

Get Discovered

How to use your metadata to conquer AI search

Considering AI in our metadata landscape

1

What's new?

What should we be thinking about now, what's new and what will stay the same?

2

Specific metadata

Keywords, sales points, descriptions, other data points you might not be populating today

3

Evolving standards

How will ONIX evolve and change with AI?

4

The long tail

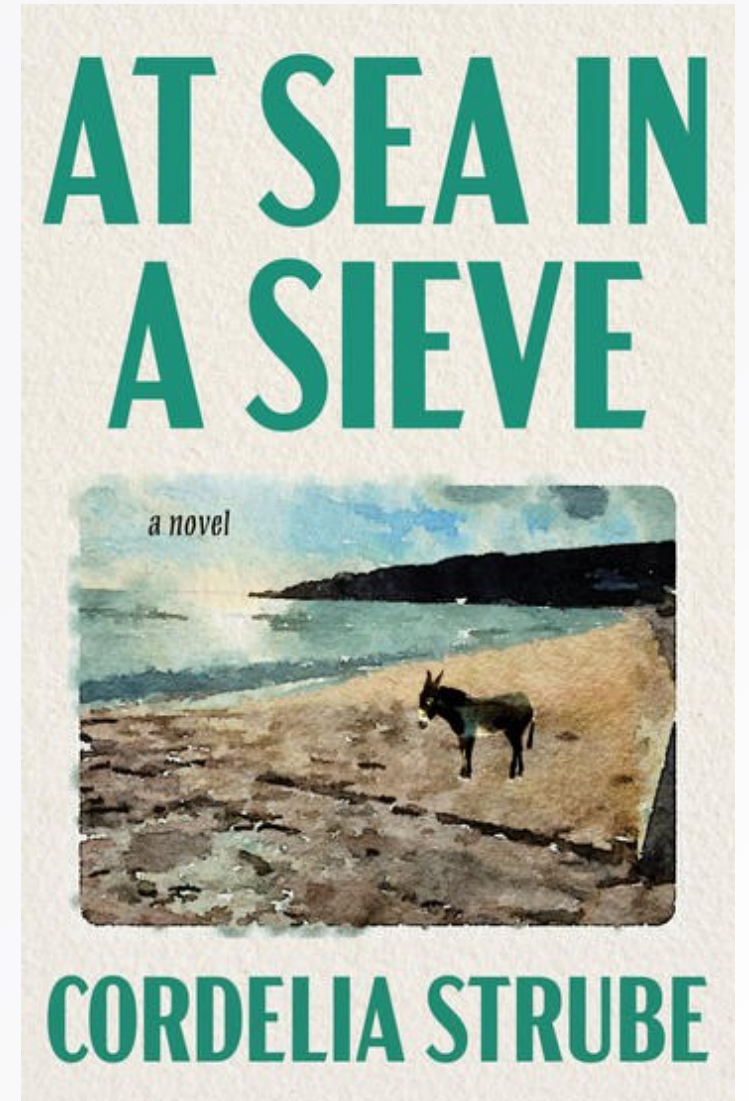
How can we get more attention on our backlist titles, and what about libraries?

What's New?

And what will stay the same?

We've been here before, eh?

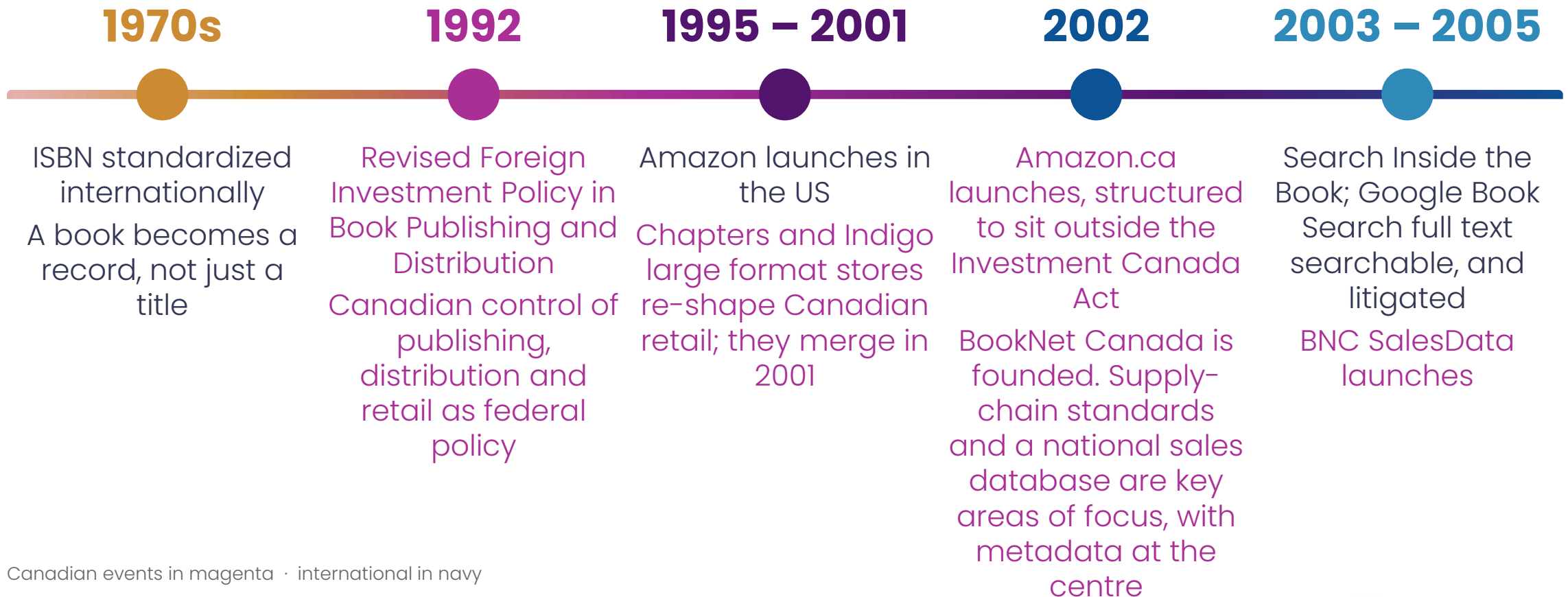
The book industry has been reshaped roughly every decade, and each new change or shift has moved discovery further from the physical book and deeper into the data



