E-Book Bibliographic Metadata Requirements in the Sale, Publication, Discovery, Delivery, and Preservation Supply Chain

A Recommended Practice of the
National Information Standards Organization

Approved: January 29, 2022

About NISO Recommended Practices

A NISO Recommended Practice is a recommended "best practice" or "guideline" for methods, materials, or practices in order to give guidance to the user. Such documents usually represent a leading edge, exceptional model, or proven industry practice. All elements of Recommended Practices are discretionary and may be used as stated or modified by the user to meet specific needs.

This recommended practice may be revised or withdrawn at any time. For current information on the status of this publication contact the NISO office or visit the NISO website (www.niso.org).

Table of Contents

Foreword

About this Recommended Practice

As we write this recommended practice, e-books are still evolving as technology and as cultural artifacts. The format, transmission, and applications of book metadata are changing. Metadata workflows and data exchange are becoming ever more automated, massive in scale, and networked. In this fast-changing environment, it is crucial for the various producers and users of e-book metadata—publishers, retailers, libraries, service providers, and preservation agencies—to develop awareness and become conscious of one another's standards, practices, and purposes.

The working group that developed this recommended practice benefited immensely from having representatives from these different types of stakeholders. Our first lesson in this project was the recognition of how little we understood about each other's work. The saying, "Librarians are from MARC; publishers are from ONIX," emerged as a useful shorthand. Our second lesson was the discovery of how many issues stem from this mutual ignorance. In many cases, decisions about metadata that seem trivial to one party cause serious problems for another and generate errors that are perpetuated across the metadata network. It became very clear to us from the experience of our own collaboration that the purpose of this recommended practice must be to create a shared understanding and, wherever practical, an alignment of practice among those of us who work with e-book metadata, to the benefit of all readers, researchers, and authors.

The proposal approved by the NISO Content and Collection Management Topic Committee (now the Information Creation & Curation Topic Committee) in May 2016, and approved by NISO Voting Members in July 2016, provided the impetus for this project. The original statement of work lists six components covering desired outputs:

1. Describe the minimal metadata requirements necessary to describe e-books in order to support sales, discovery, delivery, deaccessioning, and preservation.

2. Identify the most effective and efficient way for metadata to be moved through the entire supply chain.

3. Address how metadata records and transfer of information may be used to describe updates to metadata records and record sets.

4. Develop rules for matching e-book metadata from multiple sources for deduplication purposes.

5. Provide examples of recommended practice using common metadata standards.

6. Explore ONIX and MARC tools and identify if it is necessary to issue recommendations around conformance tools.

Several aspects of the original statement of work proved to be too ambitious for our working group. We touch on all of the proposed items except number 6, because we believe that evaluating conformance tools for the ONIX and MARC standards is better left to the very active communities of practices that exist for these standards. As we continuously learned about each other's workflows, we also realized that some workflows are bound to constraints associated with the creation of new titles and to limitations caused by the large number of participants involved in the making of e-books as objects. It became clear that trying to enforce "ideal" workflows would not be a realistic mission and could carry the risk of

making our recommendations too detached from the actual reality of e-books. Instead, we present recommendations that identify best practices within existing standards and workflows.

NISO Topic Committee Members

The Information Creation & Curation Topic Committee Topic Committee had the following members at the time it approved this Recommended Practice:

Sharon Farnel (*co-chair*)
University of Alberta Libraries

Patricia Feeney
Crossref

Stephen Flockton
IOP Publishing

Mark Heaver
Taylor & Francis Group

Cyndi Hernandez
ProQuest

Peter McCracken
Cornell University Library

Anna Neatrour
University of Utah

Kennett Rawson (*co-chair*)
IEEE

NISO Working Group Members

The following individuals served on the NISO E-book Bibliographic Metadata Requirements in the Sale, Publication, Discovery, Delivery, and Preservation Working Group, which developed and approved this Recommended Practice:

Rebecca Culbertson
University of California, San Diego

Ravit H. David (*co-chair*)
Scholars Portal, University of Toronto

Jon Elwell
EBSCO Books

Amy J. Kirchhoff
Portico

Alistair Morrison (*co-chair*)
Johns Hopkins University

Frank McDermott
John Wiley & Sons, Ltd.

Sara Newell
OCLC Online Computer Library Center

Bridget Page
EBSCO Information Services

Patricia Payton
ProQuest

Paul Swanson
Minitex

Joshua Tallent
Firebrand Technologies

Jodi Williamschen
Library of Congress

Acknowledgements

Trademarks, Services Marks

Wherever used in this standard, all terms that are trademarks or service marks are and remain the property of their respective owners.

Section 1: Purpose and Scope

The overarching goal of this recommended practice is to promote shared understanding and ongoing communication about the most fundamental e-book metadata elements among all the different types of stakeholders. This recommended practice provides a guide to creating and using e-book metadata as consistently as possible across the different types of stakeholders, who must also adhere to the existing standards and communities of practice applicable to them.

An effort to improve communication and consistency is necessary and timely. Individual organizations generally have consistent practices internally. Pairs of organizations that regularly exchange metadata generally understand each other's guidelines well enough to translate and communicate effectively. However, the scale of metadata exchange has expanded to massive numbers of records, along complex non-linear workflows, across networks of diverse organizations, and over significant periods of time. Taking a broader view to encompass this new reality, we found ambiguity and confusion over even the most common metadata fields and very different practices for seemingly simple workflows, such as announcing obsolete titles.

Five Essential Metadata Elements

We attempt first to articulate the requirements of different stakeholders and different use cases involving the same metadata elements, and then to reconcile or adjudicate among them. The goal is to develop a shared understanding of specific metadata elements among the different types of stakeholders and across the different use cases. We have narrowed the scope of our effort to five metadata elements that occur in all the major standards and that are indispensable to the most important workflows across the different types of stakeholders and the different phases of the e-book lifecycle:

- Titles
- Names
- Dates
- Book identifiers
- Subjects

Three Additional Attributes

In addition, the treatment of identifiers necessitates recommendations for three metadata elements that are inextricably involved in best practices for applying identifiers to e-books:

- Format
- Constraints on use
- Uniform resource identifier (URI) or Internationalized Resource Indicator (IRI)

Three Fundamental Functions of E-Book Metadata

This short list of metadata elements is the minimum set required to enable three basic functions of e-book metadata that apply across all stakeholder groups:

- Identifying a book in title lists, holdings records, and other record sets.
- Matching records that refer to the same book or versions of the same book.
- Distinguishing records that refer to different books or different versions of the same book.

A complete accounting of the uses of these metadata elements in the major metadata standards was compiled, along with discussions of actual uses as opposed to recommended uses. The interpretations and specific uses of these core data elements revealed some important inconsistencies among the different types of stakeholders. For example, a publisher that digitized an older print title from its backlist might use the publication date field to show when the e-version was published online because of the dictates of its publishing system or sales practices (or online platform), while a library would prefer the original (print) publication date.

E-book metadata standards must support very specific purposes and data exchange protocols such as ONIX and MARC, but must never render the metadata record ambiguous or unusable for other purposes and protocols. Where possible, a common definition of a specific datum should be shared by all providers, consumers, and contributors.

Section 2: Landscape

2.1 The Ecosystem of E-books

The landscape under consideration by the working group encompasses all agencies that exchange metadata related to e-books. The working group sought to examine all phases in the life cycle of an e-book title, from its prepublication stage through its status in an archive for preservation purposes. Included in the examination were e-book titles that may never have been nor ever will be published in print, but will be published online only. To cover all phases, the working group was comprised of representatives from e-book publishers, aggregators, vendors, jobbers and distributors, librarians, digital preservation specialists, and others involved in the management and transmission of e-book metadata.[1] Notably missing from this list are authors, editors, readers, researchers, and consumers. We have relied on the listed stakeholders to represent the interests of these most important users. Also notably missing are Google, Bing, Baidu, and other web search engines or reference services that provide the first route to discovery for many readers and researchers.

The library practices considered in this recommended practice are, unfortunately, limited to the English-speaking, Anglo-American point of view, while the service providers, publishers, and aggregators include global companies that serve those libraries. We hope that future work on this recommended practice can extend it beyond these limits and address different conventions in library practices and the challenges of non-Latin character sets.

Finally, we wanted to build our recommendations on the current, existing landscape to make it a practical guide and not merely a utopian vision of the e-book industry. In order to do that, we have familiarized ourselves with and acknowledged the work on e-books completed by institutions and organizations that are closely engaged in creating best practices and guidelines for e-books, such as EDItEUR, BISG and W3C, to mention but a few.[2]

2.2 Conceptual Framework

Any discussion of bibliographic metadata falls quickly into confusion unless there is a shared understanding of what a book is. A conceptual framework is needed to distinguish and keep straight the different possible meanings of "a book." For our purposes, the essential distinctions were made in the Functional Requirements for Bibliographic Records (FRBR) model that was introduced by the International Federation of Library Associations and Institutions (IFLA) in 1998.[3] FRBR articulates the types of concepts and distinctions that are used in modern metadata models, including Resource Description and Access (RDA) cataloging practices, BIBFRAME 2.0, and, most recently, the FOLIO Project Codex Metadata Model.[4,5,6] The IFLA Library Reference Model transfers the Resource Description and Access (RDA) from FRBR to the Library Reference Model (LRM) and is the consolidation of the separately developed IFLA conceptual models FRBR, Functional Requirements for Authority Data (FRAD), and Functional Requirements for Subject Authority Data (FRSAD).[7] It is not necessary to apply the FRBR or the LRM model to our discussion exactly, but it is necessary to bear in mind the sorts of concepts and distinctions that they and other modern metadata models make.

Work is "a distinct intellectual or artistic creation. A work is an abstract entity; there is no single material object one can point to as the work. We recognize the work through individual realizations or expressions of the work, but the work itself exists only in the commonality of content between and among the various expressions of the work."

Expression is "the intellectual or artistic realization of a work in the form of alphanumeric, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such forms." This concept is omitted by some metadata models.

Manifestation (or instance) is "the physical embodiment of an expression of a work ... Manifestation represents all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form." For our purposes in this recommended practice, a manifestation can be a digital embodiment, too, and it corresponds to the product sold by the publisher or aggregator on a specific platform or in a specific format.

Item "is a single exemplar of a manifestation. The entity defined as item is a concrete entity ... In terms of intellectual content and physical form, an item exemplifying a manifestation is normally the same as the manifestation itself."[8]

Section 3: Common Issues with the Current Workflows

The working group spent a lot of time becoming familiar with current workflows in order to identify issues that go beyond specific metadata fields since, as already mentioned above, it was realized early that stakeholders have very different understandings and ways of executing seemingly simple workflows. We think it is highly recommended for anyone who works with e-books to familiarize themselves with some of the challenges that are unique to the e-book ecosystem, and so we decided to include them as use cases that members of the group imparted during our discussions. We also thought it would be beneficial to elaborate on a few of the most pertinent issues to e-book workflows; namely, selling books before they are published, uncoordinated provision from multiple sources, and the use of different identifiers on different title lists.

