metadata record impacts both online and in-store sales. Shockingly, average sales for in-store (offline) retailers are more impacted than online retailers by having complete basic elements and an image: “The offline retail channel sees sales rising 124% for titles meeting the BIC basic standard, whereas online retail sales see growth of 48%” (Breedt and Walter 2012, 4). Booksellers make buying decisions based on metadata. This could mean that the better the metadata, the more likely the book is to end up in bookstores, and the more likely the bookseller is to recommend it. Thus, having full and accurate metadata helps sell books in online and offline channels, and has more of an impact than some may think.

More specific to this study is the comparison by genre that Nielsen performed. Since UBC Press titles would be considered by this study either specialist non-fiction or trade non-fiction, they compared sales of titles in these genres with incomplete versus complete records as well. They found that “Specialist Non-Fiction titles ... see average sales rise by 33% for titles with complete data and an image” and that “Trade Non-Fiction average sales per ISBN grow by 97% when all BIC Basic elements are present” (Breedt and Walter 2012, 4).

As noted earlier, publishers know that metadata is important, but until recently there wasn't enough data to understand just how important each element of metadata is. What is the impact when specific elements are missing? Nielsen charted four of the “enhanced” metadata elements – short description, long description, review, and author biography – to find out which had the highest impact on sales when not present. Including all genres and online and offline sales, “titles which hold all four enhanced metadata elements sell on average over 1,000 more copies than those that don't hold any enhanced metadata ... In percentage terms, titles with ... all four data elements [show an average sales boost of] 55% when compared to titles with no enhanced metadata elements” (Breedt and Walter 2012, 5). For online-only sales, as one would expect, records with increasing amounts of metadata showed increasing sales: “Records

with just one enhanced metadata element see an increase of 55% in comparison to those with none; those with two enhanced metadata elements see an increase of 71%; those with three increase 120%; and those with all four enhanced metadata elements see average sales increase by 178%” (Breedt and Walter 2012, 6). Again looking at just the genres that are appropriate to UBC Press, “Specialist Non-Fiction and Trade Non-Fiction all see the greatest negative impact on sales when the long description is omitted from the title record” (Breedt and Walter 2012, 8). It makes sense that researchers and professionals would be most interested in reading the long description of a book, so the conclusion here is that the long description has the most impact on the sales of UBC Press books.

“Book buying for UBC Press takes place in databases and online.”¹ Indeed, UBC Press’ top five customers are e-commerce (Sales figures 2016). This means that metadata is very important for the Press, because as soon as book buying moved online, “customers, browsing online, needed to know about the book: they wanted to see the cover, full description, author bios, and so on. Because they were no longer looking at the book in the store. They were no longer talking to a bookseller. They were making purchasing decisions based on what they saw on screen” (Coates 2016). In publishing, what is just as important as the number of copies sold, if not more important, is the volume of returns. The Press’ best customer in terms of returns is, not surprisingly, through their own website, none! (Sales figures 2016).

Amazon is UBC Press’ largest customer, and to that end metadata must be the Press’ utmost priority. Amazon has always been ahead of the game: “While bricks-and-mortar booksellers moved to create online options, Amazon had already invested heavily in metadata, website development, search and discovery options, and algorithms that also analyzed user behavior to expose the right titles to the right readers” (Register and McIlroy 2012, 14). Not only that, but “overall, Amazon has the best use of metadata across all retailers” and therefore title information must be comprehensive as well (Tallent 2016). There are also real-world consequences for having incomplete or incorrect metadata. For example, Amazon has strict requirements for copy, and any

¹ UBC Press staff member, in conversation with the author, June 7, 2016.

publisher failing to meet those requirements will not have its books listed for sale. Another is that the publication or release date of the book displayed on Amazon is crucial to customer satisfaction, for example through pre-orders, so if a book comes out later than committed (and Amazon is not notified in time) the publisher may be subject to extra charges.² The Press' second largest customer is Indigo. As of 2015, Indigo's buying power has pushed UBC Press into publishing for a third season, and they must now produce spring, fall, and winter books. A consequence for not having metadata complete for each of these lists is that if the Press misses the metadata deadlines (and sales meetings), Indigo may not buy *any* books from the corresponding upcoming season.³ It is essential that UBC Press have full and complete metadata ready for these important sales channels.

² UBC Press staff member, in conversation with the author, July 7, 2016.

³ UBC Press staff member, in conversation with the author, June 7, 2016.



Chapter 2. Metadata at UBC Press

Like most publishers, UBC Press used to keep all of its metadata organized in Excel spreadsheets. As the Press began to publish more books each year and its backlist continued to grow, spreadsheets became next to impossible to manage, and towards the millennium the Press switched to a database called Press Track. Press Track created a file for each book and stored basic metadata elements including author information. Around 2007, the Press again made a transition to a newer database, called Klopotek. Klopotek is very capable and robust, and is probably even more comprehensive than the Press needs. The system was originally created for journal publishers, and therefore has the ability to manage thousands of authors and articles. It is more than capable of cataloguing all of UBC Press' new titles each year, as well as its rich backlist of forty-five years of publishing. Another great feature of Klopotek is that it can handle imports from UBC Press' distributed publishers, whereas Press Track is all manual entry. Currently continuing work on making a full transition, the Press is still using Press Track for its internal production reports and for the UBC Press website, so both Press Track and Klopotek are being maintained as title databases. Klopotek is used for author management and royalties, metadata organization, and ONIX generation. This evolution of management systems and tools exemplifies the impact of metadata requirements on the Press, as the need for more robust and versatile metadata increased, so too did the need for newer technologies.

The demands of managing metadata have also affected Press staffing. A lot of the work done by the newly created position of editorial coordinator has to do with metadata. The position was created in 2015 with the overall function of bridging the acquisition and production editors. In practice, the editorial coordinator starts collecting the metadata for each title from the acquisitions editor, associated documents, and other stakeholders (authors, freelancers, endorsers) and transmits it to the production department. Another position created relatively recently, in 2009, is that of bibliographic

data coordinator (now retitled digital projects manager). When making the transition from using Press Track to Klopotek, the bibliographic data coordinator spent nearly three years cleaning the metadata and ensuring the data pathways worked, meaning that the proper data senders and receivers were both doing their jobs, and that data was being displayed correctly on supply chain partner websites. These two positions show that metadata creation and management has become increasingly important to UBC Press, and something that it takes very seriously.

Current Practices

Every title at UBC Press goes through a rigorous process to make sure as much metadata as possible is collected in the most organized way, beginning right at the start of the publication process. There are five main forms that are essential to gathering and generating metadata, and another that represents the status of all five, the List Schedule. These forms are stored in UBC Press' shared hard drive, so that all staff members can access them as needed. They are filled out by the responsible staff member(s) and emailed to outside parties (authors, freelancers) for inputting information or signing off on final drafts. Information from these documents is entered into Klopotek once finalized, and is then sent out to supply chain partners and updated as necessary. Each form is explained below in the approximate order in which it is filled during the publication process. The changes that have been made to the forms over the years have been highlighted due to the changing and increasing demands of metadata.