ECW Press, 2026

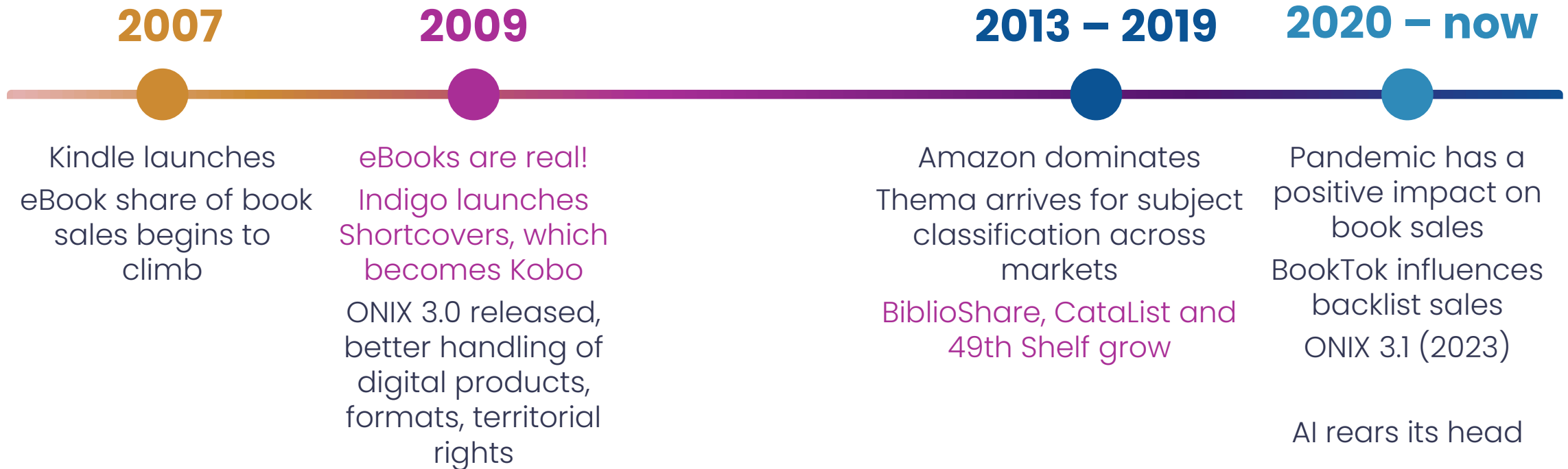
Identifiers, big box retail, and a Canadian data layer

1970s – 2005



Digital, scale, and generative AI

2007 – Today



Canadian events in magenta · international in navy

The pattern and what we've learned

The Pattern

- Discovery has increasingly, almost fully, moved off-the-shelf and into electronic product records
- Your metadata has been machine-read for 25 years. AI is not the first time machines have influenced the supply chain

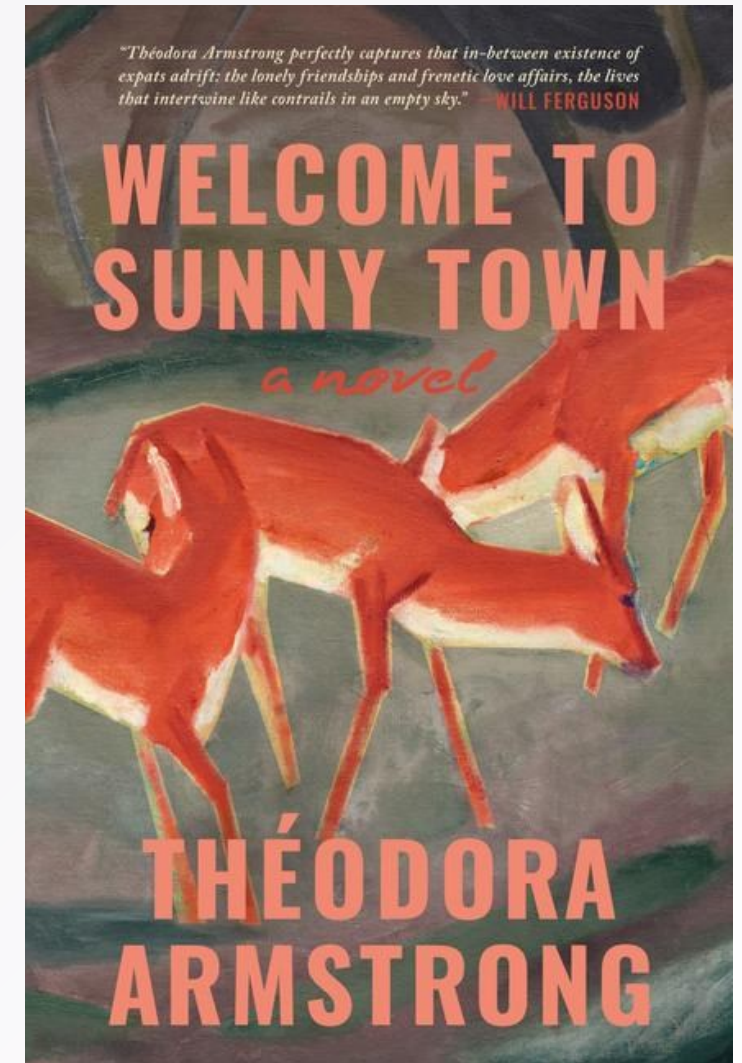
The Lesson

- For a quarter of a century, the best practice has been the same: maintain complete, comprehensive, structured, consistent, accurate, data
- Own your data and supply it early early