3.1 Selling Books Before They Are Published

Publishers have a need to market and take advance orders for forthcoming titles. Retailers, consumers, service providers, and libraries have a need to assess or make advance purchases of forthcoming titles. However, the forthcoming title list is a moving target. Many publishers plan for the release of books more than a year in advance of the authors and editors completing their work. Inevitably, some of the metadata for these books gets shared with retailers, service providers, and libraries and then later must be revised to reflect changes in titles, contributors, and content, as well as changes or delays in publishing a particular book or abandonment of the project. Correcting metadata once it has been shared is far more challenging than distributing it in the first place. A typical publisher metadata workflow (Figure 1) may bypass the publishing system to create title lists for the sales and marketing team, which results in promulgation of information that is very likely to change and with no good mechanism to change it. Once a sales brochure or a title list file is given to a prospective customer, publishers do not have a way to update it or withdraw the information.
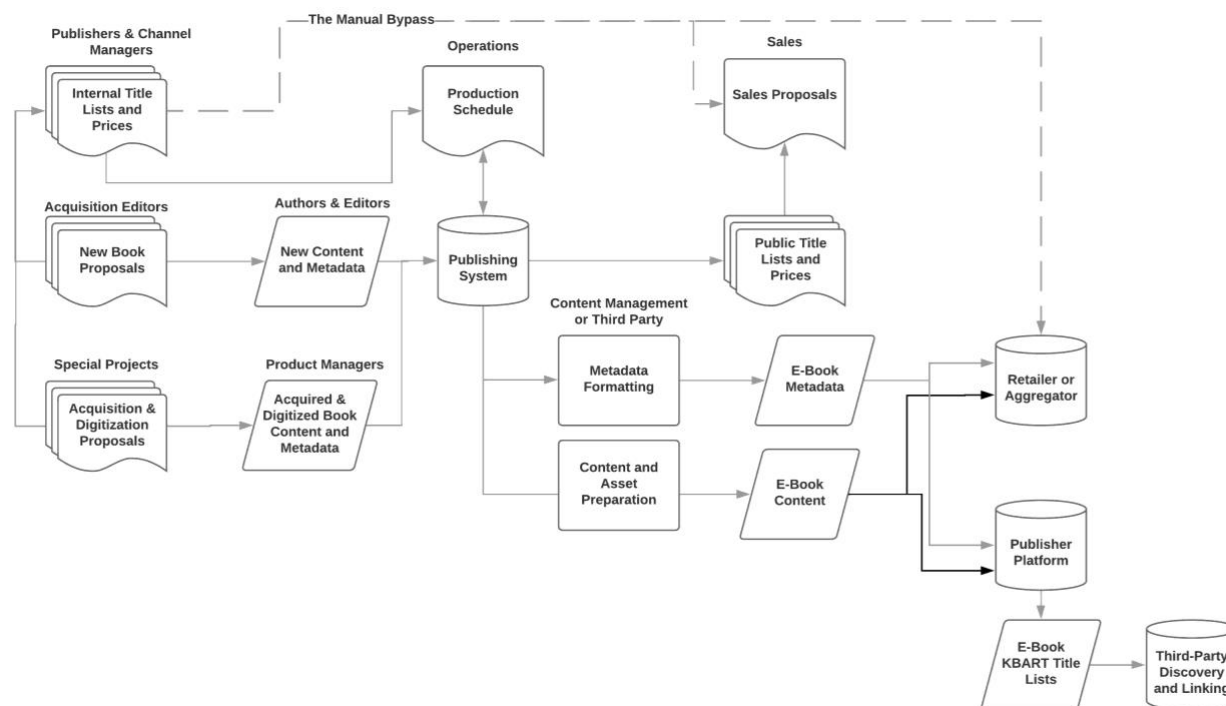
Figure 1: E-Book Publisher Internal Metadata Workflow

3.2   Uncoordinated Provision of Metadata from Multiple Sources

Discussions about metadata supply chain may be misleading, because the flow of metadata for e-books is never a linear process but more of a network, with metadata records passed around and modified by many agents. Retailer databases, library catalogs, and discovery services combine metadata from different sources and modify records to meet local requirements. The resulting record is a conglomeration, usually with no provenance for the particular data elements. In such scenarios, updates and corrections are risky. In the illustration below, the publisher provides a basic metadata record to the retailer, but the retailer also contracts with a data aggregator who enriches that record by adding subject headings, cover images, and so on. Suppose that a publisher sends Version 1 of a metadata record to both the aggregator and retailer, and the aggregator does its work and sends along an enriched record based on Version 1 to the retailer a month later as part of a regular batch delivery. However, in the meantime, the publisher discovers an error and sends a corrected Version 2 of the record to both the aggregator and the retailer. The retailer, of course, loads Version 2 into its system and awaits the enriched version from the aggregator. When the enriched version arrives, however, it is based on Version 1 and does not include the publisher's corrections. If the discrepancy is a correction to something as basic as the book title or the spelling of the author's name, reverting to the uncorrected version can cause major problems.
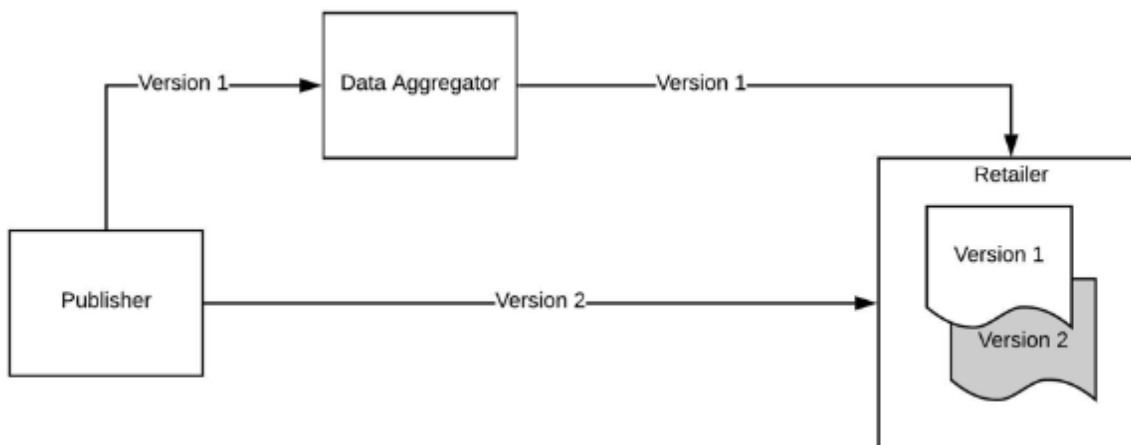
Figure 2: Publisher's Corrected Version of Metadata Is Overwritten by Older Version from Aggregator

The same sorts of problems arise when libraries process records into their catalogs or other systems, and publishers later correct them. Without knowing which data elements have come straight from the publisher and which have been modified, processing an update is a risk.

3.3   Different Identifiers on Different Title Lists

Whether in the form of a record set or a title list, e-book metadata must be exchanged in large volumes using automated processes as much as possible. Title lists and record sets control sales forecasts, marketing programs, collection management, library services such as purchasing programs, shared collection development, catalog maintenance, discovery, citation linking, and preservation.
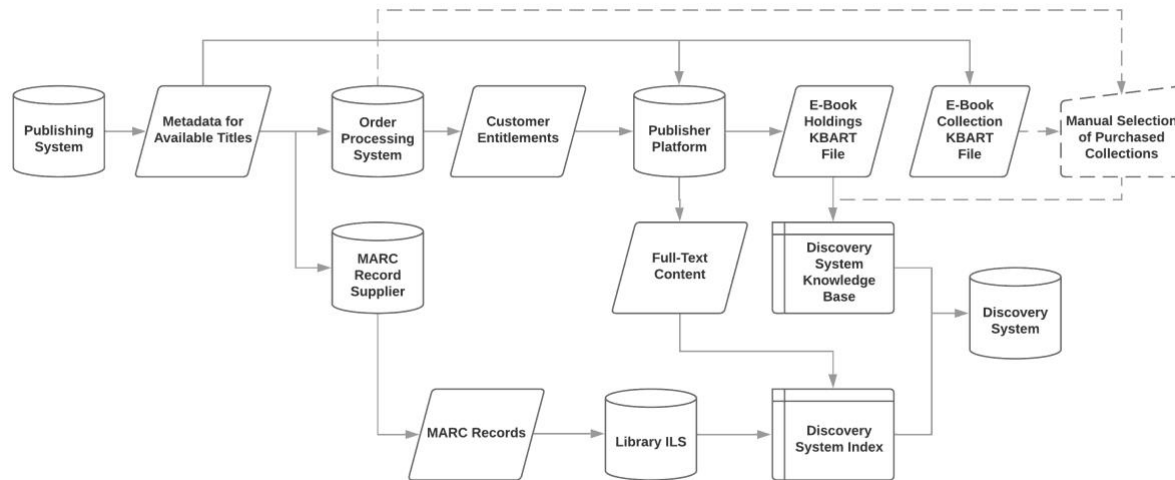
Figure 3: E-Book Publisher Discovery Metadata Flow

Often, in these scenarios, the library holdings are compared to offerings from e-book sellers or service providers. For example, libraries that rely on a library acquisition service like GOBI Library Services need to communicate which books they already hold and therefore don't wish to purchase again. Title lists are commonly a resource for e-book identifiers, most notably International Standard Book Numbers (ISBNs), and it is a common practice to use ISBNs to identify titles, compare holdings, and perform other related activities. And yet, it appears that, although ISBN has been around since the 1960s, the practice of using it varies across the board and is a great frustration to anyone who deals with e-book holdings. Since e-book distribution is a complex network of many operating parts when title lists are created by one of the stakeholders, they may not contain the same ISBNs used by another stakeholder, thus making ISBN unreliable for title list comparisons. The working group rejected the idea that ISBN or any other book identifier can be made entirely reliable, and recommends consistent handling of titles, names, and dates to provide additional data points for comparison.
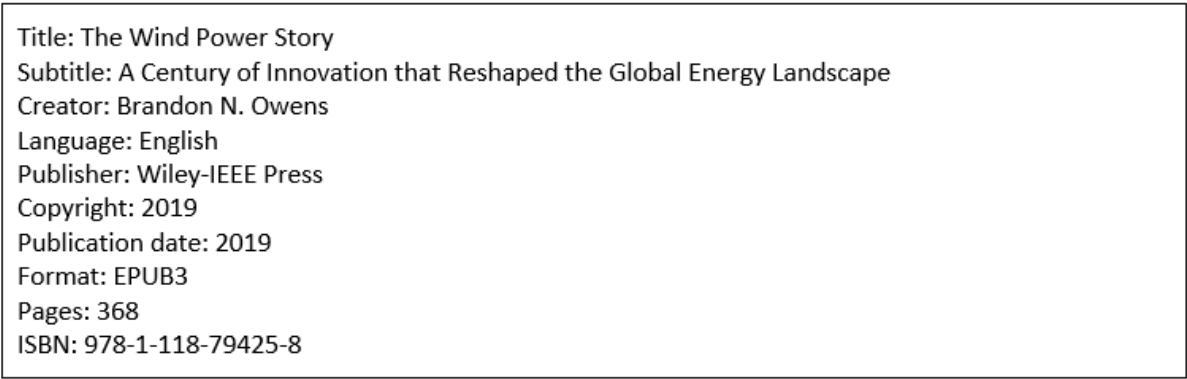
Section 4: General Recommendations

4.1 An E-book Should Be Able to Identify Itself to Human Readers with a Title Page Verso

Publishers should carry over from print production the convention of including a title page at the beginning of an e-book. The title page is an integral part of the e-book, and the metadata presented there serves as convenient, unambiguous confirmation of the corresponding elements in a metadata record. The title page and title page verso of a print book should never be omitted when that book is scanned or otherwise digitized to create an e-book.

This page should be visible to the human reader of the e-book on all platforms and devices and should include the canonical expression of the core metadata fields. Field labels should be used to avoid ambiguity, especially around dates and different types of titles and subtitles.

The metadata values shown on the title page should match the metadata in all records for the e-book.

```
Title: The Wind Power Story
Subtitle: A Century of Innovation that Reshaped the Global Energy Landscape
Creator: Brandon N. Owens
Language: English
Publisher: Wiley-IEEE Press
Copyright: 2019
Publication date: 2019
Format: EPUB3
Pages: 368
ISBN: 978-1-118-79425-8
```

Figure 4: Mockup of Proposed Title Page Verso Design for an E-Book

The title page verso should be the authority for the metadata elements that it contains.

In cases where print backfiles have been sold from one publisher to another, and the new publisher scans the print version to make a digitized version, there will be a discrepancy between the publisher listed on the image of the original print title page and the current copyright holder. In this case, the new publisher should attach a new title page at the beginning of the book to represent the metadata for this new e-book manifestation of the work. There would also be a new ISBN for the new e-book manifestation produced by the new publisher.

4.2 An E-Book Should Be Able to Identify Itself to Software Applications

The best exemplar of this principle is the EPUB 3 specification, which coordinates the structure of e-books and the functionality of e-readers.[9] It includes a small set of Dublin Core metadata fields that all EPUB packages are required to include and all EPUB reading systems are required to interpret:

- dc:identifier
- dc:title
- dc:language
- dc:date

Additional optional metadata elements include:

- dc:contributor

- dc:creator

- dc:subject

- dc:type

It is strongly recommended that publishers follow the spirit of the EPUB 3 specification by standardizing the way machine-readable metadata are embedded in e-book files, and follow its guidance on the minimum required metadata elements to be delivered, even for other file formats such as PDF, XML, or HTML that are interpreted by other types of systems such as PDF readers and web browsers.

4.3 An E-Book Should Be Discoverable Online

Discovery systems used by publishers, retailers, and libraries depend on the quality and clarity of the metadata provided to them. It is vital that all participants in metadata provision and refinement keep the search engine in mind. The basic metadata elements discussed in this recommended practice should always be present, because they are indispensable to the search, filter, and browse features that researchers and customers require in the search engines they use. Consistent definition and handling of these core metadata elements are also important.

Web-based discovery via Google and other search engines is vastly improved if the core metadata elements are tagged using schema.org. Whenever bibliographic information is presented on a public web page, schema.org tags should be used to promote proper interpretation of the metadata and improve web-based discovery.

