ISTC IN PRACTICE

Lessons learned from an early implementation

Graham Bell, EDItEUR (formerly Head of Publishing Systems HarperCollins Publishers)
The Blue Train
hardback

The Blue Train
paperback

The Blue Train
audiobook on CD
Agatha Christie’s
The Blue Train

The Blue Train
hardback

The Blue Train
paperback

The Blue Train
audiobook on CD
Agatha Christie’s
*The Blue Train*

*The Blue Train*
hardback

*The Blue Train*
paperback

*The Blue Train*
audiobook on CD

‘work’ with ISTC

‘manifestation’
with ISBN

‘manifestation’
with ISBN

‘manifestation’
with ISBN
work

manifestations
parent work

manifestations

work

manifestations
parent work

manifestations

work

manifestations
original work

many manifestations...

...and other publishers
original work  translated  adapted  translated

many manifestations... ...and other publishers
ISTC REGISTRATION

- XML messages between publisher (or other interested party) and registration agency

- request contains work type, derivation, title, subtitle, edition, work date, author name(s) and role(s) plus manifestation info

- reply will include ISTC – which may be brand new, or an ISTC previously issued to that work

- you can create an ISTC registration message from existing metadata...
IDENTIFICATION OF A ‘WORK’

• you may already have a work identifier (eg ISBN of first published version)

• does your model match that of the ISTC?
  • are abridged audiobooks new works?
  • are illustrated editions new works?
  • is an annually revised publication a succession of new works?

• do you model the concept of a parent work?
IDENTIFICATION OF A 'WORK'

- You may have a work identifier (e.g., ISBN) already
- Does your model match the ISTC?
- Are abridged audiobooks new works?
- Are illustrated editions new works?
- Is an annually revised publication a succession of new works?
- Do you model the concept of a parent work?
Recognising Derived Works from Metadata

- same title, same contributors (but what about subtitle?)
- derived works have translators, abridgers, multiple parent works etc
- three sets of Onix for Books edition types
  - abridged, adapted, enlarged imply derived
  - facsimile, unabridged, tie-in, Braille do not
- new? revised? special?
Recognising Derived Works from Metadata

- does your metadata include previous title?
- does your metadata include ‘excerpted from’, or any indication of the original language title of a translated work?
- be realistic – just how clean is your data?
Recognising Derived Works from Metadata

• Does your metadata include an excerpt from an original indication of the original language title of a translated work?

• Be realistic – just how clean is your data?
Recognising Derived Works from Metadata

- Does your metadata include the previous title?
- Does your metadata include an 'excerpt' from the original title of a translated work?
- Be realistic - just how clean is your data?
Recognising Derived Works from Metadata

• Does your metadata include previous title?
• Does your metadata include an excerpt from or any indication of the original language title of a translated work?
• Be realistic – just how clean is your data?
Matching in the registration system

- exact matches of title, contributors *etc*
- fuzzy matches to account for minor errors in metadata embedded in request message
  - fuzzy matches not at all ‘smart’ (*eg* Soundex)
  - and some are downright mystifying!
- moral of the tale – currently needs manual checking to ensure integrity 📝
Matching in the registration system

• Exact matches do not contribute fuzzy matches to account for minor errors in metadata embedded in a request message.

• Fuzzy matches are used in general and are made down-right mystifying.

• More of the tale’s current need for manual checking to ensure integrity.
MATCHING IN THE REGISTRATION SYSTEM

• Exact matches are set as title and contributor set.
• Fuzzy matches are accounted for in minor errors in metadata embedded in a request message.
• Fuzzy matches note the use of, for example, Soundex.
• And some are downright mystifying.

...morale of the tale—currently need manual checking to ensure integrity.
Recognising works from the metadata

- This standard is new, and there are kinks still to be worked out.
- Strict or loose interpretation, with regard to editing, annotations, value-added content?
- How much effort justified in handling of short stories, poetry and other compilations?
- Reliable collocation could change the way books are sold online, and the way costs and revenues are analysed.
RECOGNISING WORKS FROM THE METADATA

- The standard is new and there are still issues to be worked out.
- Strict or loose interpretation with regard to editing and annotations?
- How much effort is justified in handling short stories and poetry and other compilations?
- Reliable collocation could change the way books are sold online and the way costs and revenues are analysed.