Book Information Form and Approval To Publish

The List Schedule draws on two forms: the Book Information Form (BIF), and the Approval To Publish (ATP) form. The BIF is the first document that the author writes about their book, detailing everything from topic to potential market; it must be submitted by the author to the Press with their manuscript. The next form is the ATP, which is part of the scholarly apparatus. The acquisitions editor draws from the BIF to write a description of the book for the ATP. The acquisitions editor also adds sample chapters of the book, and includes peer reviews of the book by two experts in the respective field

and their publication recommendations. The publications board relies on the ATP to indicate to the Press that the book should be published.

The BIF has gone through many changes to facilitate better metadata capture. Questions that used to be asked at the late stages of publication (in the Author Information Form, now renamed Author Marketing Questionnaire) have been moved to this first stage. For example, the questions “Is your book topical?” and “What is controversial about your book?” are now on the BIF (AIF 2011; BIF 2015). The BIF also asks for a 200-word description of the book, with the instruction to: “set the stage for the book with some general comments about its subject matter,” “provide a short description or synopsis of the content of your book,” and “say why your book is important and who would be interested in it” (AIF 2011; BIF 2015). The last change to this form is asking for “A succinct, snappy one-line description of your book that distills its essence” (AIF 2011; BIF 2015). The answers of these questions form a rough draft for metadata production.

Asking these questions earlier is also important because it helps each stakeholder (the author, the publications board, and the Press) better understand the “why” of the book. The BIF gets the author thinking about audience and marketability; helps the publications board see the value to scholarship; and helps inform Press decisions regarding the book. All of these answers help with the associated metadata, and asking these questions earlier helps clarify and expedite the process so there will be fewer changes in the metadata elements later.

List Schedule and Transmittal

The organized metadata collection begins with the List Schedule document, created and maintained by the editorial coordinator as a reference for other departments. This spreadsheet visually represents which metadata elements have been collected for each title, and shows which information still needs to be collected. This document is mainly used by the acquisition and production editors, showing these departments the status of each title, and identifies any bottlenecks (for example, manuscript status). The editorial coordinator created the List Schedule to organize, visualize, and plan the mass amount of metadata that the Press needs to collect for every book. Some of the columns on the

form are used to track the status of specific metadata elements, like whether the title is finalized, whereas other information is Press-internal, such as potential or approved funding and the production editor responsible for the title (Spring 2017 list schedule 2016). The List Schedule tracks the information for each title, but the Transmittal Form is much more applicable to metadata generation. Each title has its own Transmittal Form (and meeting), where all of the elements from the Transmittal Form are discussed and negotiated. The elements for each title on the Transmittal Form are as follows:

1. Title (working/final?)
2. Author(s)
3. Series/imprint
4. Category: monograph or collection
5. Co-publication potential?
6. ISBNs
7. Author bios (long and short)
8. Descriptions and one liner
9. Table of contents
10. Top selling points
11. Audience and course potential
12. Subject areas and codes
13. Contract, funding sources, and permissions
14. Page count, illustrations, and tables
15. Manuscript in? BIF?
16. Schedule (needed for conference?)
17. Prices and print runs. (Manuscript transmittal form 2015)

The Transmittal meeting involves the editor who acquired the manuscript, the production editor assigned to the title, the marketing manager, and the Press' director. The Transmittal Form represents what has to be done with the manuscript and what to expect, and gives direction to all other departments. After the book record is created in Klopotek from the contract, it is from the finalized Transmittal Form that the bulk of the early metadata is entered.

Copy Form

Once the manuscript has been officially transmitted to the production department, the next document in metadata generation is the Copy Form, which includes different versions and lengths of copy for various purposes (catalogue, web, reviewers), and is managed by the production editor. The Copy Form was created around 2009, and has gone through many iterations to produce better metadata. When the form was first created, many staffers thought the different lengths of copy were simply overkill, but now it is seen as an essential document for all the different needs of the Press, and all agree that it is best for it to be together in one place.

The biggest change to the most recent form is that it has more instruction and description for desired copy, and better guidelines. For example, instead of a blank field with “promotional one liner” as its only prompt (Copy form 2009), it now reads: “sales handle/promotional one liner: Short description/annotation. This will most likely be used in the catalogue and on the web. The sales handle is 25–40 words, basically a 7-second pitch, that answers ‘What is it? Who is it for? And, so what?’” (Copy form 2015). One of the reasons for the in-depth guidelines is that if the copy is being freelanced out, the Press needs to make sure that their instructions are clear, which in turn saves time and money.

Another factor of note is that each section mentions the audience for each piece of copy: the main copy is for catalogues and librarians, the medium copy is for the book jacket, and the long copy is for the web and is therefore crucial to individual sales and must have a “hook” to draw in prospective readers (Copy form 2015). There are multiple and varied descriptions because “metadata has to be directed towards audiences.

Plural.” (Coates 2016). If all stakeholders (author, freelancers, Press staff) are clear about the audience for each section of the copy, it produces better results and it saves time both in that it is clearer how and why copy is written in a certain way. The Copy Form has to be signed off by the author, the acquisitions editor, the production editor, and the marketing manager. The changes from the 2009 form to the 2015 form have helped streamline the process, clearly targeting what the Press is looking for in each section so that metadata is ready for all different purposes.

Author Marketing Questionnaire

The final document for metadata collection is the Author Marketing Questionnaire (AMQ), which is filled out by the author when the book is being sent to the printer, and is used primarily by the marketing department as consumer-facing marketing material. The first difference between the AMQ and its predecessor, the Author Information Form (AIF) is evident from its name: the focus has clearly shifted from getting *information* about the author to getting the right information to *market* the book. The AMQ has many new questions that better demonstrate the wide array of today’s digital marketing strategies.

The AMQ asks specific marketing questions about the author’s book. For example, it asks, “is your book newsworthy or does it address current events?” (AMQ 2015), whereas the AIF only asked if the book was “topical” (AIF 2011). For reviews, the AIF asked “please list the key journals ... in which it would be particularly appropriate to advertise your book” (AIF 2011); it now asks “please list up to ten key journals *and review media* ... to which we should submit review copies of your book,” in addition to the “three most important *venues* in which we should consider advertising your book” (AMQ 2015, emphasis added). This shows that the AMQ is media agnostic, not simply advertising in journals, but also in any other form of appropriate media.

It also points out the key market for the book and asks the author to think about how their book will be used. It asks about individual consumers: “list keywords that readers would use when searching for works in your subject area,” which is, of course, important metadata (AMQ 2015). It also asks about students: “as a *course book*, what are your book’s strengths and how will it benefit the interested reader?” (AMQ 2015,

emphasis in original), rather than just asking: “if your book could be used as a textbook ... please note the type of course” (AIF 2011). These answers help inform marketing copy for both individual and course adoption sales.

The AMQ also asks marketing questions about the author. The AIF asked only if the author had a website, but the AMQ asks: “do you have a webpage, blog, Twitter, or other online presence?” (AMQ 2015). It also asks, “Are you a contributor to any newspapers, journals, or other media? Have you previously appeared on or been interviewed by the media?” (AMQ 2015). These are important sales channels in today’s market. A very interesting, and perhaps unexpected, question it asks is “please provide us with a personal statement about why you wrote this book and the experience of writing and researching it. This is helpful when we develop promotional materials, for some award submissions, and for the media” (AMQ 2015). This type of question moves the author beyond the academic and toward the emotional hook that can really help with marketing. All of the questions on the AMQ help set the tone of the promotional copy, and show where the metadata should be sent.