4.4 An E-Book Should Be Discoverable Within Purchasing Systems

Wholesalers offer library purchasing systems that provide options for searching, selecting, and ordering electronic books for libraries. These systems rely on basic metadata elements discussed in this document to link print and e-book formats and to identify reprints separately from new titles. This linking is crucial for identification of titles to be put into approval plans, standing orders, and demand-driven acquisition pools.

4.5 Metadata Records Should Contain Record Attributes

To facilitate audit and reconciliation of multiple records for the same e-book, e-book metadata records should include basic record attributes:

- Record ID

- Record created date

- Organization that created the record

- Record update date

- Organization that last updated the record

- Version control number, if the creating or updating organization uses one

Record attributes must be kept distinct from e-book metadata fields. For example, both the record created date and the e-book publication date fields should be populated even if they have the same value. Again, the EPUB 3 specification provides a good example. It requires a unique identifier for the EPUB

publication and a release identifier (composed of the unique identifier plus the last-modified date).[10] This allows for the unique identifier of the EPUB publication to persist across minor changes in content or formatting, while differentiating these versions by a date stamp—an approach well suited to electronic files.

4.6 E-Book Metadata Records Should Specify That the Item Is an E-Book and Should Not Use Fields That Are Specific to Journals

It is critically important for citation linking and other linked data applications that rely on knowledge bases that e-books are never treated as journals or journal articles. The journal article is the default form expected by most linking systems, citation tools, and other tools used by researchers. E-book metadata should not contain fields such as "Article Title" that may lead the e-book to be misinterpreted as a journal article by these systems. An exception to this is the use of International Standard Serial Numbers (ISSNs) in connection with book series, discussed below in 5.1.5.

4.7 Linked Data Should Be Preferred to Local Silos of Data

All stakeholders in e-book metadata should promote linked data and automated data exchange so that our individual stores of metadata can be continually improved by comparison to one another.

The old concept of a library catalog as an independent, free-standing authority is now being replaced by greater collaboration and sharing of records among stakeholders. Networking is more important than autonomy today.

4.8 The Development of Systems Identifying the Ontological Relationships Among Books Should Be Promoted

Various efforts are underway to create work identifiers. For many of the top needs among the stakeholders, identifying related books as being the same work or expression is crucial and currently very difficult. For marketing, purchasing, collection management (local and collaborative), and preservation, the stakeholders want to know whether one book has essentially the same content as another or is a different edition. Advances in computer-based ontology management hold the promise of identifying not just the fact that two books are related but the type of relationship that exists between them.

4.9 The Accuracy of Fields Needed by Stakeholders for Record Matching Should Be Ensured

All the stakeholders that work with large numbers of bibliographic records have an absolute requirement to be able to match records for the same book that come from different files and different sources. Typically, record-matching algorithms require all of the metadata elements covered by this recommended practice except subject classifications:

- Book identifier, if available
- Creator
- Title
- Publication date

Table 1 displays the most critical fields for record matching, but others, such as Publisher, may play an important role in record matching.

Table 1: Needs for Record Matching

Stakeholders	Needs for Record Matching
OCLC, Library of Congress, and other authorities for book identifiers	Need to match metadata records received from publishers against existing records to avoid creating duplicate identifiers for the same book
Bibliographic record providers	Need to match metadata records received from publishers against existing records to avoid duplicating effort and creating duplicate records
Jobbers and distributors	Need to match metadata records for books included in their automated purchase services against library holdings to avoid duplicating purchases
Libraries	Need to use the same basic data elements to identify duplicate records and compare their holdings to partner libraries, consortia, and union catalogs
Preservation agencies	Need to use these metadata elements to deduplicate content as it arrives
Suppliers of delivery platforms	Need to use these metadata elements to help libraries identify when they might be purchasing or licensing duplicate content

4.10 The Publisher Should Be the Top-Ranked Source for Providing and Updating Core Metadata Elements

Many metadata consumers have a ranked list of the various sources they use for metadata. Ideally, the publisher should provide the authoritative version of the core metadata elements and attributes and should be vigilant about updating and correcting its metadata. The publisher and imprint should be listed in the metadata in a consistent way. It is helpful when one publisher is acquired by another for the original name to persist as an imprint. Unless a publisher is proven unreliable, consumers of the metadata should give priority to metadata deliveries and updates from publishers. Enrichment of the metadata, beyond the elements and attributes discussed in the Recommended Practice, is often needed to meet local or other specialized requirements. Metadata enrichment should not modify the core elements.

When a strong relationship exists between a publisher and metadata recipient, it is very valuable for the recipients to report errors back to the publisher. Publishers in turn should update their metadata, make it easily accessible, and apply version control to metadata files to facilitate easy and accurate updating.

4.11 Data Aggregators Should Not Promulgate Metadata That Are at Odds with the Publisher's Version

Data aggregators have needs to modify metadata for specific applications. They should be careful not to promulgate changes to publisher-provided metadata into the supply chain.

An aggregator should not insert its name into fields in a source database; for example, by inserting the aggregator name into the publisher field or adding an in-house series name. If the metadata standard in

use provides a local title field, such information can be stored there. Such local metadata should not be promulgated into the supply chain.

Section 5: Recommendations for Specific Data Elements

5.1 Titles

5.1.1 Definition

A title is a word, phrase, character, or group of characters normally appearing in a resource that names it.[11] The title is exclusive of any subtitles and labels. The title of the book is the name of the resource and can be transcribed from the source itself. It is different from and shouldn't be confused with fields such as "preferred title," "sorted title," "subtitle," "alternative title," "collection title," and others. Every resource can have only one title in its metadata.

5.1.2 Stakeholder Needs for Titles

Stakeholders	Needs for Titles
Publishers	• Need to plan their annual title lists well in advance in order to forecast revenue and expenses and communicate prices to potential buyers, retailers, and distributors, and may need to use preliminary or placeholder title to do this • Need to announce annual title lists to potential buyers, retailers, and distributors well in advance of publication to generate interest and sales and may need to use a preliminary or placeholder title to do this • Need to sell books and complete annual title lists before some or all of the books are actually published and may need to use a preliminary or placeholder title to do this • Want to update book titles and replace placeholder titles once the final title is established, and they want the final title to replace the preliminary or placeholder title throughout the metadata supply chain and in all marketing and communication around the book and the annual title list • Want to be the authority for the titles used for their books • Want titles to be discoverable in web searches and on retailer and other platforms so that potential buyers and readers can find them
Retailers	• Need to know the size and prices of annual title lists well in advance of publication • Need to know a final title well in advance of publication so that they can announce and market books to their customers and so that customers can identify the book and place advance orders • Want to avoid revisions to the titles of a book once they have communicated it to customers
Library service providers	• Need accurate and consistent book titles in order to use book titles to match metadata records from different files and sources

	• Need titles in order to maintain an accurate knowledge base that includes all e-books in the holdings of all the libraries they support[12] • Want to receive changes to title information as soon as possible, and want book titles to be finalized well in advance of publication in order to maintain an accurate knowledge base and matching system for customers • Need accurate titles to include in OpenURL link syntax so that end users can link to their library's full-text entitlements[13] • Need accurate titles in order to support searching within a discovery service
Jobbers and distributors	• Need to know forthcoming e-book titles well in advance of publication • Need accurate titles of all e-books currently available for order from a publisher
Metadata providers	• Need accurate titles of all e-books well in advance of publication
Preservation agencies	• Need accurate titles delivered with content in order to assist in deduplication of content receipts • Need titles delivered with content to match those titles provided on forward lists of what books will be provided for preservation • Need titles to be descriptive without reliance on subtitles
Libraries	• Need accurate and consistent titles to help match metadata records from different files and sources for the same expression or manifestation of a book for acquisitions, cataloging, collection management, resource linking, usage analysis, resource sharing, and collaboration with other libraries, union catalogs, and preservation agencies • Need accurate titles for search, display, and deduplication on their discovery system or online public access catalog (OPAC) • Want the final version of the title included in the metadata record upon publication, or earlier if possible
Readers	• Need the title of an e-book to be consistent everywhere they encounter it so they can recognize it and search or ask for a known title • Want to see announcements, advertising, and reviews of titles related to their interests • Want to buy, borrow from a library, read online, or download e-books based on interesting titles • Want to see the title within the e-book itself when they are reading it

5.1.3 General Guidelines for Titles

Preliminary titles may be used in front lists. Libraries often purchase annual book collections from major publishers in advance of publication. This purchasing model enables libraries to provide their users with the most up-to-date books and usually to receive a discount off the post-publication price. However, libraries need to be aware that these advantages also come with uncertainty about whether the metadata will change and indeed whether a specific title will be published at all.

Publishers should follow the Book Industry Study Group (BISG) Best Practices by supplying at least a preliminary title 180 days prior to the on-sale date and updating to the final title no later than 120 days prior to the on-sale date.[14] The main or preferred title should not be left blank. When even a working title does not exist, a phrase such as "Forthcoming Title" can be placed in the title field. In MARC, preliminary titles can be placed in square brackets by convention to distinguish them from final titles, while ONIX provides specialized fields and practices to manage the transition from forthcoming to published books.

Publishers, aggregators, retailers, and service providers should never transmit records containing preliminary metadata after the publication date of the e-book. Libraries should treat preliminary metadata as provisional, and be prepared to update them shortly before the publication date. Final metadata should be provided before online publication or distribution of e-book files. A complete overwrite of the preliminary metadata should be done throughout all channels of distribution and communication. Post-publication changes to titles are not recommended.

5.1.4 Recommendations for Titles

5.1.4.1 Extraneous Information

Do not insert extraneous information into title fields, such as "Electronic Resource," "Kindle Version," "Third Edition," "Award Winner," or "New York Times Bestseller."

5.1.4.2 Formatting Markup

Metadata suppliers and maintainers should never attempt to finesse the display of a title by inserting non-standard markup. This can lead to corruption of the title field text. Formatting markup should follow the applicable guidelines for the metadata record format. An example of this would be guidelines provided by EDItEUR for using HTML and XHTML markup in ONIX records for descriptive copy (but not for titles).[15]

5.1.4.3 Capitalization of Titles

In some cases, the first letter of a monographic title is capitalized, but this is not a consistent rule, so the recommendation is to preserve the capitalization as first created in the metadata.

5.1.4.4 Independence of Titles and Subtitles

Metadata fields for titles should include only the titles and not the subtitles. For subtitles it is recommended to create a different metadata element. If this is not possible then it is recommended to separate the subtitle from the title by a colon. Publishers should follow an appropriate metadata standard for the definitions and proper handling of different types of titles. Titles should be descriptive and, ideally, they should not rely on subtitles for clarity.

5.1.5 Recommendations for Monographs in Book Series

5.1.5.1 Definition and Discussion of Book Series

There is a great variety of practice around titles appearing in book series. At one end of the spectrum, books may be produced as true serial publications, with an ISSN for the series, regular publication, and a volume number or date assigned to each book. Some of these series do not assign titles to individual

volumes or, worse, they assign volume titles inconsistently. By contrast, volumes in a non-serial monographic series are published and cataloged more like stand-alone books. Monographic series are usually purchased on a standing order model rather than the subscription model used for serials.[16] The trend in cataloging has been to deemphasize the series-level information in the metadata for monographic series. In 2006, the Library of Congress announced its decision to treat all new monographic series as monographs. The result is that the series title is treated only as a descriptive element of a multi-part monograph.[17]

5.1.5.2 Recommendations

5.1.5.2.1 Series Title

If the publisher provides a book title and a series title, the series title should be preserved in the appropriate field, but it should never be used as the work title or book title. The series title should be consistently used and treated in each volume of the series.

5.1.5.2.2 Series Without Volume Titles

Series that do not have their own volume title pose a challenge for record formats that treat them as books and therefore require a main title. It is recommended that if a main title must be supplied for untitled volumes, it should consist of the series title plus the volume number; e.g., "Handbook of Best Practices, Volume 15." It is worth considering whether e-book series without volume titles would be better handled as journals. In a purely online environment, the distinction between an e-book volume and a journal issue is tenuous. Ideally, publishers would phase out e-book series that do not have volume titles by transforming them into e-journals, and they would transform series with erratic publishing schedules into monographic e-book series. Consistent practices could then be developed for handling e-book serials that respect both the nature of the serial and the individual volumes.

5.1.5.2.3 Series with Volume Titles

When the e-book serial uses volume titles, it is recommended to treat the volumes as monographs. There are several types of e-book serials where the volume title is very important; examples include the "Handbook" and "Methods" serials that are common in science, technology, engineering, and math (STEM) publishing. In such cases, the volume title is as important as the title of a monograph for conveying the theme and scope of the book to the prospective reader. However, if the e-book serial is handled in the same way as a journal, then the work title or equivalent metadata field must be the series title, and then the volume title cannot be stored in that field. On some publisher or aggregator platforms, the user cannot find an e-book volume by searching in the work title field of the search interface.