Where we are now

The book industry is now faced with a distinction we haven't really needed to make before:

**Metadata in Search vs.
Metadata in the Supply Chain**



Freehand Books, 2026

Three types of metadata

Library and information-science taxonomy

Descriptive

What the thing is

- Title and subtitle
- Contributors
- Subjects and other classifications
- Language(s)

Structural

How its parts relate

- Page count
- Chapter-level info
- Table of contents
- EPUB navigation
- Accessibility features

Administrative

How it's managed

- Rights (including for AI!)
- Technical specifications
- Provenance

Descriptive metadata gets almost all the attention. Structural metadata is what makes navigation and accessibility work, and administrative metadata is where AI provenance and text-and-data-mining terms now live.

Metadata and Search

For readers: the search box is now an answer engine

SEO vs. AI in search engines

- Individual keywords > Full sentences
- Matching and ranking > Semantic retrieval
- List of Links > A synthesized answer

What does that mean?

- What should I read next? is now typed as a question, not as keywords
- Search functions on retail sites now have AI layers acting on them
- Results come back and sometimes prompt the reader / user to refine their queries

Readers are now chatting with Search engines. Your data needs to be able to respond to questions, not just individual data points, categories and classifications.

Metadata in the Supply Chain

For suppliers: systems are looking at everything

Suppliers mine your metadata

Recipients already assign subject classifications and extract the data you send, even if it's not visible on product pages



Categorization is shifting

Suppliers now want to know more about how your books are made.



Thin records get interpreted poorly

Sparse or contradictory records are read badly and AI can make things up.

Your metadata is being interpreted, not just displayed. Supply chain systems used to pass through the data you sent, but they are now evaluating everything and “deciding” what the book is, and who it's for.

Where are we

Standards and rights

- ONIX already carries vocabulary and mechanisms for communicating AI involvement, as well as the publisher's terms for text and data mining
- The ONIX standard moved before most publishers and system implementations did
- Copyright and AI Text and Data Mining policy is being developed in companies and in every major market, including Canada

PROHIBITING TEXT AND DATA MINING

```
<DescriptiveDetail>
  <EpubUsageConstraint>
    <EpubUsageType>11</EpubUsageType>
    <EpubUsageStatus>03</EpubUsageStatus>
  </EpubUsageConstraint>
</DescriptiveDetail>
```

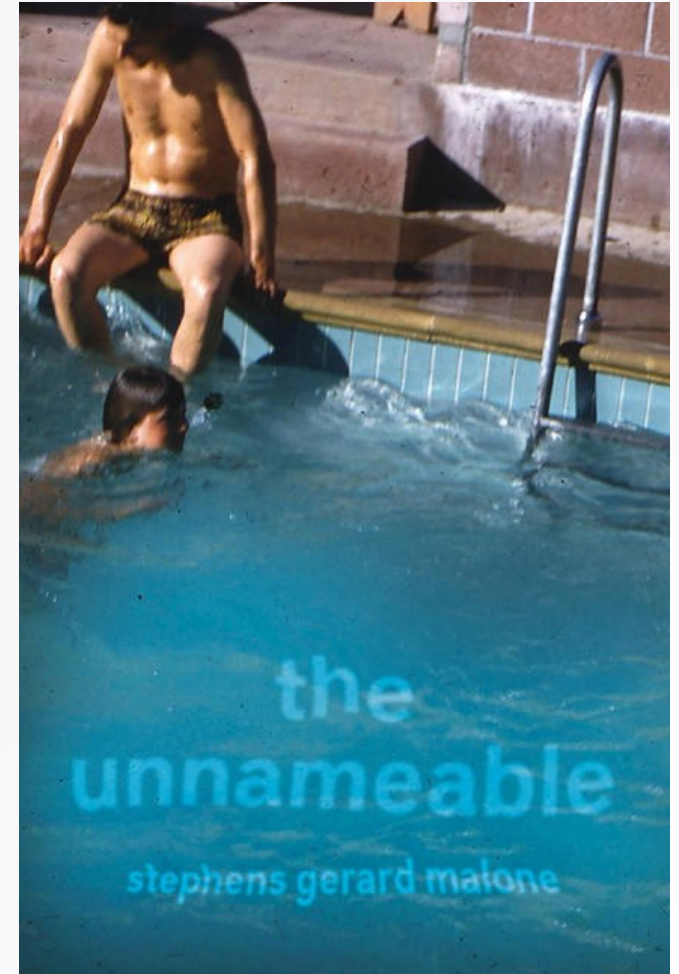
List 145, code 11 = text and data mining

List 146, code 03 = prohibited

The same composite expresses limits on copying, printing, lending and text-to-speech.