Impact of Changing Metadata Requirements

Not only does UBC Press have to deal with changing metadata types and timelines, it now also has to produce metadata for a third publishing season imposed by Indigo in 2015. The third season is the “winter list,” books published between January and April, and is for trade books only. UBC Press is getting more serious about trade books (non-peer reviewed and wider appeal) with On Point Press and On Campus Press, its trade imprints. Trade publishers have always had to have their metadata ready earlier, but this is new for UBC Press and it is a tough adjustment. One consequence for the creation of this third season is the cancelling of the “winter whip,” which was when UBC Press sales representatives would go around and re-sell the Press’ fall season books that had gotten less attention than they deserved. This used to happen right before the Christmas season and really helped boost title sales. Now there just isn’t enough time for it.

The biggest change for metadata for the Press has been the moving deadlines, and every staff member commented that this was the hardest factor to control and

manage about metadata. In 2015, there was around a seven-month lead-time for metadata for Indigo, but even in the space of one year the deadlines have moved up. In cross-referencing a 2015 Indigo buying dates document with a 2016 metadata deadlines document, we find that every deadline has moved up by two weeks or more. For example, books published between January and April (winter season) have a metadata deadline of July 15, which has moved up by two weeks; books published between May and July (spring season) have a deadline of September 15 which has moved up by six weeks; and books published between August and December (fall season) have a metadata deadline of February 15, which has moved up by two weeks (Indigo buying dates 2015; Metadata deadlines 2016). This means that metadata is due so early that very shortly after a book is transmitted, its metadata is already due. What is most stressful about the deadline moving up every year is that “it is nearly impossible to plan, because once the deadlines are adjusted, they move up *again* the following year!”⁴

Metadata crunches are brutal for Press staff, and the production department nearly shuts down for up to three weeks each season to deal with the immense backlog they cause. The crunches hit immediately before the due dates, and because of the pressure to have the metadata done – without it, the book is virtually unsalable – each production editor goes from working on one season’s books, about seven to eight titles, to working on two seasons, about fourteen to sixteen titles. This is because each editor is finishing editorial work on books for the earlier season, as well as getting metadata ready for the next season’s books.

The production department is responsible for copy with other departments providing input, because they, at this point, are the most familiar with the manuscripts. However, all of this stress and extra work can be intense; “at first it just seems like adding one or two things to your workflow but then they require more and more information, and then what you’re doing now is a completely different job than what you should be doing.”⁵ This hectic schedule also leaves no leniency for lateness at any point in the process; if the author with the manuscript, or the freelancer with the copy, or any

⁴ UBC Press staff member, in conversation with the author, July 28, 2016.

⁵ UBC Press staff member, in conversation with the author, May 27, 2016.

of the other parties are delayed in responding to the changes/comments/signing off of the Copy Form, every stakeholder has less time to complete the duties that have very firm deadlines. Finalizing everything with multiple players through email is very time consuming, but it is impossible to release metadata that is not final, as trying to recapture inaccurate metadata is unfeasible. This necessity to finalize is why a Klopotek file is created when a book becomes official, but then is updated by the production editor right before the book is printed, because many things can change in that timeframe such as page count or promotional copy.

This tight metadata schedule is also why books have been slipping from eight- to ten-month timelines to ten- to twelve-month timelines. If a book's metadata is not ready for the next metadata deadline, even if the production timeline is realizable, the book will have to be pushed to a later season. This just shows how influential metadata is: the book can be finished on time, but if the metadata does not match the care and attention paid to the final product, nobody can find or buy the book. It is now true that "hand-selling of books is a thing of the past: if retailers (even indies) can't get accurate and good metadata, they will not buy."⁶ So the metadata, in fact, dictates the list schedules, not the other way around. One of UBC Press' major selling points is that they can publish a book in eight months (many other university presses can take two years). Arguably, however, the Press needed this selling point when it was trying to prove itself, but now, it has a great reputation beyond a fast-track production schedule, and maybe lengthening the production schedule to pad out the metadata deadlines would actually be a good solution to the current scheduling issues. As attractive and simple as it sounds to elongate the timelines for every book, that is treating the *symptom* of the metadata crunches, not the actual problem. It is not that the production department needs more time for each book; it is that metadata deadlines themselves put the books on hold, so the only real solution is to manage metadata better.

Not surprisingly, due to the mass amount of information that each title needs, metadata management services have cropped up to solve the issue of in-house metadata production and dissemination at all types of publishing houses. It would be

⁶ UBC Press staff member, in conversation with the author, June 7, 2016.

much easier to have a metadata management service handle all of the Press' metadata, or to not have every department scrutinize every Copy Form, and to only have one description for each book. But UBC Press takes copy very seriously. The marketing copy, especially for UBC Press' scholarly titles, is very closely intertwined with the discipline, market, audience, and positioning, and therefore must be of the highest quality. Also, each of these books represents an author's academic career. Just as important as the creation of metadata is its dissemination, a task done by the marketing department, which involves sending specific metadata (in the correct format) to each partner and channel. This precious asset must be managed in-house, as "no one knows your books better – or cares about them more – than you do, and no distributor, agent or wholesaler will be as invested in getting it right" (BNC and Canadian Bookshelf 2010, 4).



Chapter 3. Current “Best Practices”

A good way to think about metadata is to think about parenting. Parents give as much advice to their children as they can, and set the best example they can. Children will take in the advice, but in the end, they will do what they want. For metadata, publishers should give as much data as they can, and present it in the highest quality form they can (upstream metadata). Like the children, publishing supply chain partners; such as distributors, wholesalers, aggregators, and retailers, will take the metadata and will use what they want and display it in the form that they choose (downstream metadata). UBC Press supplies the most information it can in its ONIX files, and then its distributors, such as the University of Toronto Press and, in the US, the University of Washington Press, take that metadata and add supply detail for retailers (how many copies are available) and modify the records to fit their databases (this generally means simplifying the files). Library wholesalers, like Baker & Taylor, supply “value-add” services to UBC Press metadata; such as creating and including MARC (MACHINE-Readable Cataloguing) records in UBC Press files in order to better serve the library market. Library ebook vendors and aggregators, like ProQuest, also add ebook MARC records for libraries. And when retailers receive UBC Press metadata, they ingest the ONIX file and display the information in their own proprietary form either in their in-house systems or on the web for consumers. This chapter will explain the main metadata recommendations found across various best practice guides. The chapter following, Chapter 4, will examine these best practices further in relation to UBC Press specifically and offer recommendations for the Press.

Update, Update, Update

Metadata should be created as soon as a title becomes official. The authors of *The Metadata Handbook* recommend that “metadata ... should be updated with important

changes or additions to title information prior to the official publication or on-sale date, and should receive further updates to reflect important events or activities relating to the book after publication” (Register and McIlroy 2012, 1). Many best practice guides cite that “major wholesalers, retailers, and distributors want this information six months in advance of publication” (Register and McIlroy 2012, 5; see also BISG 2013, 9; Luther 2009, 7; Smith 2013). As we know, prepublication data is fluid, so these guides recommend that metadata be released six months before the on-sale date, and all finalized and complete metadata be released four months before publication, including cover image files (BISG 2013, 50; Smith 2013).

