Background and Problem Statement:

The e-book publication space is still young, with publishing models and content models evolving day-to-day. Publishers, libraries, aggregators, preservation agencies, collection management systems and discovery services are working through the complexity of selling and buying e-books, making information about e-books available to end users, making the actual books available to end users, and preserving them.

One particularly difficult area for many members of this supply chain is receiving and working with quality e-book metadata. Book publishers and others need the metadata about e-books very early in the publication process in order to enable sales, which means that publishers are often making descriptive metadata about e-books available well before they are published. Discovery and cataloging services need metadata records as soon as the e-books are accessible to libraries and their end users. Libraries need metadata to track what e-books they have purchased, what e-books are accessible to their end users, and to populate metadata in their catalogs and discovery services. Preservation agencies and publishers need to confirm that the content received for preservation is the entirety of what was published.

Unfortunately, because the initial descriptive metadata is circulated before publication it may not be completely accurate or align with the published e-book. The proliferation of ISBN associated with the intellectual content that represents an e-book is particularly challenging. It may be that some publishers have one metadata record for a book with multiple ISBN subfields, each with an associated ISBN and a media field, but that level of detail is rarely making it through the supply chain. Currently, it is common for different members of the supply chain to have different ISBN associated with aligned, but not exactly identical, descriptive metadata for the same book and this makes it extremely difficult to match books between organizations. The International Standard Text Code or ISTC1, which could solve some of these problems, is not yet widely adopted. DOIs are not always assigned for e-books and may be assigned late in the publication process, after the initial descriptive metadata has been circulated. In addition, sometimes DOIs for books and chapters are not registered, which means they cannot be used to look up additional metadata or used to uniquely identify e-books, forcing reliance on the unreliable ISBN.

E-book records created by publishers and vendors may not accurately reflect the actual title of the final book. Often author information is coded in different fields from one record to the next, has different formatting and abbreviations from record-to-record for the same book, or the initial descriptive metadata that is circulated differs from the author (and other metadata) delivered with the actual e-book. Unlike with print books, library staff do not see the title when an e-book is “received”, so there is no mechanism for libraries to catch discrepancies between the records and the actual publication.

---

1 https://www.myidentifiers.com/istc/main
Other metadata fields, such as dates, are also erratic. Dates are a complicated area of metadata for books, as it is common for publication dates and digitization dates to become jumbled -- which makes it difficult to match an e-book to its print version, thereby leading to duplication in the collection. Editions and multi-part volumes are particularly complicated. It is common for the previous editions of e-books to become inaccessible when the current edition is published. When the new edition is published, new metadata records are created by the publishers and circulated to libraries. However, it is rare for the records associated with the previous edition to be updated to link the record for the previous edition to the record for the new edition (or simply link to the new edition). Without this relationship, the URLs used to access the older edition from the old record stop working, rather than redirecting to the newest edition. In regard to multi-part volumes, metadata is often provided to libraries as individual MARC records for each volume, when the libraries would often prefer a single MARC record for the set. Preservation agencies and others in the supply chain also struggle with multi-part volumes, because frequently, there is not enough information to string the individual books together as a volume or series and often individual book titles, without a series title, are not sufficient to identify books.

Given the difficulty in uniquely identifying books by ISBN, DOI, title and even combinations of these pieces of metadata, is difficult for libraries to manage the metadata records and ensure there is only one record per book. Libraries rely on publishers to provide descriptive metadata records for large e-book collections. These records are delivered to libraries on an ongoing basis and each new delivery can include both additions to the set, deletions from the set, and updates to records already in the set. Without unique and consistent identifiers for books and clear relationships between volumes and editions, it is impossible to maintain library catalogs.

These metadata discrepancies also cause comparisons of holdings between one organization and another to be inaccurate. Indeed, comparisons of expected holdings and received holdings within a single organization are often difficult (due, we think, to the early circulation of descriptive metadata before a book is published and the ways in which business data and actual books can diverge and then never realign).

**Statement of Work:**

Create a NISO Recommended Practice:

1. Describing the minimal metadata requirements necessary to describe e-books in order to support sales, discovery, delivery, deaccessioning, and preservation. Including, but not limited to:
   - addressing accurate title and author information,
   - editions,
   - multi-volume sets,
   - metadata updates pre- and post-publication,
   - dates,
   - identifiers, and
   - at least one controlled vocabulary subject or genre term

2. Identifying the most effective and efficient way for metadata to be moved through the entire supply chain. Including, but not limited to, the following members and states of the supply chain:
   - Publishers
   - Jobbers
   - Aggregators and other vendors
   - Discovery services
   - Libraries
   - Pre-sale and pre-publication
   - Pre-sale and post-publication
   - Post-sale and pre-publication
   - Post-sale and post-publication
   - Post-sale, post-publication, and post-deaccessioning
   - Any other combinations that come to light during the working group’s discussions
3. Addressing how metadata records and transfer of information may be used to describe:
   • Updates to metadata records for corrections
   • Updates to metadata records for new editions
   • Updates to metadata records for deletions
   • Updates to sets of metadata records to include new content in collections of e-books


5. Providing examples of the recommended practice implemented in a variety of metadata standards, such as JATS, ONIX, and MARC.

6. Exploring ONIX and MARC tools and identifying if it is necessary to issue recommendations around conformance tools

This working group will focus on descriptive metadata and identifiers for e-books (including monographic series) and will not address the encoding and transmission of rights information.

The working group, completing its work over email, video and phone conferencing, will follow a two phase process.

- Phase 1 will be an information gathering phase where the group will survey libraries, aggregators, publishers, metadata record creators, jobbers, preservation agencies and others to identify what metadata each needs and when and also what metadata each currently has and when.
- By laying our needs against known practices, we will be able to start Phase 2 to develop the full recommended practice to ensure that all parties receive and deliver the needed e-book metadata at the needed times.

The recommendations produced by the working group will clarify to all parties what robust, descriptive, e-book metadata and identifiers are required to be transmitted and recorded at each stage of the publication and delivery and deaccessioning process.

**Partners and Participation:**
Stakeholders include:
- Libraries
- E-book publishers (including large and small publishers, commercial presses, university presses, and open access presses)
- E-book aggregators (e.g., ProQuest, EBSCO, Springer)
- Book jobbers (e.g., YBP)
- E-book metadata creators (e.g., OCLC)
- E-book identifier agencies (e.g., Bowker for ISTC, Crossref for DOI, and ISBN)
- Other standards agencies (e.g., BISG and EDItEUR)
- Discovery services
- Preservation agencies

**Timeline:**
The recommendations will be published within 18 months of the formation of the working group.
- Month 1: Appointment of working group
- Month 2: Approval of charge and initial work plan (including final determination of scope)
- Months 3-7: Completion of information gathering (phase 1)
- Months 8-9: Completion of initial draft recommended practices document (phase 2)
- Months 10-16: public test and comment period
- Month 18: responses to comments and completion of final recommended practices document (target January 2018)