What nobody actually knows

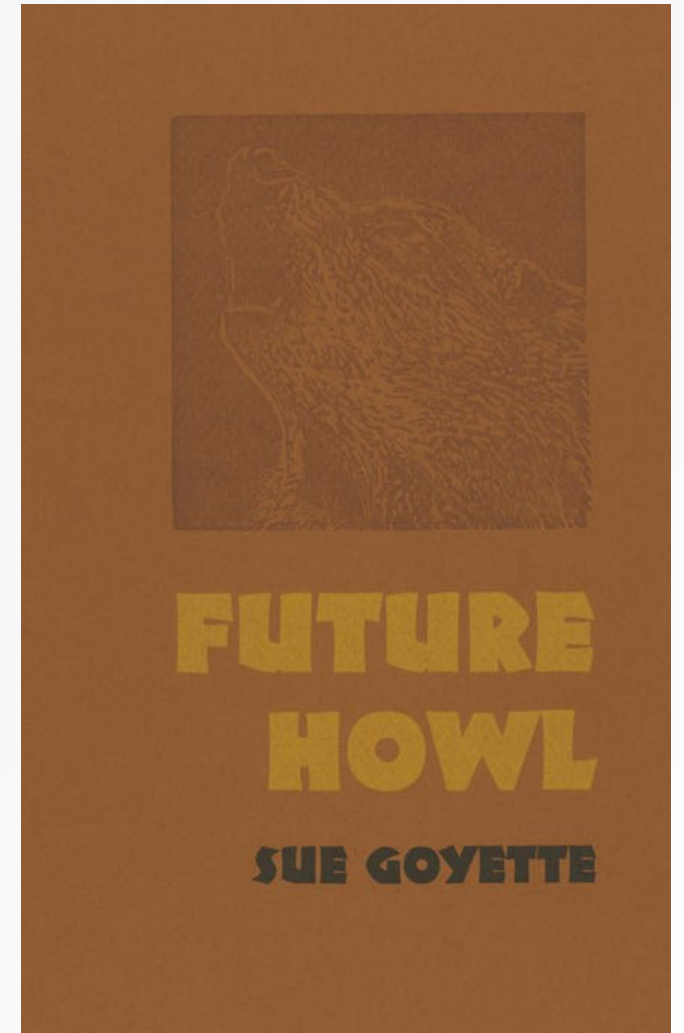
**We don't know how much
Book discovery flows through
AI channels now**



Nimbus Books, 2026

What's really new?

What should we do?



Gaspereau Press, 2025

Three things are new, now

Source material, Provenance and Potential AI Hallucination

Source

The words that make up our data are source material for discovery. This means we need to be clear with our data, concise, and leave less room for ambiguity.

Details

No previous shift asked for declaration about how the book was made, because the way books can be made has changed.

Consistency

Different metadata for different audiences used to be interesting, and sometimes worth doing. Today an AI model might find this confusing.

Specific Metadata Guidance

The devil is in the details (as always!)

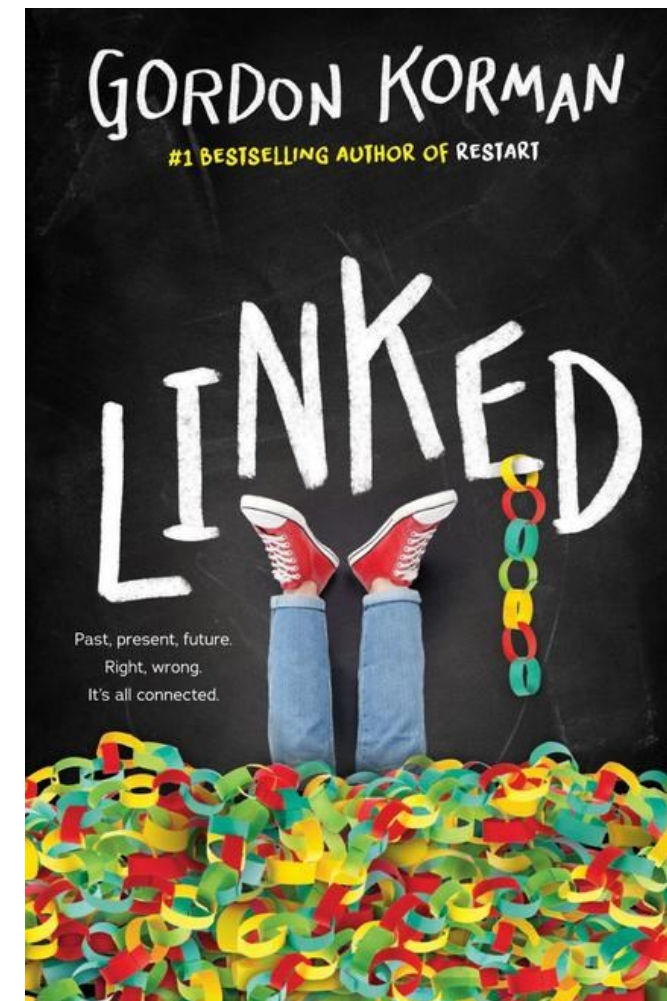
Stop treating metadata fields as optional

Supply as much metadata as you can, including:

- Copy beyond the Main Description (e.g. Table of Contents)
- Subject codes across multiple schemes, including Regional and Merchandising Themes
- Contributor biographies for all roles (with facts!)
- Identifiers
- Excerpts
- Reading Samples
- Series Name and Order

Series and Collection information

- A series is more than just a list of books in a numbered order, it is a collection of *linked* titles
- You are probably already creating your own Collections that group titles together based on marketing or other promotional efforts
- ONIX allows for proprietary collections to be sent in an ONIX message



Scholastic, 2021

Grouping titles for a seasonal promotion

A proprietary collection

A Publisher's Own Collection, identified with a Code and Title

```
<Collection>
  <CollectionType>10</CollectionType>
  <CollectionIdentifier>
    <CollectionIDType>01</CollectionIDType>
    <IDTypeName>Publisher collection</IDTypeName>
    <IDValue>SUMMER2026</IDValue>
  </CollectionIdentifier>
  <TitleDetail>
    <TitleType>01</TitleType>
    <TitleElement>
      <TitleElementLevel>02</TitleElementLevel>
      <TitleText>Summer Reads 2026</TitleText>
    </TitleElement>
  </TitleDetail>
</Collection>
```

● CollectionType 10

Publisher collection, from list 148.
Use 20 for a collection ascribed by a retailer or aggregator rather than by you.