Metadata, beyond supply detail, also needs to be updated. Joshua Tallent of FireBrand Technologies, a metadata management service, begs publishers to make a scheduled effort to update metadata, especially for backlist titles, at the very least once per year (Tallent 2016). He urges publishers to be aware of new ideas, recent events, cultural shifts, and news in order to generate more interest in new and backlist titles. Tallent provides a real-life example of how updating metadata can generate sales: in 2010, Simon & Schuster published Nic Pizzolatto’s *Galveston*, which only sold about 1,000 copies. The author later wrote a TV series, *True Detective*, so Simon & Schuster went back into their ONIX file on the book, and added a one-sentence reference in their metadata into the contributor biography field: “From the creator, writer, and director of the HBO crime series *True Detective*.” Simon & Schuster sold 37,000 new copies of the book in 2014 alone. Tallent stresses that publishers should always update metadata on their authors to include information about books they have published, even with other publishers, to help promote them – doing so may have positive reciprocal sales effects as well (Tallent 2016).

Metadata updating for scholarly publishers is just as important as for trade publishers. The specialized audiences of UBC Press titles benefit from having frequently updated metadata, to reflect new trends in academia and updated perspectives on recurring issues. Indeed, there are many benefits for research communities by updating metadata, such as helping “scholars seeking out ... obscure backlist titles” (Brand, Daly, and Meyers 2003, 6–7). Another way to ensure that metadata remains evergreen is to include keywords, as these “can be particularly useful when a consumer’s probable

search term is new, jargon, distinctive, or quite specific” (Warren 2015). UBC Press publishes scholarship in the social sciences by new and seasoned scholars, books which deal with innovative and complex ideas and would benefit from updated metadata to reflect cutting-edge research and novel terminology. UBC Press does regularly update Klopotek with awards and distinction information (for example, an author being promoted from assistant to full professor), but adding in information to the author biographies about books published with other presses could also be useful.

ONIX 3.0

It is also a best practice to update to the newest version of ONIX, version 3.0, which was released in 2009. ONIX 3.0 has many improvements over the previous version released in 2004, ONIX 2.1, which UBC Press is still using. The best and first reason to update to ONIX 3.0 is that version 2.1 was sunsetted in 2014, and is therefore no longer supported (EDItEUR 2016). Graham Bell of EDItEUR explains that “ONIX 3.0 really represents a decade’s worth of accumulated experience” in that it is more rigorous in checking XML, has better and global best practice support, and reflects today’s digital capabilities better (Bell 2015b). ONIX 3.0’s main improvements are that it is more global in scope as it is multilingual, it is better with international sales in that it is more specific with territorial rights, and it is better equipped to handle digital products and sales.

ONIX 3.0 is also more specific with its code naming fields, meaning that each metadata element has a distinctly named field to go along with it; previously (in version 2.1) important information was stuffed into fields that had no single agreed function. For example, the long description of a book used to be under the “Other Text” field, along with a bunch of other information, but now each description has its own field, in this case “Long Description,” so it is now more likely that the metadata that is sent will be interpreted and displayed correctly. ONIX 3.0 is consistent, in that there is only one recommended way of doing something; simple, in that repetitive elements have been deleted from version 2.1; and flexible and extensible, in that new elements can be added with a code list addition (Bell 2015b). Its new functionality is also desirable in that ebooks have specific fields for usage constraints and digital rights management (DRM), as well as marketing collateral material can be audience-specific (Bell 2015b).

The Book Industry Study Group (BISG) released a best practice guide in 2013 in cooperation with BookNet Canada that does a great job of detailing all of the elements (both required and recommended) that should be included in book metadata, how each of these elements can affect business operations, and how ONIX 3.0 differs from ONIX 2.1 in each element (if there has been a change). The guide specifies 32 total elements, including the “enhanced” elements that go beyond bibliographic detail (“core” elements). It is a best practice to include all of these elements in simple XHTML, as it can be validated, and it is more likely that it will be displayed properly on supply chain partner websites as XHTML transmits formatted data (BISG 2013, 82).

The major issue with updating to ONIX 3.0, however, is that it is not backwards compatible with ONIX 2.1. This has significant business implications for publishers, and thus any publisher wishing to make the transition should consult their trading partners and ask what their requirements and capabilities are beforehand. The actual task of updating to a new system is not as complicated as one may think, though, as if ONIX 2.1 has been fully updated, half of the work is already done, as all of the new functionality is optional (Bell 2015b). It is also a huge benefit that upgrades to ONIX 3.0 will be accomplished via code lists rather than upgrading to a whole new system (Bell 2015b), meaning that if publishers make this transition soon, it is very unlikely that they will need to make another major conversion in the foreseeable future.

Get Certified

There is a lot of pressure on all types of publishers to keep current with digital technology and the capabilities of metadata. There are, of course, many certification programs that publishers can subscribe to in order to ensure their metadata is being sent out accurately and to the highest standard. In Canada, BookNet had a metadata validation program based on the Canadian Bibliographic Standard, which UBC Press subscribed to, that analyzed and validated ONIX files and awarded them either a bronze, silver, or gold rating. This system was based on “a minimum set of data elements which should be available in an accurate and timely form throughout the Canadian bibliographic supply chain” (BNC 2007, 3). This meant that if a publisher fulfilled the minimum amount of data points, they could be awarded a “gold” rating.

BookNet has since replaced this standard with BiblioShare; a new system that also awards the same ratings based on metadata and the Bibliographic Standard, but is based on different criteria. Under BiblioShare, the new “focus for certification is now on data quality and making the most of ONIX, rather than on the presence or absence of particular data points” (BNC and Canadian Bookshelf 2010, 5). This means that certification is now based on quality of data rather than just quantity. The program is also more comprehensive than it used to be: in addition to certification, it now “provides regular, automated reports that identify problems in ... ONIX file[s] such as missing or invalid information. Additionally, the program helps keep [publishers] up to date by making recommendations for following best practices ... as the market changes and develops, BNC adds the information so the feedback [publishers] get is always current” (BNC and Canadian Bookshelf 2010, 5). The program foregrounds the importance of validating files, as data receivers will validate files before they accept them, so pre-validating the files with BookNet first saves publishers from having to fix errors later, and “make[s] [their] titles visible right away” (BNC and Canadian Bookshelf 2010, 6). UBC Press sends its files to BiblioShare, and includes enhanced metadata, as recommended by BiblioShare, by “going beyond the BNC Gold Standard to include the same kind of information [the Press] would include in [its] catalogue or on a detailed web listing” (BNC and Canadian Bookshelf 2010, 5). These enhanced elements are things like full contributor biographies, links to the author’s website, jacket copy, excerpts, tables of contents, reviews, prizes, and media mentions.

Thema

Thema, launched in late 2013, is “the international subject category scheme that aims to reduce the duplication of effort needed to support the plethora of nationally focused subject schemes in what is increasingly a truly global book trade” (Bell 2015a, 30). There are many different subject schemes throughout the world, but Thema is most similar to the BIC subject categories already used in the UK. Thema is managed by EDItEUR, which is also responsible for ONIX. Basically, Thema is “a hierarchy of subject categories. Each category has a code, plus a heading that describes what books

assigned that code are about ... and includes a range of qualifiers that can add important detail” (Bell 2015a, 30).

