New Developments in Bibliographic Services: A Report from Bowker

Angela D’Agostino
Senior Vice President
Business Development
Why is Bowker here – changes within the publishing landscape

Identifiers for digital content
  – Recommended standards for in copyright material
  – Grant program for out of copyright material

Metadata
  – Digitized books
  – E-book collections
  – Using full-text to enhance bibliographic records

What’s in it for You
  – Standards
  – Improved search and discoverability of your digital collections
Why Bowker?

• ISBN agency for U.S. and Australia
• Board member of American Association of Publishers (AAP)
• Registration Agent for International DOI Foundation (IDF)
• Founding member of International Standard Text Code (ISTC)
• Committee member on multiple Book Industry Study Group (BISG) working groups devoted to metadata, identifiers and other standards
Implications of Digital Publishing and Mass Digitization

- Premise: Digital publishing and digitization projects produce more products from the same content
  - Whole books
  - Fragments of books (e.g. chapters or sections)
  - Fragments of journals (e.g. articles)
  - Full-text hosted books in multiple locations

- Regardless of the choice of product identifier used, there will be:
  - More products to describe
  - More products related (often equivalent) to one another

- There will be a greater need to aggregate this information in order to manage and use it
Some Issues

• How granular should publishers make content?
  – What impact on a book’s sales if component chapters available individually?

• What is the most popular digital format going to be?
  – Will consumers purchase digital content if not available in their preferred format?

• What if there is no predominant format?
  – Must publishers make content available in every format, or could they provide retailers with raw content?

• How should multiple editions of the same content be presented?
  – Consumers must be able to discover and select the “right” product

• How are libraries identifying digital works and enabling them to be searched and discovered
Identifiers for Digital Content
Identifiers -- Overview

- ISBNs are a standard that work for books – over 250 million have been assigned worldwide since 1968
- ISBNs are assigned to e-books – different ISBN per platform or reader
- ISBNs are assigned to every new manifestation -- including digital
- ISBNs enable common framework for cataloging, distribution and discoverability
- ISBNs can and should be assigned to chapters and “fragments”
- ISBN + meta-data enriches catalogues for patron discoverability via online platforms
- Digital Object Identifiers (DOI)
  - Applicable to networked books, wiki’s, blogs, cross-reference-able book citations, images and other dynamic content
  - Persistent linkages and identification which can be used with any standard numbering system
  - DOI + ISBN = actionable ISBN; resolve to full-text or snippets
  - DOIs are not the standard numbering schema for pieces and parts
Recommendations for uniquely identifying In Copyright Material

- ISBNs assigned to every manifestation of digital work – by publisher
- ISBNs assigned to every chapter or piece of a book that is made available for sale – by publisher
- DOIs assigned to book material that is dynamically changing (users adding comments or other regular updating of online content) – by content owner
Grant Program for Out of Copyright Works

• Bowker is donating 3 million ISBN assignments to libraries engaged in digitization projects
• Libraries to provide Bowker with metadata on digitized works
• Criteria for grant:
  – Library owns book
  – Books is Out of Copyright
  – English-language material
  – Book will be made publicly available
• In discussions with Google partners, Microsoft partners, Internet archive and other OCA members
Introducing the ISTC

- **International Standard Text Code**
- Unambiguously identifies a textual work, even though it may be published in many different forms
- ISTC is an *attribute* of a product, *not* a product-identifier
- For use in administering copyright, licensing, collocation, improved discovery services and sales analysis
- Identify content separately from the products which contain it
- Identify the relationships between these items of content
- Identify the relationships between content and products
Metadata is mandatory in order to prevent duplicates. It includes:

- Language of text
- Title
- Contributors
- Source (if derived from another work)
- Registrant & their role
- Date of registration

Optional metadata includes:

- Alternative titles
- Nominal date of creation
- Type of work
- Date first published
ISTC Structure & Capacity

- Structure allows up to 4,096 distinct registration agencies
- Each agency could register up to 4,294,967,296 annually
- Provides a theoretical capacity to register up to 17,592,186,044,416 each year
Who assigns?