● CollectionIDType 01

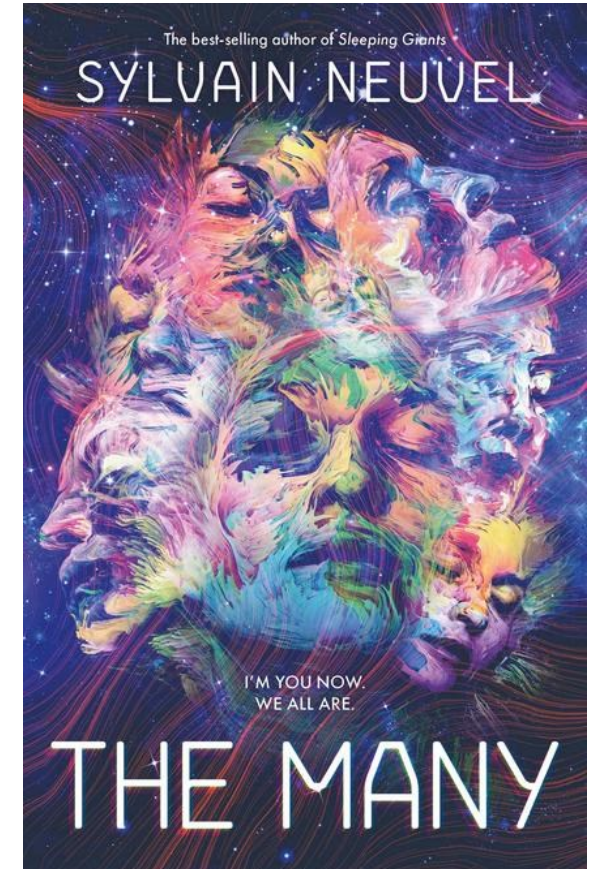
Proprietary, from list 13.
<IDTypeName> provides more information about who or what the collection belongs to.

● TitleElementLevel 02

Collection, from list 149. It marks the title text as the collection's name, not the title itself.

One set of keywords, many recipients

- Use phrases instead of single words, and put your most important keywords first because different recipients have different limits
- Write your keywords for the recipient with the tightest limit and make all your keywords or phrases count. Don't use "bestseller" or deliberate mis-spellings
- Your keywords should not exist anywhere else in your metadata, that other information is already being ingested and indexed (with one caveat about series)
- Don't do keyword stuffing. If a structured field exists for data, put data in that field



Solaris Books, 2026

More to consider, when adding to your metadata

Language and terms you can add to keywords and everywhere

01

Sub-genre subjects and categories

02

Tropes (think Romantasy)

03

Mood and tone

04

Place and Setting, including regional specificity

05

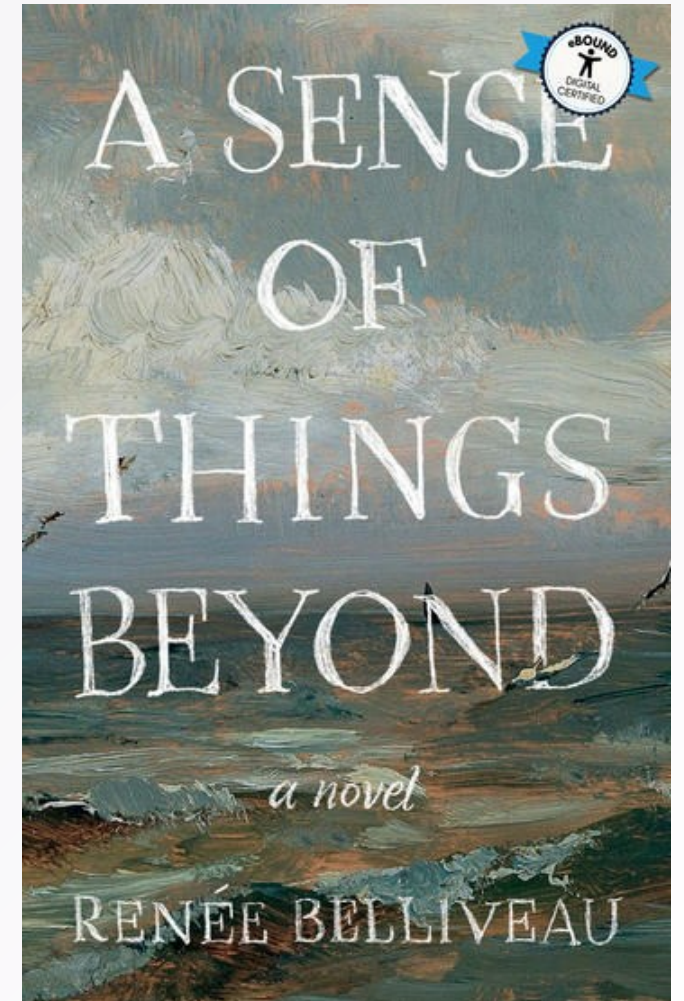
Holiday, occasion and gifting language

06

Relevant current events (but refresh often!)

And believe it or not ...