5.2 Names

5.2.1 Definition

This recommended practice is concerned primarily with the names of e-book creators, but the principles can be applied to other types of individuals and groups that occur in metadata. There are two types of creators who are responsible for the creation of the content according to Library of Congress (LC) authorities: those with personal names, including authors, editors, performers, photographers, and artists; and those with group names, including corporate entities, government bodies, conferences, and jurisdictions.

5.2.2 Stakeholder Needs for Names

Stakeholders	Needs for Names
Publishers	- Need to identify the same creator in multiple internal systems, including the publishing system and accounting system - Need to track books by the same creator - Need to market books based on their creators and previous works by the same creator
Libraries	- Need to catalog or have a discovery system to display and make retrievable content by a particular creator based on a personal name, preferably a controlled form of the name - Need a name authority to identify the same creator despite differences in the name used with a particular book, such as pseudonyms, translations, transliterations, and variations - Need a name authority to distinguish contributors with similar names, a particular problem for corporate names - Need to use names in the contributor field to help identify and disambiguate books with similar titles - Want to report on publications by creators associated with the institution
Preservation agencies	Need for consistency in delivery of names in metadata records and e-books
Library service providers	- Need a name authority in order to identify the same creator across many works despite differences in the name used with a particular book - Need a name authority in order to distinguish creators with similar names - Need accurate names in order to support search/retrieval within the discovery tool and knowledge base
Retailers	- Want to promote titles based on their authors, especially for popular authors - Want to be able to distinguish authors with similar names
Readers	- Want the e-book they are reading to show the name of the creator - Want to search for e-books by author or find additional works by a specific author - Want to use creator to distinguish books with similar titles - Want to distinguish authors with similar names

5.2.3 General Guidelines for Names

Personal and group names may occur in numerous places in e-book metadata, most commonly in book and chapter contributor fields. Accuracy in names is important to all stakeholders, as these names are often used to disambiguate similar titles, to identify duplicate books, and to identify a book or corpus of books written by a single author. Most importantly, most library services index personal and group names, and consumers and library patrons often want to locate books written or edited by a specific author.

5.2.3.1 Creator

At least one creator, such as an author or an editor, is mandatory for every e-book.

5.2.3.2 Chapter Contributors

Books with chapter level-contributors should provide a creator name for every chapter, in addition to the name(s) of the e-book editor(s).

5.2.3.3 Name Authorities

For all personal and institutional name fields, we recommend that names be used from official authority files such as the Library of Congress Name Authority File.[18] If the names needed are not available or are not included in an official authority file, linked data sources are preferred, such as Virtual International Authority File (VIAF), International Standard Name Identifier (ISNI), ORCID, or Wikidata.[19,20] Wherever possible, an identifier for a person or organization should be included (e.g., ISNI or ORCID). If the names are not available in an official authority file or a linked data source, other well-known lists may be used, such as IMDb or Wikipedia.[21]

5.2.3.4 Name Consistency

Publishers are strongly encouraged to use the same rules and name formats for all books and metadata in the publisher's catalog, to facilitate searching, sorting, grouping, and other rule-based operations employed by recipients of the corpus. In instances where no external authority is used, it is recommended that the same text string(s) for any given personal or institutional name should be used across the publisher's entire corpus (e.g., if an author publishes two books with a single publisher, both books would ideally use the same text string for the author's name).

5.2.4 Recommendations for Names

5.2.4.1 Follow Cataloging Practice

Metadata creators should conform to current RDA cataloging practices related to names, in addition to the rules of the metadata standard they are using.[22]

5.2.4.2 Designate Name Authorities

When using an authority file, use the exact same text (including capitalization) as given in the authority file and leverage the appropriate indicators in the metadata record to specify which authority file is being used. Some metadata record formats (e.g., MARC) use special authority fields for names which originate from standardized sources and names that are not from a standardized source. If the personal or institutional name is from a standardized source, put it in an authority field in the metadata record if the schema provides one. If the name is not from a standardized source, put it in a non-authority field. It is critical to avoid confusing authority terms with names not established by an authority.

5.2.4.3 Avoid "All Caps" in Names

Do not use all capital letters in a personal or institutional name unless the official form of the name includes capital letters.

5.2.4.4   Require First and Last Names

The minimum metadata expected for a personal name are given name and surname. When additional information is available, it should be included. Single-word names (e.g., Madonna), names including titles (e.g., Dalai Lama, Prince Philip), or other cases that don't fit neatly into given name categories should use what is available. It is highly recommended that content providers use the same text string to represent all instances of a single individual or organization in all metadata for all books they provide.

5.2.4.5   Make First and Last Names Separate Elements

It is preferred to place the elements of personal names into separate tags (e.g., <first name>, meaning given name, and <last name>, meaning surname or family name). When not possible, it is preferred that first names and last names be ordered and delimited across the entire corpus in a consistent manner. If neither is possible, we recommend personal names be in the format of last name, first name within a single tag/field, and multiple names be separated by a semicolon. It is also recommended that names reformatted for sorting be stored in separate fields or separated from the regular name by a semicolon when the same field must be used.

5.2.4.6   Use Identifiers

Providing unique identifiers for individuals, institutions, and organizations is becoming a common practice to disambiguate names and to prepare data for semantic web applications. We highly recommend adding identifiers to names when available and possible, but under a different tag/field. Name identifiers for individuals include ORCID and ISNI. For institutions, Research Organization Registry (ROR) IDs are stored with additional metadata about the organization, such as alternate names/abbreviations, external URLs (e.g., an organization's official website), and the Open Funder Registry, a common taxonomy of over 13,000 international funding organization names together with unique IDs for each.[23]

5.3   Dates

5.3.1   Definition

The Library of Congress identifies the following types of dates related to resource creation:

- Date created: the date of creation of the resource

- Date issued: the date that the resource was published, released, or issued

- Date captured: the date on which the resource was digitized or harvested

- Copyright date: a date on which a resource was copyrighted

This limited set of dates is sufficient for the needs of library systems and most of the needs of service providers. However, the workflows of publishers and retailers are more complex. ONIX 3.0 provides for a much more complex set of dates and statuses related to resource creation and availability. Here is just a selection of the dates given in Codelist 163:[24]

- Publication date

- Sales embargo date

- Public announcement date

- Trade announcement date

- Date of first publication

- Reprint date

- Publication date of print counterpart

- Date of first publication in original language

In addition to providing a rich vocabulary for the various dates associated with the publication of a book, ONIX also provides various status and role fields that can be used to clarify the meaning of these dates and other attributes, including Publication Status, Market Publication Status, and Supply Date Role. For example, the publication date of a forthcoming title is understood to be an expected publication date.

5.3.2 Stakeholder Needs for Dates

Stakeholders	Needs for Dates
Publishers	- Need to assert copyright on intellectual property - Need to manage publication schedule for production work, marketing, and coordination with retailers, aggregators, and customers - Need to base budget on forecasted publication dates and financial reporting, such as revenue recognition, on actual publication dates - Need to group books into annual title lists and may want to assign publication year based on the year of the title list rather than the year in which the book is actually made available, especially around the beginning and end of the year
Retailers, jobbers, and distributors	- Need to plan sales and marketing activity around the availability of new e-books - Need to base budget on forecasted publication dates and prices - Need to honor publisher restrictions on when a book is made available for sale - Want to use date to distinguish different versions and editions of an e-book
Aggregators and metadata providers	- Need to manage workflows based on various dates communicated by the publisher - Need to trigger events in workflows based on metadata transmissions and various dates included in the metadata - Want to use date to distinguish different versions and editions of an e-book - Want to match e-book to print records based on date
Library service providers	- Need dates in order to support searching/retrieval and ranking results within the discovery tool; end users subsequently use dates to filter, sort, and evaluate the resource within the discovery tool's interface - Need to identify e-book records to be placed in approval plans, standing orders, and demand-driven acquisition pools based on date
Preservation agencies	- Need dates so that end users can correctly identify the book or version of a book they would like to read - Need at least a machine-readable date, and preferably a string date as well
Libraries	Need bibliographic information in a catalog and discovery system that will search, filter, and sort content by publication date

	• Need to distinguish different versions and editions of an e-book based on date
	• Need to match e-book records to existing holdings based on date
Readers	• Need to know how current the material is that is presented in an e-book; for a digitized version of an older print book, want the publication date to refer to the print version
	• Want to search, filter, and sort e-books by publication date

5.3.3 General Guidelines for Dates

The preceding table shows a wide variety in the types and purposes of dates that are tracked by different stakeholders. Serious problems arise when the same date field is used by different stakeholders who understand it differently. Our recommendation is that all stakeholders should preserve and keep distinct these four main date attributes:

1) Publication date—required

2) Copyright date—required if applicable

3) Format date—highly recommended

4) Update date—required if the e-book can be updated

5.3.4 Recommendations for Dates

5.3.4.1 Publication Date

Publication date must be included. It is the date that this expression of the work is first made available to the public, regardless of medium or format. Publication date is generally the only date that can be used in search applications ranging from library catalogs and discovery systems to online bookstores. Due to the fact that e-books follow a long tradition of books being published in print, in most cases the publication date should refer to the first time the resource was made available in print. For example, if a publisher digitized a print book from its backfile that was published in 1975, and made the digital version available in 2015, the publication date should be given as 1975. However, many e-books are now born digital and many publishers may have a digital expression of books as the first format available to the public. In either case, the common sense understanding is that publication date reflects the currency of the content. This expectation should be honored; thus, publication date is applied to only content publication, revision, and updating. Publication date is not dependent on the date that the content is available for sale, nor does it depend on the date of securing copyright, nor the date in which it is available in a specific format or via a specific provider.

BISG's *Best Practices for Product Metadata* call for publication dates to be supplied for all titles as early as 180 days prior to the availability of a resource. While there is no consensus in the U.S. book trade on a single definition of publication date that would apply to all books and related products, the *Best Practices for Product Metadata* recommend that if the publication date is defined as the date a product may be made available for sale to consumers, the Strict On-Sale Date field (Sales Embargo Date code in ONIX 3.0) should be supplied along with Publication Date to avoid confusion. (Should the publisher's definition of publication date be more of an indicator of approximate availability, this additional field need not be supplied.) This practice supports the need for a retailer to know when to move a book from pre-order or forthcoming status to actual invoicing of content. In addition, purchasing, merchandising, and marketing plans are built around a product being available at a specific time.[25] Unfortunately, the BISG *Best Practices* document usage guidelines for this data element do not differ between digital and physical products, and the practice of matching the Strict On-Sale Date field with the Publication Date field creates

a conflict for librarians. Under this practice, e-books could be given a publication date that is much later than the print publication date, leading to confusion and purchase of multiple copies of the same work.

As illustrated with the example given earlier in this section, of an e-book created by digitizing an older print book from a publisher's backfile, it is necessary to use a conceptual framework such as FRBR to distinguish publication date from dates related to specific manifestations of the expression.[26] The same expression may be released in multiple manifestations, such as paperback, trade, and hardback print versions, and electronic versions in PDF, EPUB 3, Mobi, and HTML. These different manifestations do not warrant different publication dates.

5.3.4.2 Copyright Date

Copyright date refers to the date of legal copyright. It is an attribute of the intellectual property embodied by an e-book and understood within the framework of national and international copyright law. The copyright date should be the date from which the publisher claims copyright of the work or expression, not the release date of a particular manifestation, such as an e-book. For books that are still published or offered for sale by the publisher, the authority for copyright date is the publisher. Other parties in the supply chain should not add copyright data to the e-book metadata without authoritative information from the publisher.

In cases where an e-book is created by digitizing an older print book that is out of copyright or otherwise in the public domain and the creator of the e-book cannot assert copyright, this date should be omitted from the metadata. However, if copyright is claimed, the date of the claim should be used.

Even in cases where the copyright date is identical to the publication date, both fields should still be included in the e-book metadata.

5.3.4.3 Format Date

Format date is highly recommended and is especially important for e-books. Format date conveys the date on which the title first became available as an e-book. Format date should be maintained separately from the publication date, which is tied to the content, rather than the expression of the content as an e-book. Various tags can be used for format date for online platforms, including "First Published Online."

The format date is a type of publication date and it is not equivalent to the file creation date of a PDF or EPUB. For example, if an EPUB file is created but not released for sale until six months later, it is preferable for the format date to show the latter date when the book is actually published in the EPUB format, while the file creation date shows the date the file was created. In EPUB 3, the last modified date refers to the date of the rendition, while the publication date is a `dc:date` element that refers to the publication date of the content.[27]

5.3.4.4 Update Date

Update date will become increasingly important as publishers take more advantage of the ease of updating e-books, compared to the process for traditional print books. The update date can potentially provide a version identifier when the content of an e-book has been corrected, expanded, or otherwise changes.

