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
work → derived work → manifestations

manifestations
parent work → work

manifestations → manifestations

deriveddwork → deriveddwork

work → manifestations
original work

many manifestations...

...and other publishers
original work  translated

many manifestations...  ...and other publishers
original work  translated  adapted  translated

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

• XML message from publisher or other party to registration agency

• request/reply ‘conversation’

• request contains work type, derivation, title, subtitle, edition, work date, author name(s) and role(s) plus info on one or more manifestations

• 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

for example, you can derive the ‘work type’ reliably from the BIC (but maybe not BISAC) subject codes you already have

mostly prose, but also lyrics, poetry, scripts

but there are some tricky areas...
IDENTIFICATION OF A ‘WORK’

• you may already have a work identifier (e.g., 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 already have a work identifier (e.g., ISBN of first published version)
- does your model match the ISTC?
- are abridged audiobooks new works?
- are illustrated editions new works?
- is any revision of a work a new work?
- 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?
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?

but also possible to submit too much data…
RECOGNISING DERIVED WORKS FROM METADATA

- Does your metadata include previous title?
- Does your metadata include an excerpt from the original language of the translated work?
- Is your data realistic—just how clean is your data?
- But also possibly to submit too much data…

Recognising Derived Works from Metadata
• Does your metadata include previous edition title?

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

• Is it realistic – just how clean is your data?

• But also possible to submit too much data...

Recognising Derived Works from Metadata

THOMAS EIDSON

The Last Ride

‘A gripping novel’ Sunday Express

THOMAS EIDSON

THE MISSING

NOW A MAJOR MOTION PICTURE FROM RON HOWARD,
STARRING TOMMY LEE JONES AND CATE BLANCHETT
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?
- but also possible to submit too much data...
Recognising Derived works from Metadata.
• does your metadata include previous title?

• does your metadata include an excerpted from’, or a different language translation of your work?

• be realistic – do you need all your data?

• but also please be aware that submitting too much 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 of title and contributors etc.

• Fuzzy matches to account for minor errors in metadata embedded in the request message.

• Fuzzy matches not all 'smart' like Soundex, etc.

• And some are downright mystifying.

• Moral of the tale – currently needs manual checking to ensure integrity.
MATCHING IN THE REGISTRATION SYSTEM

- Exact matches of title and contributors etc.
- Fuzzy matches to account for minor errors in metadata embedded in the request message.
- Fuzzy matches not dataloggers' 'smart' regex
- And some are downright mystifying.
- Moral of the tale: currently needs 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?
- ISTC of *this work* may not be as useful as the ISTC of the *parent work* or the *ur-work*
RECOGNISING WORKS FROM THE METADATA

- This standard is new and the red are 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?
- ISTC of the work may not be as useful as the ISTC of the parent work or the art-work.