Thema has been accepted in many areas of the world; “in North America, ... the Book Industry Study Group ... now maintains mapping between its own BISAC subject scheme and Thema, and BookNet Canada has developed an automated version of this mapping that can bring BISAC-coded backlists up to date with Thema almost instantly” (Bell 2015a, 31). One of the reasons Canadian publishers are among “the most enthusiastic proponents” of Thema is that, as Canada is always positioned between the US and the UK, Canadian publishers are currently “faced with using both BIC and BISAC” subject categories (Bell 2015a, 31). Thema would cut this effort down to one scheme. Before Thema, like before ONIX, every national market (or even every publisher) was doing things differently. But with Thema, like with ONIX 3.0, the book trade is global, reducing duplication of labour, as everyone in the supply chain will be communicating in the same language and format. As a result data sharing and comparing will become much easier. The good news is that Klopotek already supports both ONIX 3.0 and Thema (Bell 2015b).



Chapter 4. Recommendations

UBC Press has adjusted to the needs of metadata in many aspects of its business, chiefly through technology, staffing, and general organization and workflow. The recommendations in this chapter examine how UBC Press can benefit further from its progress in these areas. It is clear that these recommendations also will take time if implemented; most of them would take a matter of years. Taken together, these suggestions help integrate metadata processes more fully into regular workflow, which will help smooth out metadata spikes in the years to come. These recommendations show how UBC Press can get more out of its metadata; how staffing is needed to accomplish its metadata goals with efficiency; and how the overall organization and workflow of metadata generation and distribution can be improved. Overall, the goal is to reduce the repetition of work, and in the process, save time and money.

Technological

One of the best ways that UBC Press can improve metadata workflow is to get rid of Press Track, and to run everything through Klopotek. Press Track is outdated and very slow, and currently only has two functions: producing production reports and feeding data to the UBC Press website. Klopotek is much newer and more comprehensive, and its capabilities far outstrip Press Track. The Press will also save a lot of time in that data will only need to be entered into one database in order to be reflected on its website and passed on to trading partners. This will also reduce potential error. As the marketing department sets all final prices, subject codes, and publication dates for titles in Klopotek, these elements often remain incorrect in Press Track (and are hence reflected incorrectly on the web and in production reports as well), which is a constant source of confusion. As UBC Press is currently working on launching a new website (which will take data from Klopotek) it is already moving towards this solution. It is much easier to

keep everything consistent on a single interface, which limits extra work and saves time when changes do have to be made, because they only have to be made in one location (Tallent 2016).

Another recommendation is to begin the migration from ONIX 2.1 to ONIX 3.0. As mentioned earlier, ONIX 2.1, which the Press is currently using, has been abandoned as a standard by EDItEUR and is no longer supported. ONIX 3.0 has many new features that will benefit the Press. One of the most useful upgrades is that different content can be targeted to different audiences. Considering that the Press already solicits different descriptions for different audiences on the Copy Form, the hard work has already been done. So instead of just including one description, “ONIX 3.0 provides a more streamlined and flexible way to make marketing content available to a variety of audiences. Descriptive text can be targeted to the end consumer, to a trade customer, or to a librarian. Links can be provided to material provided by the publisher, like images or ... online reviews” (BISG 2013, 122). The audience is specified through the “Content Audience” field and determines how each description will be used. Marketing copy cannot be underestimated; every member of the supply chain uses it, from online individual consumers to “buyers for libraries, wholesalers, distributors, and retailers,” and each of these players “need to understand what they are being asked to purchase, and they can make good use of textual descriptions of products. Branch librarians and in-store booksellers can also use this information to help their patrons” (BISG 2013, 113).

Another new feature of ONIX 3.0 is that it supports the use of keywords. Again, keywords are already being created for UBC Press titles, so adding them into the ONIX feed is the easy part. Keyword stuffing into descriptions is not only sounds awkward but is also “detected and punished by search engines,” so this dedicated field is very useful (Register and McIlroy 2012, 48). Keywords are designed to mimic the actual searches done by potential end consumers, as “many searches ... are comprised of natural language queries that describe different elements of a book, such as its setting, characters, theme or an emotional response to its content. Keywords ... allo[w] people knowledgeable of the book to specify additional terms by which to find it” (Sim 2016). It is also crucial to include keywords as “recently, Amazon started supporting keywords into its searches,” which serves as a “prime example that if publishers hadn’t been

including this type of metadata, keywords, Amazon wouldn't have enough data to work off of" meaning that the more good quality metadata included in ONIX, the better (Tallent 2016).

ONIX 3.0 is also much better at dealing with digital products. UBC Press is expanding its sales of digital books through the launch of its new website, which will serve as a direct sales channel for Press titles to end consumers. Yet the largest benefit to do with ebooks is ONIX 3.0's ability to include usage constraints, the "limitations on the use of a product and its contents by the customer (or licensee)"; and DRM, "a technical method to monitor or enforce usage constraints" (BISG 2013, 26–27). The embedded usage constraints and DRM within the digital file are important because of library sales, especially as UBC Press titles have obvious university appeal. Institutional libraries are now pushing for full-text searchable metadata, meaning all of the text of the book would actually become part of the metadata. Instead of purchasing ebooks ahead of time and making them available on their lending platforms, libraries are instead fighting for patron-driven acquisitions (PDA). Currently, the purchase price for UBC Press ebooks for institutions is the cost of the hardcover book, which is usually \$95 to \$99. PDA, conversely, will only trigger a purchase of a title from the Press when either a patron uses the book for more than a certain amount of time, or when a certain number of users access the book.

Full-text metadata and PDA aid libraries in two ways: first, with full-text metadata, patrons can search through an entire book to find pertinent information much quicker; and second, with PDA, libraries can cut spending and risk as they do not have to buy a digital copy of a book before the book is actually used. This is a dangerous thought for publishers, as they have already been hit hard by cuts to institutional library budgets over the past few years. UBC Press and other publishers are at present negotiating the specific terms of PDA with libraries – or, in fact, whether they will stay in PDA at all (some presses are disallowing PDA for frontlist titles). Using ONIX 3.0 would better communicate the DRM and usage constraints applied to each UBC Press title for individual sale, or to specify if a title was eligible or ineligible for PDA.

Another benefit of ONIX 3.0 is that previously, in ONIX 2.1, “exchanges [were] based on sending complete records, either as new information or to replace information previously supplied,” meaning that even if there was only one minor change to a record, say, the changing of supply detail, a whole new record would have to be generated and sent out to every trading partner (EDItEUR 2016). In ONIX 3.0, “a more flexible and granular approach to updating is supported,” which eliminates the need for sending a whole new record (EDItEUR 2016). This is helpful for many reasons: the file size is smaller, there is less room for introducing other errors when sending new messages with small updates, and there is less of a chance for error on the trading partner end. While using ONIX 2.1, the issue of sending out new metadata if there is an error or change is that staff have to follow up with respective vendors and make sure that they receive the second round of metadata and that it is full and correct; and because the publisher is forced to send a whole new file, a lot of times retailers can think it is an entirely new book rather than just a corrected record. The new organization of ONIX is through six distinct “blocks” so that only the modified block will be re-sent. The blocks are as follows: Product description; Marketing collateral detail; Content detail; Publishing detail; Related material; and Product supply (Register and McIlroy 2012, 29–30). Less needs to be changed, and therefore it is less likely that issues will occur.