- In-copyright works will be assigned by the rights-holder or their representative: publisher, author or agent
- Most registration agencies will be operated by bibliographic agencies and Reproductive Rights Organizations (ex. CCC)
- Out-of-copyright works may be assigned by anyone, but most likely libraries, who could operate their own registration agencies
- Links between works (ISTC) and products (ISBN) will be captured on bibliographic databases
A Note on Costs

- As with ISBNs, ISTCs will not involve any ongoing costs for support or maintenance.
- Initially they are likely to be made free of charge, although cost-recovery will probably become necessary.
- Individual registration agencies may charge for value-add services.
Current status

- ISO draft standard is approved & pending formal publication Winter 2007
- International authority being formed by consortium of Nielsen, RR Bowker, IFRRO and CISAC
- Web-based registration system already developed and undergoing system testing
- Expect registration agencies to start operating early 2008
- Bowker working with BISG metadata committee, establish best practices around hierarchical linking of parent (ISTC), primary commercial elements (book/ISBNs) and component parts (chapters/ISBN with linking methodology)
Bowker’s Identifiers 2.0 Model

- Identifier Registration and Auto-Assignment
- Content “chunking” solutions with metadata capture and assignment
- Digital “bar code” creation
- Channel Distribution of DBCs (search engines, retailers, libraries, e-tailers, e-book registries, etc.)
- Other Discoverability Services (‘deep indexing’ profiles, enriched metadata, user tagging, digital rights management)
- “Digital Books In Print” for digitized works
Metadata
“Books” In Print?

• Currently Bowker’s Global Books In Print Database is over 16 million records – in print and out of print books, audiobooks, e-books, videos, DVDs

• Plan is to extend the Books In Print to include metadata about any digital content (alongside the physical):
  – E-books available for free or purchase, sold individually or as part of collections (ex. Safari, Netlibrary, Ebrary, Sony, etc.)
  – Print on demand titles (already have Lightning Source, extend to other vendors)
  – Digitized books through Google, Microsoft, libraries and others (public domain and in copyright)
  – Digital documents (ex. O’Reilly PDFs)
  – Work elements (ex. Thomson’s and Taylor & Francis’s chapters or other component pieces of books)
Why Books In Print?

• “Database of Record” for Barnes & Noble, Borders, Follett, Sony, Microsoft
• Extensive reach throughout public and academic institutions worldwide as well as retailers, online shoppers and consumers
• Currently track who has physical and e-book available for sale (ex. Ingram, B&T, Blackwell)
• Facilitate collection analyses through Bowker’s Book Analysis against core lists such as Resources for College Libraries
• Syndetics Solutions will index content relevant to library’s collection (ex. end users can search table of contents of books held in your library)
• Enriching bibliographic record with keywords, commonly used phrases from full-text – proof of concept underway
Benefits to librarians and patrons

• Search:
  – Enhance discoverability of relevant content through OPAC (searchable TOCs, descriptions, profiles, keywords)

• Find:
  – Links to the full-text of the digital work
  – One-stop availability of multiple links/availability of content (physical, e-book chapters, PDFs, etc.)
  – Compare digital assets to print assets for weeding or collection development (books in Safari compared to physical collection)

• Consume:
  – Read, download, purchase, listen, view
  – Professionals can order through wholesalers/distributors
  – Consumers can order from retailers or other online sites
What’s in it for you?
Standards

- ISBNs to digitized works to help identify and provide another “hook”
- Potentially a grant of ISBNs to those of you who qualify
- Links to other manifestations of works potentially through ISTC or through metadata algorithms
- A means to easily enter your content into supply chain when you are ready to monetize
Search and Discoverability of Digital Collections

• Adding enriched metadata to OPAC to help in relevancy and targeted searching
• Adding links to MARC records to point users to full-text
• A means of identifying works available in digital format in or out of copyright
• A means to enable wider discoverability of your collections via Books In Print or DOI registry, complementary to other bibliographic and search tools
Libraries & Colleges
Angela D’Agostino
Senior Vice President, Business Development

- Phone:  (908) 219-0150
- Email:  angela.dagostino@bowker.com

- Corporate information for Bowker
  - R.R. Bowker LLC
  - 630 Central Avenue
  - New Providence, NJ 07974, USA
  - Tel.: +1 (908) 464-6800
  - Toll-free (US and Canada): +1 (888) 269-5372
  - Web:  http://www.bowker.com