In the future, we expect that "living books" that undergo constant updating will become more common. They exist today, but are generally presented as databases rather than e-books, primarily for online encyclopedias and reference works. For large reference works of this type (e.g., *Britannica Academic)*, it is preferable to track changes at the article level. When small changes are made to shorter monographic e-books and the publisher does not want to treat the new version as a separate edition, it is highly recommended that an update date be maintained to differentiate the versions of the e-book. The evolution of such living books is difficult to predict. If they become ubiquitous, publishers, their customers, and other stakeholders will have to reach consensus on the best bibliographic description for a constantly changing e-book. In traditional publishing, the general rule followed by publishers is that the addition or

revision of at least 10 percent of a print book is required for a new edition. A new edition has its own metadata, with a new publication date, new identifiers and changes to other elements, such as the contributors, as needed. If the concept of editions is used with books that also undergo smaller updates, a convention analogous to that of major versus minor software versions may be needed. Only the update date would be changed for minor versions of a book, while a major version change would be treated as a new edition.

5.3.4.5 Date Format

We recommend that all date fields used for e-books follow the International Organization for Standards (ISO) 8601 standard for expressing dates, which specifies the formats shown in Table 2.[28] Note that all values are numeric, four digits are used to express the year, and the format YYYYMM (without a hyphen) is not allowed.

Table 2: ISO 8601 Standard for Expressing Dates

Date Type	Format	Examples
Year	YYYY	2010
Year and month	YYYY-MM	2010-06
Specific date	YYYY-MM-DD or YYYYMMDD	2010-06-05 or 20100605

5.3.4.6 EPUB Package Version Date

EPUB packages carry a different date attribute than the ones described above. The date in the EPUB package refers to the last time the package was modified and is tied closely to a unique identifier that is assigned for the package as a whole. To identify a specific version of a packaged publication, a package identifier can be constructed by combining the unique identifier with the last modified date of the publication. Changes between versions may include minor typographic or markup corrections without affecting the unique identifier. Significant revisions to the content that result in a new edition require a change of the unique identifier.

The following example shows the Unique Identifier element for a publication:

```
<package … unique-identifier="pub-id">
    <metadata xmlns:dc="http://purl.org/dc/elements/1.1/">
        <dc:identifier id="pub-id">9781487504267</dc:identifier>
        <meta property="dcterms:modified">2015-12-15T09:18:00Z</meta>
        …
    </metadata>
</package>
```

Every metadata section must include at least one identifier element containing an unambiguous identifier for the publication. Multiple identifier elements are permitted, but only one can be marked as the unique identifier via the package element Unique Identifier attribute. This specification makes a distinction between the unique identifier for an EPUB publication and the identifier that uniquely identifies a specific version of it (i.e., to be able to differentiate EPUB publications containing different versions of the same manifestation). Two copies of an EPUB that are bit-for-bit identical are the same version and must retain the same last-modified date. If they are not bit-for-bit identical, they represent different versions, and must have different last-modified dates.

5.4 Book Identifiers

5.4.1 Definition

A book identifier is an unambiguous reference to the resource within a given context that can be communicated among stakeholders. (Identifiers used by a single organization or within a single system are not considered here.) Identifiers consist of unique numeric values or alphanumeric strings and are stored in specific fields in most metadata standards. A book identifier should be a formal, standard identifier issued and maintained by a registration agency, and may be formatted as a Uniform Resource Identifier (URI), Internationalized Resource Identifier (IRI), or a unique number or code. BISG's *Best Practices for Identifying Digital Products* lists content, format, and usage constraints as the three factors that determine the need to assign a unique identifier.[29] Associated identifiers, such as chapter-level DOIs, may provide additional access points to the content or link to additional metadata details created by external organizations.[30]

5.4.2 Stakeholder Needs for Identifiers

All stakeholders need a unique, unambiguous identifier for each e-book in all contexts and for all applications, but this simply does not exist. The complex nature of books, as articulated in the discussion of the conceptual model, the divergent needs of stakeholders, and the weight of historical practice make it difficult, if not impossible, to establish such a perfect identifier. A more realistic goal is to consistently follow best practices for identifiers and other e-book metadata elements that allow stakeholders to have a shared understanding of what an identifier identifies and the ability to trace the network of relationships among related books with different identifiers.

Stakeholders	Needs for Book Identifiers
Publishers	- Need to manage e-book production schedule and work of disparate staff and departments, using a product code - Need to associate expenses, revenues, and sales performance with product identifiers that correspond to business practices - Need to recognize different formats of the same book, such as hardbacks, trade paperbacks, EPUBs, Kindle versions, and online e-books
Retailers, jobbers, and distributors	- Need an e-book identifier that is unique and unambiguous - Need to track e-books as products available for sale - Need to associate expenses and revenues with a specific e-book product - Need to recognize different formats of the same book
Aggregators	- Need an e-book identifier that is unique and unambiguous - Need to recognize different formats of the same book in order to base new metadata records on existing records; e.g., basing an e-book record on an existing record for a print version
Metadata providers	Need to recognize different formats of the same book in order to base new metadata records on existing records; e.g., basing an e-book record on an existing record for a print version.

	• Need to assign their own books identifiers to the records they create using their own matching algorithms based on author, title, publisher, date, and other metadata elements rather than following the publishers' decisions about ISBN or product code assignment; e.g., there is no one-to-one mapping among OCLC number, Library of Congress Number, and ISBN
Library service providers	• Need identifiers in order to support known-item searching/retrieval within the discovery tool • Need ISBNs/eISBNs for all e-books in order to maintain an accurate knowledge base that covers all the holdings of all the libraries they support • Want a DOI in order to support linking out to the full text on the publisher's site • Want a unique identifier that can be passed in the OpenURL standard in order to support linking out to the book or chapter full text on the publisher's site • Need to recognize the same e-book on multiple platforms and associate electronic formats with print versions in order to remove duplicate records
Preservation agencies	• Need to use identifiers to identify duplicate content from the same provider • Need to use identifiers to identify duplicate content from different providers • Need to use identifiers to uniquely identify different versions of the same book
Libraries	• Need to recognize different formats of the same book in their catalog or holdings records • Need a reliable identifier to communicate library holdings to jobbers and retailers; preservation agencies; and other libraries, such as those in a consortia • Need a reliable identifier to guard against duplication of resources
Readers	• Need to know about restrictions or special requirements for an e-book; e.g., it can only be read in a Kindle reader or on Overdrive • Want to link from a discovery service directly to the e-book • Want a simple way to cite and retrieve a known item

5.4.3   General Guidelines for Book Identifiers

5.4.3.1   Preferred Identifiers

The most important book identifiers, which should be recognized by all stakeholders, are International Standard Book Number (ISBN), OCLC Control Number (OCN), Library of Congress Control Number (LCCN), and, for online books, Digital Object Identifier (DOI). We recommend that metadata include all of these identifiers if they are available.

5.4.3.2   ISBN

Book identifiers should be assigned to e-books by publishers and registration agencies. Publishers should register ISBNs following the ISO 2108:2017 standard and BISG guidelines for e-books and online content.[31,32] The ISBN of the e-book should always be included in its metadata record. When other identifiers, such as LCCN and OCN, are available, they should also be included. Publishers may include their internal identifiers for e-books in metadata records. Because identifiers are frequently used for

automated resolutions, an identifier should never have extra text or characters appended to it or otherwise deviate from its specified format. Identifiers should consistently appear in the same metadata field.

5.4.3.3 Use Separate Fields for Identifiers

Do not concatenate multiple identifiers in the same field; instead, provide each identifier in its own field.

5.4.3.4 Note Reason for a Unique Identifier

Include in the metadata record any characteristics of the e-book that were used as factors in determining that a unique identifier was required for the e-book. For example, if the publisher registered a unique ISBN for the EPUB 3 version of an e-book, there should be a format field in the metadata record specifying EPUB 3. If the identifier is for a specific platform, that platform must be listed in the metadata record.

5.4.3.5 Book Identifiers and Platforms

When the same e-book is made available in the same format with the same use restrictions on multiple retail or aggregator platforms, it is not necessary to specify the platforms in the metadata record.

The broadest distinction among e-book format classifications is that between online and downloaded e-books. Online e-books are generally hosted on a publisher or aggregator platform where they can be read. Downloaded e-books are retrieved, usually from a retailer platform, and then read with an e-reader application or device. The same e-book may be available in multiple formats on a single platform. For example, a publisher's platform may give the reader the option of reading the e-book online in HTML format, downloading it as a PDF, or downloading it as an EPUB 3 file, and each of these formats should have a different ISBN. In such a case, the publisher should refer to the ISO 2108 and BISG guidelines and provide one record for the e-book that uses the ISBN of the primary format, probably the online version in this example. The ISBNs of the other formats should be included in the record as identifiers for related books. If that EPUB 3 version is also sold on a retail platform, the metadata record provided to the retailer should use the EPUB 3 ISBN as its identifier.

E-books that are available on the publisher's online platform or preservation platform should include a URI/IRI, preferably a DOI, in the metadata record.

Include the identifiers of related books in the metadata record, especially the hardback ISBN, in the case of e-books that are published in conjunction with a print book counterpart. This helps retailers, libraries, and others in the supply chain make the connection between different manifestations of the same expression of the work. It is also helpful to include identifiers for other formats of the same book, if available.

5.4.4 Recommendations for Identifiers

5.4.4.1 ISBN Is Critically Important to All Stakeholders

The ISBN is the most important identifier of a book or e-book in the publishing supply chain. Refer to the National ISBN System to learn more about ISBNs and how they are created and used.[33] All e-book stakeholders are dependent on ISBNs for sales, inventory, reconciliation, title lists, and bibliographic records management. ISBN is used by many systems such URN resolution services to identify and retrieve e-book metadata. We have therefore decided to focus heavily on it and to make our recommendation as detailed as possible.

5.4.4.2 Use a Unique ISBN for Unique Content

All unique, specific content should have unique ISBNs registered and assigned by the publisher or its designee. Third parties that are not designated by the publisher should not register for ISBNs. In select cases, the international ISBN standard allows ISBN assignment by third parties. Since this may lead to

confusion within the supply chain, publishers are encouraged to understand stakeholder needs for unique ISBNs and implement the assignment at the content creation process.

5.4.4.3 Follow ISO 2108 and BISG Guidelines for Formats

Follow ISO 2108 and BISG guidelines by using a different ISBN for each format of the e-book.

5.4.4.4 Use a Unique ISBN for Each Edition

New editions of a book require a new ISBN. While e-books are often updated more frequently than print books, an update to an e-book is not generally understood to be a new edition, just an updated version. (As suggested in the Date section of these recommendations, it is helpful to track the update date for e-book updates in the e-book's metadata.)

5.4.4.5 Metadata About ISBNs

Metadata should clearly indicate which ISBN belongs to the main product, its relationship to all other versions, and all available formats of the e-publication (such as print, PDF, EPUB, etc.). When referencing format types, it is preferable to use a consistent naming convention (e.g., "EPUB" or "e-book") throughout produced records. Additionally, it must be clear which ISBNs are assigned to the same intellectual content (i.e., in different formats) versus which ISBN is assigned to related intellectual content (e.g., different edition, title in a series).

5.4.4.6 ISBNs for Print, Online, and EPUB

The ISBN format could be broadly divided into three main attribute categories: print, online, and EPUB. If the metadata schema allows, it is best to sort all print versions of a resource and include them under the broader category of print. Notably, ONIX does not support this type of grouping of ISBNs by format. For instance, hardback, paper, and paperback ISBNs could be included under the print type of ISBN, and electronic, digital, or PDF ISBNs could be included under the online type of ISBN. For EPUB, as this is a separate package and not just a format, it is best to have its own category/type of ISBN, as EPUB.

Note that we recommend adding these three broad categories in addition to, and not instead of, the format of each ISBN; i.e., paperback, PDF, etc. Try to harmonize the names used to describe the format of the ISBN and be consistent. For example, if you have decided on "Hardback," then follow this term and don't introduce "Hardcover" as an alternating or additional term. Also, normalize the writing form of the term throughout; avoid alternating between lowercase and capital letters or adding hyphens and other typographic symbols.

5.4.4.7 ISBN for Open Access Titles

A distinct ISBN should be added for open access (OA) versions of an e-book to distinguish it from related retail sale versions and facilitate matching the e-book to its OA license. If possible, the OA license statement should be included in the e-book metadata.

5.4.4.8 Use ISBN in E-Book File Names

The name used for e-book files should include the ISBN for that rendition of the e-book. Naming e-book files using an ISBN for a different rendition makes it difficult to match the metadata record to the e-book file.