With all of these advantages in mind, it is also important to consider that many of UBC Press’ trading partners still accept ONIX 2.1, and some may not yet accept ONIX 3.0. BookNet Canada explains that, although no longer supported, ONIX 2.1 is still being used, most widely in North America (BNC 2016). BookNet warns that eventually ONIX 2.1 will no longer be used, so it is necessary, at minimum, for “all publishers and data recipients to familiarize themselves with the changes” in ONIX 3.0 (BNC 2016). Graham Bell of EDItEUR notes that it is possible to run both ONIX 2.1 and 3.0 concurrently to meet the needs of various trading partners, and in this way the migration can be done as a gentler, safer, long-term project in which the new functionality of 3.0 can be updated and reflected incrementally (Bell 2015b). BookNet also advocates this implementation process, explaining that the benefit of running both systems can be that once trading partners are aware of the option, it “facilitates retailers and other aggregators developing ONIX 3.0 support” (BNC 2016). Though running two systems is more work, it is vital that metadata be delivered in the format that the data recipient needs and can use.

Staffing

It is essential that UBC Press hire a new digital projects manager (they are already in the process of doing so) to fill this newly open position as soon as possible, as this person will be responsible for all metadata distribution. Perhaps also having a metadata trainer come in and coach the applicable departments could be an option. As the Canada Council (and other funding bodies) is very interested in the possibilities of digital publishing, perhaps this trainer could be covered by a digital publishing grant. These types of grants are already available in the United States, such as Washington State Library's Metadata Enhancement and Remediation Grant, though this grant focuses on institutional metadata (Washington State 2016). UBC Press already takes part in BookNet's BiblioShare program, which validates and certifies metadata, but it is possible that they could also take part in other publisher metadata certification programs. Although at this point data is supplied from the Press to their distributors, and then from there to other channels, it may be useful to see how those distributors are doing in terms of metadata fitness through certification programs outside of Canada. There are programs through the BISG, the Publisher Certification Data Program; or through BIC, the Product Data Excellence Awards (which gives detailed feedback to publishers); or even through retailers, such as Barnes & Noble, which grade publishers based on metadata completeness, and also shares these records with their trading partners (Luther 2009, 11).

There is a huge need at UBC Press to hire an in-house copywriter. There are two major reasons for this. One is that the book descriptions are vitally important to the sale of Press titles, and freelancers have not worked out as well as the Press had hoped.⁷ Various issues with freelancers include lateness (especially because communication is done through email), differences in opinion of how a book should be described (the Press' vision versus the freelancer's vision), and the issue of budget (the freelancer must be paid regardless of whether the Press uses the copy or not). The Press needs to be in control of its own metadata, one of its largest assets for the sales and promotion of its

⁷ UBC Press staff members, in conversation with the author, July 7, 2016; July 8, 2016; July 21, 2016; and July 28, 2016.

titles, which has led to many of the production editors writing all of the copy for their books themselves. This is simply unsustainable during metadata crunches when, as mentioned earlier, each editor can be responsible for up to sixteen titles. If there was an in-house copywriter, copy could be written earlier and on a more staggered basis for each title rather than the pressure of other staff members adding more and more to their workflow. Another benefit is that the Copy Form, the main document used to facilitate the generation of different lengths of copy for each title, would only need to be sent out of UBC Press once. The only stakeholder outside of the organization that would need to sign off on the copy would be the author, instead of each party communicating through the middleman at the Press (currently the production editor). This would save time, potential stress of looming deadlines, and a lot of emailing. If any of the parties signing off on the form in-house; such as the acquisitions editor, the production editor, or the marketing manager had a change that needed to be made, they could simply talk to the copywriter and have the issue solved, rather than having to send emails to a freelancer and waiting for them to make the changes remotely.

Organization and Workflow

At UBC Press, the production department is primarily responsible for the metadata creation process. This happens because at the metadata deadline, the production department is most familiar with the manuscripts and therefore has a good understanding of how the book should be described. In addition, UBC Press is more editorial than market-driven (as are many university presses). At other publication houses, it is more common for the marketing department to create and manage metadata, as it is a selling tool, and as such, it has been recommended to “encourage both marketing and editorial department input on metadata” (Ruffilo 2011). Graham Bell of EDItEUR has commented that “some publishers are ... recognizing [that] metadata is now the key element of their business-to-consumer communication, and [are] shifting metadata ownership from editorial or production into sales and marketing” (Bell, qtd. in Register and McIlroy 2012, 65).

Ideally, at the Transmittal, all departments would contribute to Klopotek: acquisitions with comparable titles; marketing with author information, subject codes,

and sales points; and production with a manuscript assessment and specs. If this interdepartmental view became a reality, a clear, visual, charted workflow should be created, a process map that shows “the flow of information, recording details such as what personnel or department inputs data, handoffs, data repositories, and to whom metadata is exported. A process map may reveal problem areas such as disconnects, bottlenecks, redundancies, rework loops, and decision delays” (Warren 2015). Basically, any type of restructuring in terms of metadata workflow would need to take into account who inputs the data, where it goes, and to whom, because “how it’s done in each organization has implications for workflow and staff responsibility, and it’s important to discuss this across divisions and work groups” (Register and McIlroy 2012, 57). This does not mean that everyone needs to go into Klopotek and add information, which might in fact confuse things, but that each department deliberate and come up with information that can be added to Klopotek by one dedicated staff member, most likely the digital projects manager.

There should also be more onus on each author to be timely with the metadata they are responsible for, such as title decision, which can be quite a long process. As mentioned earlier, one of UBC Press’ major selling points is the quick production timeline of eight to ten months. Instead of lengthening this schedule due to metadata crunches, perhaps instead a slightly more flexible schedule could be put into place and this could be articulated in the author’s contract (timelines are currently not present in the contract, but rather communicated by the acquisitions editor). For example, the book could be on an eight-month schedule, with a ten-month option if certain elements are late. This could mean that if the associated metadata for a title cannot be agreed upon (say, the descriptive copy) in a reasonable amount of time for the next metadata deadline, the book will be pushed to the next season. This would make the author directly responsible for delays in publication dates, and thus more likely to give as much information as they can on each of the forms that are used to collect metadata, and are more likely to respond faster when it comes to making comments or signing off on marketing copy or other decisions.

It may also be worthwhile to take a deeper look into Klopotek and see if it can facilitate the production of some of the various forms that different departments are

working on. The List Schedule, Transmittal Form, Copy Form, and Author Marketing Questionnaire currently share certain information, yet each form must be filled out manually. Klopotek has the ability to manage a lot of information, and it seems likely that it should be able to reorganize and redistribute information on the forms that are used to collect metadata, considering that it currently auto-fills information from the database into catalogue templates, brochures, and order forms. The forms mentioned above could all be completed faster by having bibliographic title information filled in automatically from Klopotek. This way, the metadata can be entered once, used for multiple documents, and changed only when necessary, resulting in saved staff time.