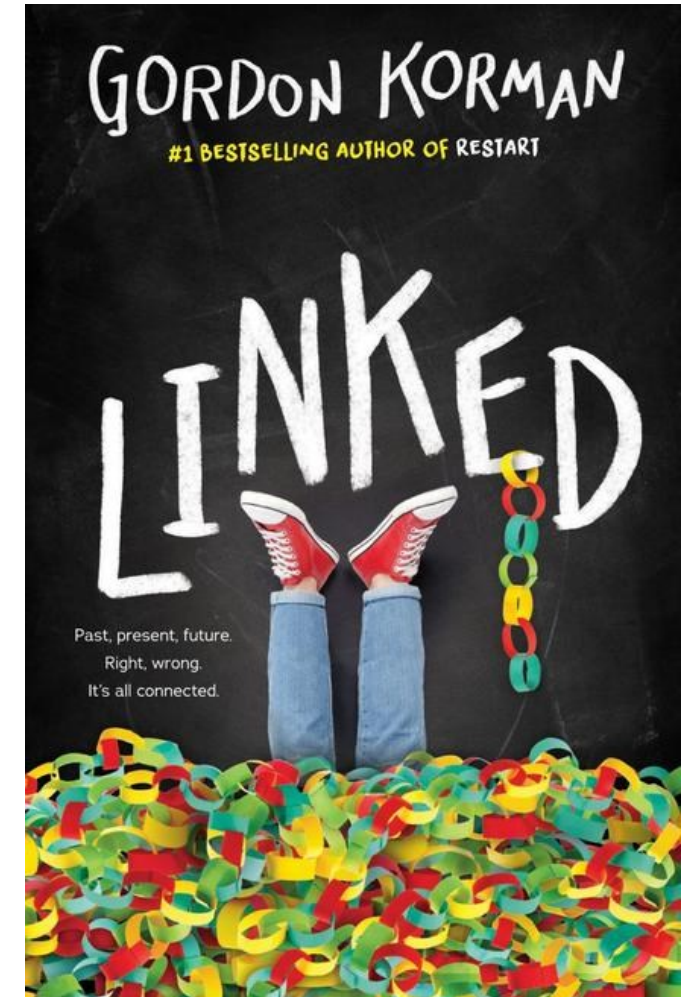
Publishers are adding emojis to their metadata.



Nimbus Books, 2026

Sales Points

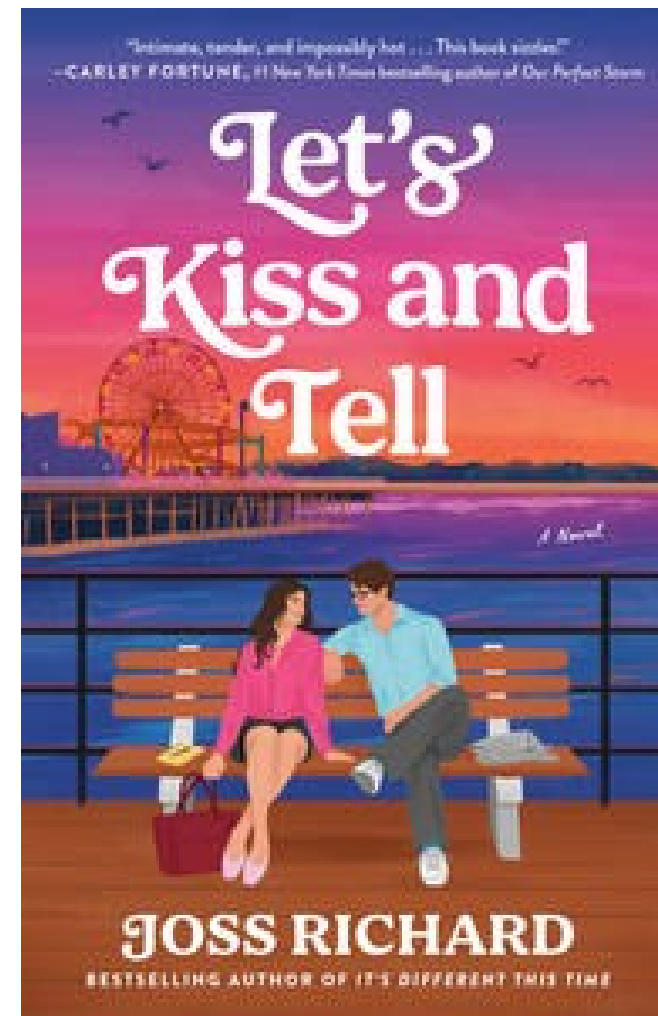
- Sales Points have traditionally been written with buyers as the audience, and with the assumption that you or your sales team will be in the room when titles are presented
- With AI in the proverbial room, we should attempt to write sales points that bridge the gap between a person in the room and the data we circulate. Assume the buying decision is being made without you.
- Take care with your metadata, even when it is not public-facing.



Scholastic, 2021

Experiment (and share!)

- Metadata is a renewable resource
- Maintain your own guidelines for writing keywords and copy. Know what you're trying to do, and how it fits the overall mission of the press or title
- Update your metadata frequently and track what happens when you send it out into the world
- Share what you know (please).



Penguin Canada, 2026

Evolving Standards

How ONIX and other industry standards help us cope

How will ONIX evolve?

Declaring AI involvement

- ONIX provides codes for communicating the involvement of AI, such as where contributors are assisted by AI
- This example shows two named, human contributors and a third, unnamed person which indicates AI
- Sequence is important here. It's the second contributor, the illustrator, that has been assisted by AI