5.4.4.9 Include Associated ISBNs in E-Book Metadata

Reference all related ISBNs assigned to the intellectual content of the book in all versions of the book (e.g., metadata accompanying the print version of the book should include the ISBN for the electronic version, and vice versa). The importance of inclusion of all the ISBNs associated with a manifestation of a work cannot be stressed enough. In particular, when trying to deduplicate title lists, all the ISBNs should be available in searches, in order to unambiguously identify titles.

5.4.4.10 ISBN Format

Data suppliers should provide ISBN-13 and omit hyphens. Use of ISBN-10 is not recommended unless it clearly references the ISBN as obsolete, in which case its inclusion will help locate older copies of the resource. ISBN-13 should never be abbreviated by omitting the first 3 numbers.

5.4.4.11 Use ONIX 3.0 Capability to Provide Book Format for an ISBN

Publishers should take advantage of the capability of ONIX 3.0 to show the format associated with each ISBN listed in the record.[34]

5.4.4.12 Universally Unique Identifiers (UUIDs)

Do not utilize UUID as the unique identifier for an e-book, as UUIDs are not issued and maintained by a registration agency, as recommended above in 5.4.1.

There is a common practice in EPUB creation (due to how InDesign exports EPUB files) to utilize an auto-generated UUID as the EPUB file's Unique Identifier, with the assigned ISBN included in a separate <dc:identifier> element. This practice breaks the relationship between the EPUB package and the correct, assigned identifier (the ISBN) as well as the relationship between the identifier and the modified date as explained above in 5.3.4.6, and could cause problems for some ingestion systems.

5.4.5 OCLC Control Numbers

5.4.5.1 Description

WorldCat is a union catalog that itemizes the collections of 72,000 libraries and is operated by OCLC. Subscribing member libraries collectively maintain WorldCat. OCLC automatically assigns OCLC control numbers (OCN) to items that are loaded into WorldCat. Because of the breadth and prevalent use of WorldCat by libraries and vendors around the world, supplying OCLC numbers in metadata records when they are available is vitally important. OCLC numbers are used as a match point between agencies when bibliographic and associated information is exchanged. OCLC numbers are also used for setting holdings or for removing holdings settings for a particular institution in WorldCat. Librarians also use OCLC numbers to match local metadata records to WorldCat records. Unfortunately, OCLC control numbers are not reliable as a unique identifier. There may be multiple OCLC control numbers for a specific manifestation of a work, making it unreliable for record matching.

5.4.5.2 Usage of OCLC Control Numbers

OCLC numbers in the metadata should not include "OCN" or any other prefix.

5.4.5.3 When OCLC Control Numbers Are Required

OCLC control numbers are required when libraries notify OCLC that they hold specific titles in their collections; when vendors notify OCLC that the title is available from them; when OCLC sends updates to libraries or vendors; and when vendors supply metadata to libraries who have subscribed to the OCLC metadata service.

5.4.6 Digital Object Identifier (DOI)

5.4.6.1 Uses of DOIs

DOI should be applied at the title level. DOIs are used to uniquely identify electronic content. Publishers can work with one of several agencies worldwide to register content and receive DOIs, including Crossref, Airiti, Inc. (Taiwan), CNKI (China), and KISTI (Korea), to name just a few.[35]

A DOI consists of two elements separated by a slash: a prefix, issued by the registering agency, and a suffix, chosen by the publisher for each item.

DOIs are globally accepted but not globally created for books and book chapters. In general, agencies and publishers create DOIs for scholarly and professional content as opposed to trade books. Agencies may charge membership or per-item fees, which may prove prohibitive for some publishers. Likewise, agencies may have content delivery requirements that may prove difficult for some publishers to apply.

DOIs for books and book chapters may not have a clear relationship beyond the prefix, which will be the same for anything published by the same organization. For example:

> Book: https://doi.org/10.1016/c2014-0-02431-8 (Parent)
>
> Chapter 1: https://doi.org/10.1016/b978-0-08-100151-6.00001-9 (Child)
>
> Chapter 6: https://doi.org/10.1016/b978-0-08-100151-6.00006-8 (Child)

Metadata consumers, such as aggregators and link resolver services, may also use DOIs to enhance metadata or to create persistent links to the content. DOIs include a prefix that identifies the publisher, followed by a suffix that is unique for each item. It is highly recommended to create DOIs for all scholarly electronic content, such as books, book chapters, and reference works, in addition to other already existing identifiers such as ISBNs.[36]

5.4.6.2 DOI Formatting

All punctuation and special characters in a DOI must be retained in transmission.

Although the DOI system allows spaces, most Registration Agencies do not allow them, as they cause significant linking problems.[37] Therefore, data suppliers should ensure that DOIs in their metadata contain no spaces, and that DOIs listed in published works (e.g., in a PDF or EPUB file) contain no spaces.

Data suppliers should isolate the DOI in a field and omit URL elements such as "https://doi.org/" in the metadata, except in fields that provide hyperlinking.

5.4.7 Library of Congress Control Number (LCCN)

LCCN identifiers are also significant: The Library of Congress assigns an LCCN to all items held by the Library. The Library of Congress provides guidance on LCCN format and usage in its MARC standard.[38]

5.4.7.1 Usage of Library of Congress Preassigned Control Numbers

Data suppliers should not include preassigned control numbers (PCNs) in metadata.[39]

5.5 Subjects

5.5.1 Definition

A subject term is a word or phrase that reflects the subject matter of the content, entered as free-text keywords. Subject terms describe what the resource is about, including topics, personal/corporate names, work titles, geographic information, or time periods/culture. This element may be used for free-text keywords, tags, or local project-level category terms that may not draw from a formal controlled vocabulary. When using subject terms from a controlled vocabulary, applying structured term entries provides greater consistency in search results for users. Subjects also group related resources and enable exploratory searching/browsing.

5.5.2 Stakeholder Needs for Subjects

Stakeholders	Needs for Subjects
Publishers	• Need to track and predict market interest in specific subject areas

	• Need to sell subject-based title lists, and may need to assign a book to a specific list for sales and marketing reasons rather than simply on subject coverage. For example, a publisher may assign a book that covers chemistry, material sciences, and chemical engineering more or less equally to a chemistry title list to meet the title count target on that list. Such title list assignments should not be construed as definitive subject classifications.
Jobbers and distributors	• Need to help libraries find similar books from multiple publishers based on consistently applied subject classifications
Library service providers	• Want to use subject classifications to improve searching/retrieval and results ranking and filtering • Want to sell subject-based title lists
Preservation agencies	• Subjects are typically not used by preservation agencies, other than to preserve them for posterity
Libraries	• Need to track and predict patron demand for e-books by subject area • Need to manage acquisition work and budgets by subject area • Want to use subject classifications to improve search and results filtering in their catalog or discovery system
Readers	• Want to locate and access e-books on topics of interest
Researchers	• Need to perform comprehensive searches for e-books related to their research topics

5.5.3   General Guidelines for Subjects

5.5.3.1   Consistency with Other Formats

Any special considerations needed for e-books are essentially the same as those encountered when classifying intellectual property across formats, such as hardback, paperback, and audio—consistency is paramount. In most cases, if a book was originally published in print format, the same subjects would be assigned to the e-book. However, attention should be given to whether an e-book contains significantly different content than an earlier version of the same title, if newer and more appropriate subjects are available in the authority in use, or if previously assigned subjects have been deprecated.

5.5.3.2   Subjects vs. Keywords

Some metadata providers refer to keywords and subjects as the same thing and use one bibliographic field for both. This is not recommended. Keywords and subjects are not the same thing: Keywords are a more specific form of topic that could be put into search, whereas subjects are a higher level of keywords. In this document we specify the most common controlled vocabularies for subjects or subject schemas, suggesting stakeholders to choose at least one of them in addition to any internal subject or keyword schema they might be using, in order to find common ground with other links in the e-book supply chain

and in order to ease the process in cases when there is a need to harmonize metadata that is coming from separate and different databases.

Always give priority to subjects from official authority files such as Library of Congress Subject Headings (LCSHs), Book Industry Standards and Communication (BISAC) Subject Codes, or Thema.[40] LC subjects are library-oriented and may take some training, but BISAC and Thema were specifically developed as book industry schemas rather than cataloging tools and thus may be easier to learn. Be aware of whether the authority you are using is conveyed as codes or text. If both are called for, make sure the data is properly fielded. LC subjects are generally transmitted as text. BISAC and Thema have unique codes associated with text and usually both are transmitted. Dewey Decimal Classification and LC Classification are typically transmitted as codes only.[41]

There are many other standards as well, some of which are specialized in nature, such as Medical Subject Headings (MeSH). In fact, ONIX now recognizes over 100 different subject schemas.[42] Keep in mind the target audience for the data. A medical library would be more likely to use MeSH subjects than a local bookstore. Similarly, the subjects from national libraries of countries not involved in the buying or selling of a given book are probably not needed.

5.5.3.3 Subject Authorities

When using an authority file, assign the same text as given in the authority file, and indicate which authority file is being used in the appropriate field. Do not use all capital letters in the subject, unless required by the standardized source. Make sure you indicate the version or edition of the authority you are using, if appropriate. This is especially important for authorities that issue numbered updates (e.g., Dewey, BISAC, and Thema) rather than authorities such as LCSH, which are continually updated. Data recipients need to know what version is being used to set up any code validation processes they may have. It is also suggested to dedicate a separate tag field for each subject and never to cluster them together, even if separated by semicolon.

5.5.4 Recommendations for Subjects

5.5.4.1 LCSH

Include LCSHs: Library of Congress Subject Headings are typically assigned by catalogers at the Library of Congress and become part of the Library of Congress catalog record for that ISBN. LCSH is a controlled vocabulary consisting of standardized headings and subdivisions that are put together as a string (e.g., Agriculture--Economic aspects--United States--History--20th century--Bibliography).

Although Library of Congress subjects are usually associated with MARC records, they can also be transmitted in other formats, such as ONIX (Code List 27 value 04).

5.5.4.2 BISAC Subjects

Include BISAC subjects: BISAC subjects are maintained by the Book Industry Study Group (BISG) and have been the book industry standard for North America for nearly 25 years. Issued annually in the fall, the list started with roughly 2,000 subjects originally designed as a shelving guide for physical stores (each section being an attempt to mirror a bookstore section). It has grown to over 4,600 subjects in over 50 sections and has been adapted as a key tool for browsing online stores. One of its strengths is that the BISAC Subject Committee consists of publishers, retailers, and book aggregators, who are in touch with what is happening in the marketplace. Every year roughly 100 new codes (and occasionally entire sections) are added as a reflection of the needs of the industry. Phrasing of subjects is also updated to reflect common usage.

5.5.4.3 Thema Subjects

Include Thema subjects: Thema subjects are maintained by EDItEUR and were designed as a scheme that could be used globally by the book industry. If you are familiar with BIC (Book Industry

Communication), you will recognize the coding and structure of Thema, since its roots are in BIC.[43] BIC runs the UK Thema National Group, which feeds directly into the International Thema Steering Committee. Thema is, in effect, a greatly expanded and "internationalized" version of BIC and was developed by a steering committee that meets regularly and has members from several countries.

Thema is hierarchically coded and consists of subjects that can be qualified by codes from six different lists. The codes are of variable length, with subject codes starting with a letter and qualifier codes starting with a number that indicates the type of qualifier. Subjects may be assigned without qualifiers, but qualifiers may not be assigned without subjects. Qualifiers are not "attached" to subjects but are fielded separately and essentially qualify all the subjects assigned.

5.5.4.4 Other Subject Schemas

It would be unwieldy to describe every possible subject schema that is available (as mentioned above, ONIX recognizes over 100 of them). These range from well-established, library-oriented systems that generally take training to assign correctly (Dewey Decimal Classification is ONIX value 01, Library of Congress Classification is ONIX value 03) to schemas closely identified with certain markets (CLIL is a French system of four-digit numbers and is ONIX value 29) to schemas that are specific to certain subject matter (Getty Art and Architecture headings are ONIX value 08).

Whatever subject scheme is chosen, be aware of the structure and usage rules for that scheme and be aware of how this corresponds to the metadata format in use. Always try to use at least one of the recognized industry-standard subjects and then add the subjects from other schemas as they relate to the intended data recipient. Accurately assigned subjects will be rewarded with better discoverability and sales.