The Press should, at all costs, try to always be the source from which all other trading partners receive metadata. ONIX records are not simply released into the world and displayed exactly as they are sent on other websites; instead, “a vendor or retailer record may consist of elements from many sources. For example, Amazon and other major online sellers accept data feeds from all publishers with titles listed on the site, plus multiple feeds from data aggregators such as Bowker and Ingram” (Register 2013, 41). At UBC Press, all applicable metadata is entered into Klopotek, which then generates ONIX files for a variety of channels. It sends data to Bowker in the US and to BookNet in Canada. An ONIX file is also sent to CoreSource (owned by Ingram), a digital asset manager that stores UBC Press ebooks and their associated metadata. Klopotek also sends data to the University of Toronto Press (UTP), UBC Press’ Canadian print book distributor, and the University of Washington Press (UWP) in the United States.

Problems arise, however, when these supply chain partners add to and modify the files after they have been sent out. Bowker and BookNet ingest the data and alter it to fit their proprietary form. From CoreSource, e-vendors and ebook aggregators (for both retail and library markets) such as ProQuest, a library wholesaler, add MARC records as well as adjust the ONIX files to meet their requirements. UTP does even more modification; as they take UBC Press ONIX files and pare them down to fit their simple metadata system, as well as adding information, such as publishing status and availability. UWP makes similar modifications for their national market. These print distributors then send their versions of the ONIX files to retailers (bricks-and-mortar and

online), wholesalers, and BookManager (from which indie stores purchase books). That is a lot of modification to the original files. One way to ameliorate this problem to some degree is to send the metadata from the Press directly to as many partners as possible, rather than relying on distributors to do it (Tallent 2016). Every effort should be made to have less downstream modification of metadata, and one way to do that is to ensure that upstream metadata “becomes good enough and is updated frequently enough” to reduce the need for supply chain intervention (Register 2013, 43). That being said, metadata that UBC Press does not have control over, for example supply detail by their distributors, or MARC records for libraries, must be added by downstream partners and are vital to business operations.

All of these recommendations play into one factor: better integrating metadata collection, generation, and distribution into regular workflow, one that captures and maintains everything about a book. It is important to remember, “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” (Register and McIlroy 2012, x). The metadata process needs to be as automated as possible by generating more from the same information, as data entry kills productivity.



Chapter 5. Conclusion

Mike Shatzkin, digital publishing expert, states rather ominously that “metadata work for publishers is, really, a bottomless pit, since it is, in effect, ‘information about the book’ and there is no limit to that. There will be no end to the categories of quality, interest, and association each book can have attached to it” (Shatzkin 2010). In a less menacing way, Laura Dawson assures us that the best quality a publisher can have in the digital age is that of flexibility, and that having “different audiences for metadata” is necessary, as “not everybody needs to know the same thing about a book. Much as there’s no one way to describe an elephant, there’s no one way to describe a book. Developing the workflows that capture and maintain the range of descriptions that ‘describe a book’ will be critical in a world in which ‘discovery’ increasingly means ‘found it online’” (Dawson 2012).

Deeper than just describing a book as an entity, metadata is getting so complex and deep into a text that it can now be associated all the way down to the chapter or even *paragraph* level, known as “granular metadata.” This granular metadata is very useful for selling individual chapters of books, and thus publishers “have begun to assign metadata [in their backlist titles] to elements below the title level to provide greater access to book chapters and to prepare for future revenue opportunities” (Luther 2009, 2). In fact, in its 2014 self-study, UBC Press noted that “digital processes facilitate the process of delivering books in parts (chunks). This is particularly attractive for making material available for courses and electronic course packs” (Press review 2014, 41). Indeed, adding granular metadata to UBC Press books makes sense, as the “specialized needs of scholarly and professional readers and the rigorous selection criteria of academic libraries make it important that the metadata is as complete and granular as possible” (Register and McIlroy 2012, 6–7). It would definitely help scholars and students find the information they are looking for faster and with more accuracy. But there is one big problem created when books can be chopped into pieces and delivered

separately. How do you price them? Is it based on word count? In the case of a chapter in a collection, is it based on the author? And of course, there is also a ton of work involved in the creation of metadata per section or (God forbid) paragraph, so much so that the monetary benefit would have to significantly outweigh the costs.

On the horizon for publishers and metadata is the potential of “linked data.” What this means is basically better use of data through linking, and “refers to a way of connecting related data on the Web through linking pieces of data across the current silos created by individual web pages and discrete databases” (Register 2013, 42). As it stands now, each book record is updated individually, and thus when one record is updated, nothing is done to the others; for example, “if review information is added to the paperback version of a book, it doesn’t automatically update a hardcover version with the same content. If contributor biographical or award information is added to a record, it doesn’t automatically get added to the records associated with that contributor” (Register 2013, 43). Instead of duplicating every record for each book in multiple databases, all are linked together, so that “in a linked data model, different metadata elements referring to the same ‘thing’ can be created and/or maintained separately and be linked together as needed to describe the ‘thing’ rather than always staying together in one stand-alone record” (Register 2013, 43). As a result, downstream updates to records could in fact benefit upstream players (publishers), as well as saving time from updating multiple records.

In *The Metadata Handbook*, the authors interviewed publishing industry stakeholders and asked what trends they could see happening in the future, and one of the common themes was the push for “open data” and the belief that all industry stakeholders should adopt systems that “support the potential of shared metadata, linked data, and increased metadata exposure” (Register and McIlroy 2012, 62). One important development is the intention of libraries to move away from MARC records, which is a huge step towards making ONIX files more useful – and for the sharing of metadata. MARC is very specialized and less extensive than ONIX, as it focuses more on bibliographic detail than descriptive content. Stakeholders such as EDItEUR have worked toward crosswalking ONIX and MARC, but currently this still requires a lot of duplication (Register and McIlroy 2012, 32). In 2011, the Library of Congress announced

that they would be moving away from the MARC 21 standard – which was introduced in 1997 – with other national libraries following suit, including Canada. Even more promising, “in 2012, [the Library of Congress] ... announced a strong focus on transitioning to a Linked Data model ... [which] indicate[s] a greater consideration of integration and interactivity between library and publisher metadata” (Register and McIlroy 2012, 20).

Another indication towards the overall cohesion of sharing and discoverability is the development of the International Standard Text Code (ISTC). Established in 2009, the ISTC is similar to an ISBN, but instead of identifying a single product, it identifies a work, “regardless of the publisher or format of a particular product.” So instead of a code for one copy of *Moby-Dick* published by Penguin Classics, one single code is issued to every published copy of *Moby-Dick* (Register and McIlroy 2012, 25). The purpose of the ISTC is “to enhance the effectiveness of metadata and improve discoverability in a search environment. Internet retail booksellers and wholesalers find it valuable to expose different versions of the same title in search results” (Register and McIlroy 2012, 25). The same is being done with authors with the International Standard Name Identifier (ISNI) which is helpful for authors with pseudonyms, corporate authors, but also just for authors with their names represented in different ways (such as with or without a title, like “Dr.”, or with or without a middle name or initial). Both are useful for the end user, and are gaining acceptance as ONIX 3.0 supports both, but some publishers may push back against the trend because both identifiers promote discoverability of all books rather than just those by one publisher (and it creates more work for them). The benefits outweigh the drawbacks though, as “exploiting the potential of identifiers to link bibliographic metadata, including names ... ultimately benefits all through increased discovery, a better user experience, and efficiencies in metadata management across the supply chain” (Register and McIlroy 2012, 62).