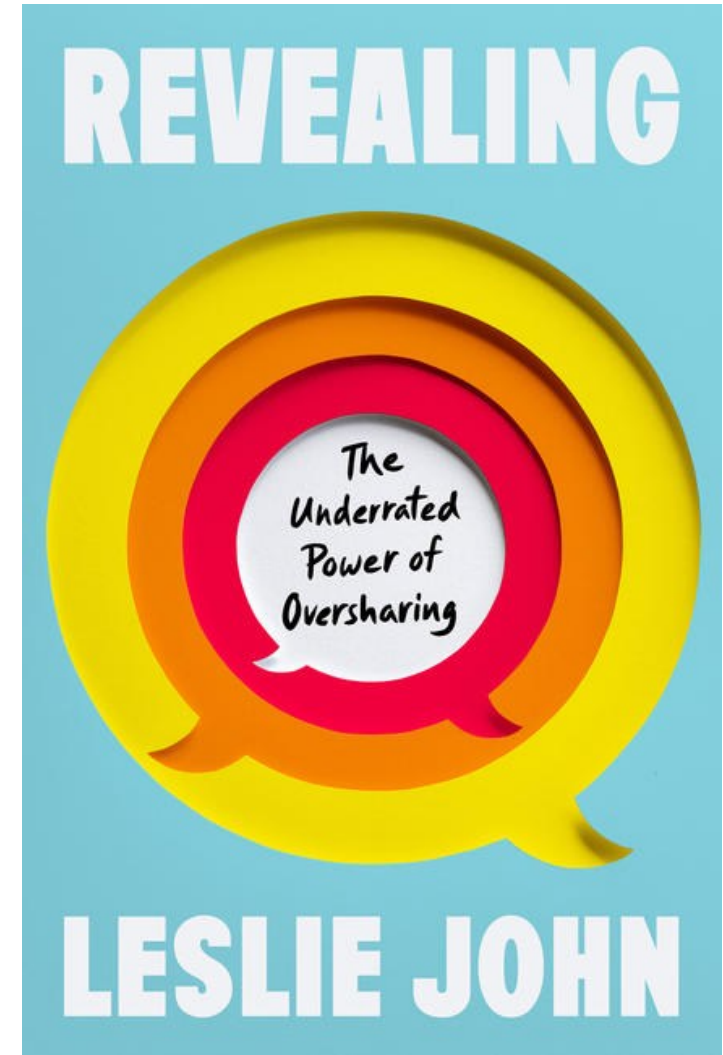
AI Involvement, worked example from EDITEUR

```
<Contributor>
  <SequenceNumber>1</SequenceNumber>
  <ContributorRole>A01</ContributorRole>
  <PersonName>James Green</PersonName>
</Contributor>
<Contributor>
  <SequenceNumber>2</SequenceNumber>
  <ContributorRole>A12</ContributorRole>
  <PersonName>Fiona Brown</PersonName>
</Contributor>
<Contributor>
  <SequenceNumber>3</SequenceNumber>
  <ContributorRole>Z01</ContributorRole>
  <UnnamedPersons>09</UnnamedPersons>
</Contributor>
```

How will ONIX evolve?

Provenance and rights

- ONIX is expanding to hold more information about the books we write, make, and read. It is no longer just our marketing message
- The ONIX message can contain not only information about AI involvement or permissions, but also about accessibility and specification (POD)
- EUDR (deforestation regulations) and GPSR (product safety) data are expected to be communicated in ONIX



Penguin Canada, 2026

Is Metadata Content?

What's the difference now? Who has rights to my data and content?

Your Metadata

- You want it circulated as widely as possible
- ONIX is the carrier for your metadata, and sending ONIX results in an implied license to ingest and mine your metadata
- This means your metadata can always be mined, even if you opt out of text and data mining for a book

Your Content

- You want it used on your terms
- Express constraints in ONIX, and on your website
- Consistency is key, it limits your exposure