5.5.5 Final Thoughts

A Google search for "the future of e-books" immediately retrieves a snippet stating ominously that "the future of e-books is to adapt or die." E-books have been adapting for a long time now. In their simplest form, e-books have been around since at least 1971, when Project Gutenberg was launched. And yet, it was only with the first e-book reader launching in 1998, the Rocket, that a whole new workflow for creating and distributing books in digital formats began to inform and reform the publishing industry.[44]

The perception of e-books by both creators and readers has dramatically changed since their inception. The early visions of e-books included high expectations for the new format as a transformational technology. We were told that books would come to life as e-books, incorporating multimedia features, interactivity, automated updates, and more knowledge than was ever printed even in multi-volume encyclopedias; there were experiments with hypertext fiction (Judy Malloy) and books published on CD-ROM.[45] A dozen innovative e-readers have come and gone, and the most creative experiments in format and user experience have mostly not survived, because of the limits of technology, the psychology of readers, market forces, piracy concerns, lack of standards, and perhaps other reasons.

Nevertheless, the evolution and adoption of e-books proceeded steadily. They did not make print books obsolete, but they have forced publishers, distributors, retailers, libraries, and consumers to develop new business models, sales models, channels, and purchasing, borrowing, and reading habits. It has taken two decades, but e-books are now central to all conversations about metadata standards, discoverability, accessibility and online reading trends and usage. We conclude our work on this recommended practice confident that e-books will continue to adapt, not die.

Our recommended practice is a small step in the development of e-books but it is our hope that it is a meaningful step towards building consensus on standardizing how metadata describes e-books, and a collegial environment where best practices are shared—and negotiated—by all stakeholders.

Like the first appearances of some very ambitious e-books, our group too, began the conversation on e-books with the latest developments in our field—semantic web vocabularies and data models like

BIBFRAME. We quickly realized we needed to take a step back. It became clear that at this point in time the most needed work had to be done on basic bibliographic fields and on some highly complex workflows that we knew little about.

And yet, as this document began to form into a recommended practice that we all trust, and one that we want to share with e-book creators, distributors, and users, we are confident that e-books will continue to evolve and adjust to fit into new information systems and data models. There are at least three subjects pertinent to books that we could not include in the first version of this recommended practice, but we believe should be addressed in future versions: linked data for books, accessibility, and open access.

5.5.5.1 Linked Data

As we were applying the comments received during the public review of this recommended practice, OCLC published a detailed report titled *Transforming Metadata into Linked Data to Improve Digital Collection Discoverability*, a pilot project with five different institutions.[46] The project demonstrated the amount of work involved in the transition to linked data, but also that the tools exist and the workflows can be developed. Clearly, institutional investment in transforming metadata to linked data, partnering to develop data models that can be shared across institutions, and investing in web services and user interfaces that could leverage the power of linked data are challenges that will necessitate better communication and collaboration among the different types of e-book stakeholders.

5.5.5.2 Accessibility

In 2020, the Balisage Series on Markup Technologies focused on accessibility. Although the XML-based standards JATS and BITS already have elements to describe accessibility, focusing on accessibility during the annual gathering highlighted the vast amount of work that still needs to be done in addressing accessibility features on websites, in metadata statements, and as a requirement in various formats such as HTML, EPUB, and ONIX. Ensuring accessibility for web content is complex, and the multiplicity of metadata schema used for e-books presents a significant challenge. As with the metadata elements addressed in this recommended practice, there is a great need to build consensus across the various schema on how to describe the accessibility features of e-books.

5.5.5.3 Open Access

University presses and higher education institutions are seeking publishing models that would allow research to be shared more freely. Recent developments in the area include inventing new forms of peer-reviewed scholarly works, new infrastructures for open-source collaborative publishing processes, and new funding models. Recent development in OA e-book publishing will certainly require the metadata standards to catch up. Already, work on e-book usage reporting is confronting the need for standardizing the practice for assigning ISBNs for OA e-book titles. Clearly, there will be a need for a common understanding of how to describe OA e-book titles and how to distinguish them from regular front list titles.

5.5.5.4 Representing Diverse Perspectives

As mentioned at the beginning, this document is limited to the English-speaking world and has been written from an Anglo-American point of view. Future work on this draft must be expanded to review the appropriateness of the current recommended practice to e-books in non-English languages and content from self-published, independent, small, and local producers. It would also be beneficial in future drafts to reference the work currently underway to address the marginalization and underrepresentation of certain groups in e-book metadata fields such as subject headings. As e-book workflows become more automated and more complex, stakeholders may also need to investigate bias in their systems (or what we may call algorithmic bias) and encourage broad and equal access in infrastructure and discovery.

Appendix A: Metadata Examples

A.1 Titles

MARC

The title proper is reflected in the MARC 245 field (245 $a). Other titles of importance are the sub-title (245 $b) and variant titles (246 $_1). Also, related/analytical uncontrolled titles are in the MARC 740 field. The title proper includes the numerical designation of a part/section (245 $n) and the name of a part/section (245 $p). For resources that contain series, include either: 490 0# (Series descriptive statement) or a pair of fields: 490 1# (Series descriptive statement) and an 830#0 (Series authorized access point). Note that 490 1# and 830#0 often have the same content but fulfill different functions

ONIX

ONIX for Books defines fourteen different types of book titles in Codelist 15, as well as subtitle and collection titles. ONIX 3.0 merges the handling of the series and sets used in ONIX 2.1 into a single codelist for collections. This is consistent with the trend towards deemphasizing series metadata. ONIX 3.0 includes features to improve the handling of e-books. All of the examples below use ONIX 3.0.

EPUB 3

The EPUB 3 specification for title information to be embedded in the e-book record supports six basic types of titles:

- main
- subtitle
- short
- collection
- edition
- extended

A main title is required. If the book has a subtitle, it is also required.

BITS (Book Interchange Tag Set)

The metadata standard BITS includes the collection-meta section, which supports capturing metadata for either a series or a book set, with optional tags that include:[47]

- collection-id
- title-group
- volume-in-collection
- issn
- issn-l (Linking ISSN)
- isbn

A.2 Names

General:	The Absent City Ricardo Piglia Tr. Sergio Walsman Duke University Press, 2000

BITS:	```
	<contrib-group>
	    <contrib contrib-type="author" xlink:type="simple">
	        <string-name>Ricardo Piglia</string-name>
	         <role>Author</role>
	    </contrib>
	    <contrib contrib-type="translator" xlink:type="simple">
	        <string-name>Sergio Waisman</string-name>
	        <role>Translated by</role>
	    </contrib>
	    <bio xlink:type="simple">
	        <title>Biographical note</title>
	         Ricardo Piglia lives in Argentina and is the author of nine Spanish-language
	           books, two of which have been previously translated into English:
	                <italic>Artificial Respiration,</italic> also published by Duke University
	           Press, and <italic>Assumed Name. The Absent City</italic> has been performed as
	           an opera in Argentina and Piglia's books have been translated into Portuguese,
	           French, Italian, and German. His fiction has won the Casa de las Américas Prize,
	           the Boriz Vian Prize, and the Premio Planeta.
	         Sergio Waisman is Assistant Professor of Spanish and Portuguese at San Diego
	           State University. His previous translations include Piglia's <italic>Nombre
	                falso</italic> (<italic>Assumed Name</italic>), which received the
	           Meritorious Achievement Award in the 1995 Eugene M. Kayden National Translation
	           Contest. In addition, Waisman was awarded a National Endowment of the Arts
	           Translation Fellowship to support his translation of <italic>The Absent
	                City.</italic>
	    </bio>
	</contrib-group>
	``

MARC:	One Option: 100 1 a| Piglia, Ricardo. 600 1 0 a| Waisman, Sergio Gabriel. 0| http://id.loc.gov/authorities/names/n95049240 Another Option: 100 1_ |a Piglia, Ricardo. 600 10 |a Waisman, Sergio Gabriel.
MARC XML:	``` <datafield tag="100" ind1="1" ind2=" "> <subfield code="a">Piglia, Ricardo.</subfield> </datafield> ```
ONIX:	``` <Contributor> <SequenceNumber>1</SequenceNumber> <ContributorRole>A01</ContributorRole> <PersonName>Ricardo Piglia</PersonName> <PersonNameInverted >Piglia, Ricardo</PersonNameInverted> <NamesBeforeKey>Ricardo</NamesBeforeKey> <KeyNames>Piglia</KeyNames> </Contributor> <Contributor> <SequenceNumber>2</SequenceNumber> <ContributorRole>B06</ContributorRole> <PersonName>Sergio Waisman</PersonName> <PersonNameInverted>Waisman, Sergio</PersonNameInverted> ```

```
    <NamesBeforeKey>Sergio</NamesBeforeKey>
    <KeyNames>Waisman</KeyNames>
</Contributor>
<ContributorStatement>by Ricardo Piglia, translated by Sergio
Waisman</ContributorStatement>
```

General:	Public Relations and the Corporate Persona Burton Saint John III Taylor & Francis, 2017
BITS:	One option: ``` <contrib-group> <contrib contrib-type="author"> <string-name>Burton Saint John III</string-name> <aff> <institution>Old Dominion University, USA</institution> </aff> <role>Author</role> </contrib> </contrib-group> ``` A more robust option: ``` <contrib-group> <contrib contrib-type="author"> <string-name>Burton Saint John III</string-name> <name name-style="western"> <surname>Saint John</surname> <given-names>Burton</given names> <suffix>III</suffix> </name> <aff> <institution>Old Dominion University, USA</institution> </aff> <role>Author</role> </contrib> </contrib-group> ```
MARC:	``` 100 1_ |a St. John, Burton, |d 1957- |e author. ```
MARC XML:	``` <datafield tag="100" indl="1" ind2=" "> <subfield code="a">St. John, Burton,</subfield> <subfield code="d">1957-</subfield> <subfield code="e">author.</subfield> <datafield> ```
ONIX:	``` <Contributor> <ContributorRole>A01</ContributorRole> <PersonName>Burton Saint John III</PersonName> <PersonNameInverted>Saint John III, Burton</PersonNameInverted> <NamesBeforeKey>Burton</NamesBeforeKey> <KeyNames>Saint John III</KeyNames> <ProfessionalAffiliation> <Affiliation>Old Dominion University, USA</Affiliation> </ProfessionalAffiliation> </Contributor> ```

A.3 Dates

Format date is tied with on-sale date, by which a certain title becomes available via specific sale channels, such as the publisher website, or available online on a publisher or aggregator platform.

BITS

The following example demonstrates how in detailed XML-based metadata all three date attributes (`pub-date`, `pub-date format`, `copyright year`) can be found:

```
<pub-date iso-8601-date="2019" publication-format="print">
   <year>2019</year>
</pub-date>
<pub-date iso-8601-date="2019" publication-format="online">
   <year>2019</year>
</pubdate>
<isbn publication-format="print">9781487504267</isbn>
<isbn publication-format="online">9781487519780</isbn>
<publisher>
   <publisher-name>University of Toronto Press</publisher-name>
   <publisher-loc>Toronto [Ontario]</publisher-loc>
</publisher>
<permissions>
   <copyright-statement>Limited to one user at a time</copyright-statement>
   <copyright-year>2019</copyright-year>
</permissions>
```

MARC

Publishers often provide only the copyright date, in which case the publication date is put in square brackets to indicate that it has been inferred. The inferred publication date should be the same as the copyright date. If the publisher specifies both a publication data and a copyright date, they can be different.

MARC Publication Date:	`(Inferred)  264 #1 $c [2019]`
MARC Copyright Date:	`264 #4 $c © 2019`

A.4 E-Book Identifiers

ISBN

BITS:	`<isbn publication-format="print">9780773557147</isbn>`
	`<isbn publication-format="print">0773557148</isbn>`
	`<isbn publication-format="online">9780773558113</isbn>`
	
	Note: if 10-digit ISBN is supplied as well as the 13-digit ISBN, the 13-digit ISBN should be listed first.