In a world where metadata is becoming dangerously close to replacing actual books, it is important to remember that “a product is something a consumer only experiences in usage,” thus the event of reading a book can never be replaced (Kircz 2007, 558). Metadata has its importance, though, in this event, as descriptive metadata that reflects the knowledge and experience of a book will be “properly wrapped into

metadata ... [and] unwrapped by the consumer. In other words, the metadata wrap or shell is essential for the ultimate capacity of consumption of the product” (Kircz 2007, 558). It is the metadata creator’s responsibility to make sure that the ethos of the book is conveyed in its description, because that is how a consumer will be moved to purchase and read. Providing the consumer with rich, appropriate, and accurate information gives them every tool they need to make an informed buying decision, because “with no physical item, metadata *is* the user experience” (Register and McIlroy 2012, 43, emphasis in original).

References

- AAUP (Association of American University Presses). 2015. Digital book publishing in the AAUP community: Survey report. New York: AAUP.
- Bell, Graham. 2015a. Making cents of subjects. London, UK: Book Industry Communication.
- . 2015b. Navigating the transition from ONIX 2.1 to 3.0. Webinar, January 26. Toronto: BNC. <https://booknetcanada.wistia.com/medias/1nrl6owlh3>.
- BISG (Book Industry Study Group). 2013. Best practices for product metadata: Guide for North American senders and receivers. Developed in cooperation with BookNet Canada. New York: BISG.
- BNC (BookNet Canada). Canadian bibliographic standard. 2007. Toronto: BNC Bibliographic Working Group.
- . The Canadian book buyer 2015. 2015. Toronto: BNC Research.
- . ONIX user documentation. 2016. Toronto: BNC Staff Wiki. <https://booknetcanada.atlassian.net/wiki/display/UserDocs/ONIX>.
- BNC (BookNet Canada) and the Canadian Bookshelf. Increase sales and lower costs with better metadata: A step-by-step ONIX guide for publishers. 2010. Toronto: BNC Research.
- Brand, Amy, Frank Daly, and Barbara Meyers. 2003. Metadata demystified: A guide for publishers. Sheridan, WY, and Baltimore, MD: Sheridan Press and NISO Press.
- Breedt, Andre, and David Walter. 2012. White paper: The link between metadata and sales. Surrey, UK: Nielsen BookScan.
- Dawson, Laura. 2012. What we talk about when we talk about metadata. In *Book: A futurist's manifesto*, ed. Hugh McGuire and Brian O'Leary. Montreal: O'Reilly Media.
- EDItEUR. 2016. ONIX for books. <http://www.editeur.org>.

- Greenfield, Jeremy. 2012. Nearly 100% of publishers have seen e-booksellers get their metadata wrong. *Digital Book World* blog entry, May 3. <http://www.digitalbookworld.com/2012/nearly-100-of-publishers-have-seen-e-booksellers-get-their-metadata-wrong/>.
- Jackson, Millie. 2008. Using metadata to discover the buried treasure in Google Book Search. *Journal of Library Administration* 47, 1–2: 165–73.
- Kircz, Joost. 2007. Creation-driven marketing: Integrating metadata into the production process. *New Library World* 108, 11–12: 552–60. <http://dx.doi.org/10.1108/03074800710838281>.
- Lichtenberg, James. 1997. Metadata for the millennium: Publishers must begin now to reshape their business for the new century. *Publishers Weekly*, April 21: 44–46.
- Luther, Judy. 2009. White paper: Streamlining book metadata workflow. Baltimore, MD, and Dublin, OH: NISO and OCLC.
- Moffat, Malcolm. 2006. Marketing with metadata: How metadata can increase exposure and visibility of online content. *New Review of Information Networking* 12, 1–2: 23–40. <http://dx.doi.org/10.1080/13614570601136248>.
- Register, Renée. 2013. The importance of metadata for e-content. *eContent Quarterly* 1, 1: 30–44.
- Register, Renée, and Thad McIlroy. 2012. *The metadata handbook: A book publisher's guide to creating and distributing metadata for print and ebooks*. Columbus, OH: DATAcurate.
- Reid, Calvin. 2010. Accurate metadata sells books. *Publishers Weekly*, July 5: 4.
- Ruffilo, Nick. 2011. Five degrees of metadata: Small changes can mean big sales. *Publishers Weekly* blog entry, January 24. <http://www.publishersweekly.com/pw/by-topic/columns-and-blogs/soapbox/article/45844-five-degrees-of-metadata-small-changes-can-mean-big-sales.html>.
- Shatzkin, Mike. 2010. Metadata is the new most important thing to know about. *Shatzkin Files* blog entry, June 8. <http://www.idealogy.com/blog/metadata-is-the-new-most-important-thing-to-know-about/>.
- Sim, Chris. 2016. Who uses keywords in metadata? *Digital Book World* blog entry, March 4. <http://www.digitalbookworld.com/2016/who-uses-the-keywords-in-metadata/>.
- Smith, Mason. 2013. Dude, we need to talk. Ingram presentation at AAUP conference, Boston, June 21. <https://vimeo.com/70627705>.

Tallent, Joshua. 2016. What your metadata does when you're not looking. Webinar, June 14. New York: BISG.

Warren, John. 2015. Zen and the art of metadata maintenance. *Journal of Electronic Publishing* 18, 3. <http://dx.doi.org/10.3998/3336451.0018.305>.

Washington State Library. 2016. Current grant cycles. <https://www.sos.wa.gov/library/libraries/grants/grants.aspx>.

W3C (World Wide Web Consortium). 2015. Digital Publishing Interest Group Metadata Task Force report. January 8 version. <http://www.w3.org/TR/2015/NOTE-dpub-metadata-20150108/>.

Unpublished Documents Consulted

AIF (Author information form). 2011. Vancouver: UBC Press internal document.

AMQ (Author marketing questionnaire). 2015. Vancouver: UBC Press internal document.

Bibliographic file distribution. 2011. Vancouver: UBC Press internal document.

BIF (Book information form). 2015. Vancouver: UBC Press internal document.

Coates, Laraine. Metadata presentation notes. 2016. Presentation delivered on May 12, Vancouver.

Copy form. 2009. Vancouver: UBC Press internal document.

-. 2015. Vancouver: UBC Press internal document.

Indigo buying dates. 2015. Vancouver: UBC Press internal document.

Spring 2017 list schedule. 2016. Vancouver: UBC Press internal document.

Manuscript transmittal form. 2015. Vancouver: UBC Press internal document.

Metadata deadlines. 2016. Vancouver: UBC Press internal document.

Metadata distribution. 2013. Vancouver: UBC Press internal document.

Press review: Self-study. 2014. Vancouver: UBC Press internal document.

Sales figures: April 2016 to July 2016. 2016. Vancouver: UBC Press internal document.