Written by authors, not AI

**How can we indicate our titles are written by humans,
and not by AI?**

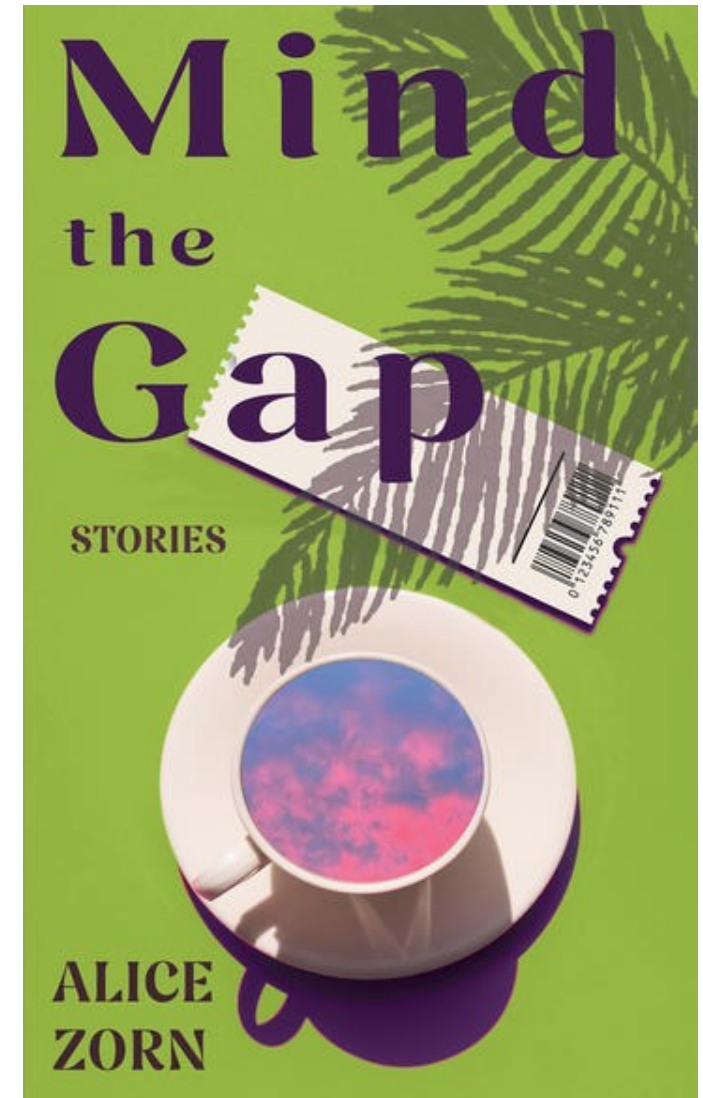
The Long Tail

What to do for backlist titles and where libraries fit in

The Long Tail

Backlist title approach

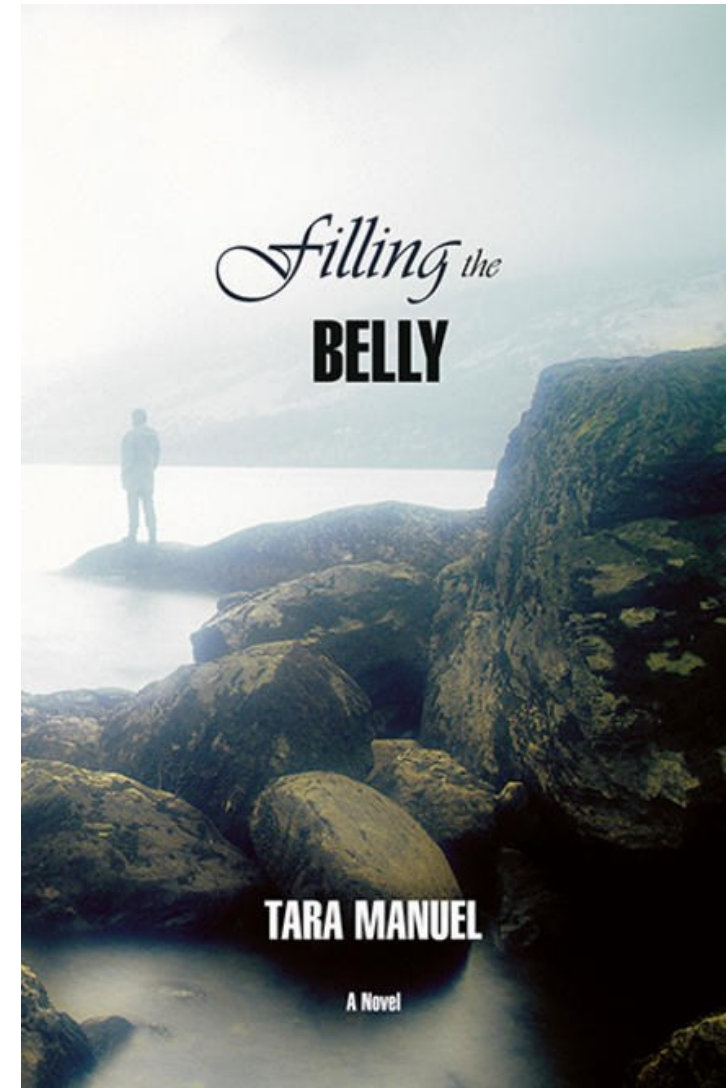
- Backlist titles often have the least complete metadata
- Create and run compliance reports for metadata completeness
- Look for incorrect or incomplete metadata, not just what's missing
- Work the gaps, not the individual titles.



NeWest Press, 2026

What are some of the gaps?

- Add Awards won since publication
- Links to Related Products (Works and Editions), Series, Contributors. These links are what pull your titles along when new books come out
- Current hooks, updated language
- Anniversaries (everybody loves a comeback!)

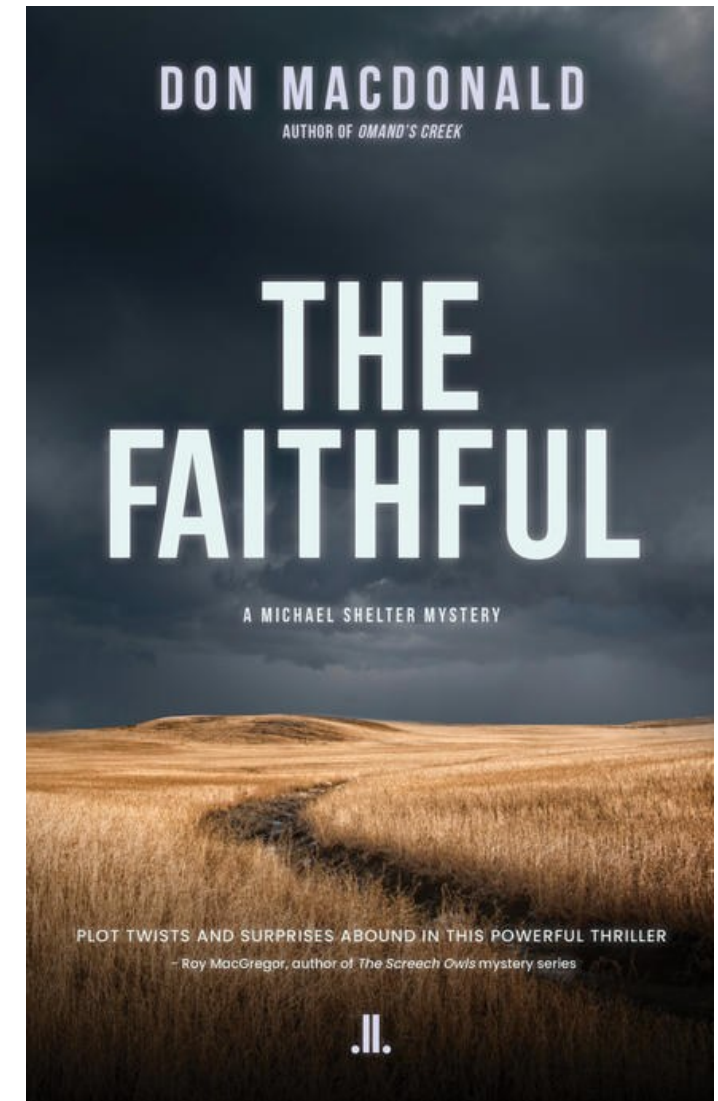


Thistledown Press, 2023

The Long Tail: what can Libraries do?

Authority and discipline

- One consistent form of the author or contributor's name, everywhere
- A consistent series name, everywhere
- Use identifiers: The International Standard Name Identifier (ISNI) and Orcid IDs help to disambiguate individuals
- Curation, look to programs like Lone Stars



Linda Leith Publishing, 2026

Summary

Three take-aways

- Fill the fields you've been skipping, not just for frontlist but also for your backlist titles. Take care to craft the data and get curious about what happens when you send it out.
- Keywords should contain information the rest of your data doesn't. Nothing else.
- Say what's true about provenance in the structured fields for that data.



ECW Press, 2026

Thank you.

Tricia McCraney
tricia@virtusales.com