MARC—valid:	020 ##$a9783141592641 (valid ISBN) Qualifying information can be provided in a subfield $q following the ISBN: 020 ##$z9783141592641 $q (v.2 ; $q alk. papers)
MARC—invalid:	020 ##$z9783141592641 (CD-ROM) (on record for the online version)

MARC XML:	`<datafield tag="020" ind1=" " ind2=" ">`
	    `<subfield code="a">9783141592641</subfield>`
	`</datafield>`

ONIX:	`<ProductIdentifier>`
	      `<ProductIDType>15</ProductIDType>`
	      `<IDValue>9783141592641</IDValue>`
	`</ProductIdentifier>`
	
	or:
	
	`<productidentifier>`
	      `<b221>15</b221>`
	      `<b244>9783141592641</b244>`
	`</productidentifier>`

OCLC Number

BITS:	`<book-id book-id-type="OCLC">60820242</book-id>`
MARC:	001 60820242 (valid OCLC number) 019 ##$a62874189##$a244170452 (merged OCLC numbers)

MARC XML:	`<marc:controlfield tag="001">60820242</marc:controlfield>`
	
	`<marc:datafield tag="019" ind1=" " ind2=" ">`
	   `<marc:subfield code="a">62874189</marc:subfield>`
	  `</marc:datafield>`
	`<marc:datafield tag="019" ind1=" " ind2=" ">`
	   `<marc:subfield code="a">244170452</marc:subfield>`
	  `</marc:datafield>`

ONIX:	`<ProductIdentifier>`
	      `<ProductIDType>23</ProductIDType>`
	      `<IDValue>60820242</IDValue>`
	`</ProductIdentifier>`
	
	or:
	
	`<productidentifier>`
	      `<b221>23</b221>`
	      `<b244>60820242</b244>`
	`</productidentifier>`

DOI

BITS:	`<book-id book-id-type="doi">10.1000/100</book-id>`

MARC:	`856 40  ##$uurn:doi:10.1000/100`
	
	`856 40  ##$u https://doi.org/10.17226/25354`
	`856 40  ##$u https://doi.org/10/7551/mitpress/9957.001.0001`

MARC XML:	`<datafield tag="856" ind1=" " ind2=" ">`
	`     <subfield code="u">urn:doi:10.1000/100</subfield>`
	`</datafield>`

ONIX:	`<ProductIdentifier>`
	`        <ProductIDType>06</ProductIDType>`
	`        <IDValue>urn:doi:10.1000/100</IDValue>`
	`</ProductIdentifier>`
	
	or:
	
	`<productidentifier>`
	`        <b221>06</b221>`
	`        <b244>urn:doi:10.1000/100</b244>`
	`</productidentifier>`

LCCN

BITS:	`<book-meta>`
	`<book-id book-id-type="LCCN">11025047</book-id>`
	`<book-title-group>...</book-title-group>`
	`...</book-meta>`

MARC:	`010     ##$a###85153773#` Note: Only use on record for the right format. For instance, do not use on a record for the e-version, if the LCCN is applicable to the print version record. Do not try to add additional LCCNs to the field.

MARC XML:	`<datafield tag="010" ind1=" " ind2=" ">`
	`     <subfield code="u">85153773</subfield>`
	`</datafield>`

ONIX:	`<ProductIdentifier>`
	`        <ProductIDType>13</ProductIDType>`
	`        <IDValue>85153773</IDValue>`
	`</ProductIdentifier>`
	
	or
	
	`<productidentifier>`
	`        <b221>13</b221>`
	`        <b244>85153773</b244>`
	`</productidentifier>`

A.5 Subjects

MARC

```
650_0 |a Soccer |z Africa.
650_0 |a Soccer |x Social aspects |z Africa.
650_0 |a Soccer in popular culture.
650_0 |a Mass media and sports |z Africa
61120 |a World Cup (Soccer) |n (19th : |d 2010 : |c South Africa)
651_0 |a Lansing (Mich) |x History.
650_7 |a SOCIAL SCIENCE / Popular Culture. |2 bisacsh
650_7 |a SOCIAL SCIENCE / Developing Countries. |2 bisacsh
650_7 |a SOCIAL SCIENCE / Media Studies. |2 bisacsh
650_7 |a SPORTS & RECREATION / Soccer. |2 bisacsh
```

For the above, the second indicator of 7 means "Source specified in subfield $2."

In a MARC record, a subject can appear as:

650_0 $a Agriculture $x Economic aspects $z United States $x History $y 20th century $v Bibliography

Genre headings are shown in 655 field, for example:

655#4 Dissertations (UCLA)

Note that OCLC is converting all genre headings that had a local genre heading, e.g. 655#2 Dissertations (UCLA) $2 local to 655#4.

A subject term for a meeting name is listed in the 611 field, for example:

611 20 Conference on Biomarkers in AKI $n (10th : $d 2011 : $c Dublin, Ireland)

Note that when transmitting via ONIX or other more textual formats, "—" or "--" is used to separate subjects and subdivisions:

Agriculture -- Economic aspects -- United States -- History -- 20th century -- Bibliography

The subfield designations are lost but the text is more readable compared to MARC.

ONIX

```
<subject>
    <b067>01</b067>                          <!-- subject scheme (Dewey) -->
    <b068>23</b068>                          <!—edition -->
    <b069>338.10973</b069>                   <!—code -->
</subject>
<subject>
    <b067>03</b067>                          <!-- LCC -->
    <b069>HD1751.A7</b069>
</subject>
<subject>
    <b067>04</b067>                          <!-- LCSH -->
```

```
    <b069>sh85047239</b069>
    <b070>Farm produce</b070>                <!-- text -->
</subject>
<subject>
    <b067>04</b067>                          <!-- LCSH -->
    <b069>sh2007100879</b069>
    <b070> Agriculture--Economic aspects--Australia</b070>
</subject>
<subject>
    <b067>10</b067>                          <!-- BISAC subject -->
    <b068>2017</b068>
    <b069>BUS070010</b069>
    <b070>BUSINESS & ECONOMICS / Industries / Agribusiness<b070>
</subject>
<subject>
    <b067>11</b067>                          <!-- BISAC regional theme -->
    <b068>1.1</b068>
    <b069>6.0.1.0.0.0.0</b069>
    <b070>Australia<b070>
</subject>
<subject>
    <b067>93</b067>                          <!-- Thema subject category -->
    <b068>1.3</b068>
    <b069>KNAC</b069>
    <b070>Agriculture, agribusiness & food production industries<b070>
</subject>
<subject>
    <b067>94</b067>                          <!-- Thema geographical qualifier -->
    <b068>1.3</b068>
    <b069>1MBF</b069>
    <b070>Australia<b070>
</subject>
```

Endnotes

[1] Jobbers (also called Ordering Portals) provide numerous services around digital and print content acquisitions to libraries. They consolidate the ordering print and digital content from multiple aggregators and various distribution channels, and ordering, invoicing, LTS services, duplication control, enhanced curation support, and automation in acquisition workflows.

[2] BISG: https://bisg.org/; EDItEUR: https://www.editeur.org/8/ONIX/; IDPF: http://idpf.org/

[3] IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records,* IFLA Series on Bibliographic Control 19, Munich: K.G. Saur Verlag, 1998.

[4] See http://www.rda-rsc.org/.

[5] Library of Congress, *Overview of the BIBFRAME Model,* https://www.loc.gov/bibframe/docs/bibframe2-model.html

[6] FOLIO Project, *The Codex Metadata Model,* https://wiki.folio.org/pages/viewpage.action?pageId=1415393

[7] For more on the LRM model, please see the IFLA document: https://www.ifla.org/publications/node/11412

[8] For more on FRBR model please see IFLA document: https://www.ifla.org/files/assets/cataloguing/frbrrg/frbr-bibliography.pdf

[9] EPUB 3.2 Packages 3.2: Final Community Group Specification, 8 May 2019, https://www.w3.org/publishing/epub3/epub-packages.html

[10] See https://www.w3.org/publishing/epub3/epub-packages.html#sec-metadata-elem-identifiers-pid

[11] Unless noted otherwise, we use the Library of Congress "Master data element list" to define concepts we discuss in this document: http://www.loc.gov/standards/mdc/elements/

[12] NISO RP-9-2014, *KBART: Knowledge Bases and Related Tools Recommended Practice*, available at https://www.niso.org/publications/rp-9-2014-kbart

[13] ANSI/NISO Z39.88-2004 (R2010) *The OpenURL Framework for Context-Sensitive Services*, available at https://www.niso.org/publications/z3988-2004-r2010

[14] BISG. *Best Practices for Product Metadata: Guide for North American Data Senders and Receivers*, 2015.

[15] https://www.editeur.org/files/ONIX%203/APPNOTE%20HTML%20markup%20in%20ONIX.pdf

[16] Jones, E. (2018). The evolution of the serial work, the FRBR conceptual model, and RDA. *Cataloging & Classification Quarterly, 56*(2–3), 128–145. https://doi.org/10.1080/01639374.2017.1388325

[17] Spidal, D., & Marshall, L. D. (2013). Monographic series: The continuing debate. *The Serials Librarian, 65*(3–4), 350–373. https://doi.org/10.1080/0361526X.2013.818614. See also the Library of Congress title authority heading file for book series names: https://authorities.loc.gov/help/title-auth.htm.

[18] http://id.loc.gov/authorities/names.html

[19] https://viaf.org

[20] https://www.wikidata.org/wiki/Wikidata:Main_Page

[21] https://www.imdb.com/ and https://www.wikipedia.org/ respectively.

[22] Resource Description and Access: https://www.oclc.org/en/rda/about.html

[23] For additional information on all the above-mentioned identifiers, see their corresponding websites: ISNI: http://www.isni.org/about; ORCID: https://orcid.org/about; ROR: https://www.ror.community/about/; Open Funder Registry: https://support.crossref.org/hc/en-us/articles/214360886-The-Open-Funder-Registry.

[24] ONIX Codelist 163: Publishing Date Role: https://onix-codelists.io/codelist/163

[25] According to the BISG *Best Practices*, "This date is usually agreed upon in an affidavit signed by both publisher and bookseller. It is typically the date on which a book is put on sale in traditional bricks-and-mortar bookshops, but in cases where a book is sold online or via mail order prior to its appearance in physical stores, this is the date the consumer will receive the book" (p. 169). These guidelines also reinforce the need for honoring this date as it ensures that the title is released in coordination with publisher marketing activities and ensures that no materials provider or market gains an unfair competitive advantage from early sales of the title. For the full report see: BISG. (2015). *Best practices for product metadata: Guide for North American data senders and receivers*. https://static1.squarespace.com/static/550334cbe4b0e08b6885e88f/t/55d2277be4b0a2c68568aaec/1439836027770/BISG_Best_Practices_for_Product_Metadata_6.1.15.pdf.
Data recipients need to be able to rely on publication date as a version identifier for different ISBNs (manifestations) related to the same book (expressions).

[26] For FRBR definitions and relationship of entities, see http://vocab.org/frbr/core#Manifestation.

[27] EPUB 3.2 Packages 3.2: Final Community Group Specification, 8 May 2019, https://www.w3.org/publishing/epub3/epub-packages.html

[28] For more information on ISO 8601: https://www.iso.org/iso-8601-date-and-time-format.html

[29] BISG, *BISG Policy Statement POL-1101: Best Practices for Identifying Digital Products*, December 7, 2011, revised February 25, 2013. https://c.ymcdn.com/sites/bisg.org/resource/resmgr/Files/Publications/Free_Documents_-_Supply_Chain/BISG_Policy_1101-2.pdf

[30] Emory University, Core metadata project. http://metadata.emory.edu/guidelines/descriptive/identifier-standard.html

[31] International Organization for Standardization, *ISO 2108:2017, Information and documentation – International Standard Book Number (ISBN)* https://www.iso.org/standard/65483.html

[32] BISG, *BISG Policy Statement POL-1101: Best Practices for Identifying Digital Products*, December 7, 2011, revised February 25, 2013. https://c.ymcdn.com/sites/bisg.org/resource/resmgr/Files/Publications/Free_Documents_-_Supply_Chain/BISG_Policy_1101-2.pdf

[33] https://www.isbn-international.org/content/what-isbn

[34] See ONIX 3 specification, P.23.5 Related product form code

[35] For a full list of registering agencies, see https://www.doi.org/registration_agencies.html

[36] See Crossref's Best practice for book content for more information on DOIs.

[37] https://www.doi.org/hb.html

[38] https://www.loc.gov/marc/bibliographic/bd010.html

[39] https://www.loc.gov/publish/pcn/about/

[40] Library of Congress: http://id.loc.gov/authorities/subjects.html; BISAC:

https://bisg.org/page/BISACEdition; Thema: https://www.editeur.org/151/Thema

[41] Dewey Decimal Classification: https://en.wikipedia.org/wiki/List_of_Dewey_Decimal_classes

[42] The full list can be found at https://ns.editeur.org/onix/en/27, where MeSH subjects are indicated by the value 06.

[43] https://www.bic.org.uk/58/About-BIC/; https://www.bic.org.uk/86/THEMA/

[44] Hillesund, T. (2001). Will E-books Change the World? *First Monday, 6*(10). http://firstmonday.org/issues/issue6_10/hillesund/index.html

[45] Berens, K.I. (2014). Judy Malloy's seat at the (database) table: A feminist reception history of early hypertext literature, *Literary and Linguistic Computing*, 29(3), 340-348, https://doi.org/10.1093/llc/fqu037

[46] Bahnemann, G., Carroll, M., Clough, P., Einaudi, M., Ewing, C., Mixter, J., Roy, J., Tomren, H., Washburn, B., & Williams, E. (2021). *Transforming Metadata into Linked Data to Improve Digital Collection Discoverability: A CONTENTdm Pilot Project.* OCLC Research. https://doi.org/10.25333/fzcv-0851

[47] Book Interchange Tag Set: https://jats.nlm.nih.gov/extensions/bits/tag-library/2.0/