• Challenges for Libraries, Publishers and Vendors
• Strategies for Meeting the Challenges
• “Next Generation” Cataloging and Metadata
Challenges for libraries, publishers and vendors

• We must admit that the current models are unsustainable
  • LC Working Group on the Future of Bibliographic Control recommendation
    • 1. Increase the efficiency of bibliographic production and maintenance
      • 1.1.1 Make more use of bibliographic data earlier in the supply chain
But the publisher supply chain experiences challenges in metadata creation and management too!!

We need to work together to increase the efficiency of bibliographic production and maintenance for the publisher supply chain as well.
Challenges for Libraries, Publishers and Vendors

- Too much stuff!
  - Explosion of materials and formats
  - Faster publishing cycles
  - Multiple sources for metadata
  - Multiple metadata formats and standards
- Users expect fast web exposure of new materials, easy information retrieval and immediate access to materials
Challenges for Libraries, Publishers and Vendors

• Metadata creation is expensive and labor intensive but the danger of hidden materials is greater
  
  • For libraries:
    • No access or limited access to materials
    • Delayed patron access to in-demand materials
    • Implications for collection analysis, selection and reporting
  
  • For publishers:
    • No metadata, no sales!!!
    • Incomplete or incorrect metadata = missed sales, poor business intelligence
Challenges for Libraries, Publishers and Vendors

• There is still enormous duplication of effort in work on metadata for the same set of titles
  • In libraries:
    • Complex local practices, local editing and other manual manipulation of existing metadata
    • Not all metadata creation or enhancement is shared
    • Fewer staff to create metadata upon receipt of items
    • Multiple libraries (and vendors) may be doing work on the same titles before a record appears in cooperative databases such as WorldCat
Challenges for Libraries, Publishers and Vendors

- In the publisher supply chain
  - Staff and systems for publisher creation and enhancement of metadata
  - Staff and systems for extensive review and manipulation of metadata at retail, wholesale and metadata aggregation vendors
  - Staff and systems to add library-specific metadata used in library vendor programs and ordering tools: web-based ordering, selection lists, approval plans, etc.
Challenges for Libraries, Publishers and Vendors

- In the publisher supply chain
  - Staff for MARC record creation
    - Many library vendors create MARC records or outsource for MARC record creation in addition to creating, enhancing and manipulating data used for marketing and ordering
    - Many of these records end up in proprietary systems and are not shared
Publisher Supply Chain Data Flow Today

U.S. Publishers

Publisher Metadata Staff, Systems, Data feeds, etc.

End users

Data Aggregators, Bowker, Nielsen, Ingram, etc.

Metadata Staff

Libraries

Material Wholesalers, B&T, Ingram, Border, Etc.

Metadata Staff

Retailers, Barnes & Noble, Amazon, Borders, Independent Bookstores, Etc.

Metadata Staff

End users

Library of Congress

Libraries
Strategies for Meeting the Challenges

- We can’t continue to “silo” library metadata and metadata practices
  - Re-mix and re-use existing metadata
  - Increase collaboration and cooperation between library and publisher supply-chain communities
  - Break down barriers between metadata used for acquisitions and metadata used for discovery, business intelligence and collection management
  - Become more involved in upstream metadata creation processes, integrate available metadata into workflows upstream and allow the metadata to evolve over time
Strategies for Meeting the Challenges

- Metadata management workflows and practices must change to allow easy ingest and use of existing metadata
  - Reduce practices that require manual manipulation of existing metadata
  - Allow different levels of metadata based on material type, user needs
  - Allow metadata for new titles to “grow up” over time
Strategies for Meeting the Challenges

- Solutions must be interoperable and easily shared - inside and outside the library community
  - The library community must extend expertise, as well as our cooperative and collaborative practices, to include publishers and publisher supply chain partners
  - We must find ways to create, ingest and share multiple types of metadata
    - Libraries must become more open to the use of non-MARC data, non-library vocabularies, etc.
    - Publishers must find ways to leverage library data including classification, terminologies and authorities
“Next Generation” Cataloging & Metadata Services
Data Flow
“Next Generation” Cataloging and Metadata Creation Pilot

- Automated capture, crosswalk and enhancement of publisher ONIX metadata
  - Output in MARC and ONIX to benefit both library and publishing communities
  - OCLC pilot program with publishers, vendors and libraries January-June 2008
  - Press release and additional information here: http://www.oclc.org/productworks/nextgencataloging.htm
“Next Generation” Cataloging and Metadata Creation Pilot

- Library Pilot Partners
  - Phoenix Public Library
  - The Ohio State University Libraries
  - Chicago Public Library
  - MIT Libraries
“Next Generation” Cataloging and Metadata Creation Pilot

- Publisher/Vendor Pilot Partners
  - Ingram Book Group
  - Princeton University Press
  - Hachette Book Group
  - Taylor & Francis
“Next Generation” Cataloging and Metadata Creation Pilot

Advisory Board

Paul DeAnna, National Library of Medicine
Phil Schreur, Stanford University Libraries
David Williamson, Cataloging in Publication Division, Library of Congress
John Chapman, University of Minnesota Libraries
Michael Norman, University of Illinois, Urbana-Champaign
Laura Dawson, Consultant to the publishing industry
Nora Rawlinson, Publishing industry veteran and consultant to libraries on collection development
Kevin Clair, Penn State University Libraries
Richard Stark, Barnes & Noble
Marlene Harris, Chicago Public Library
“Next Generation” Cataloging and Metadata Creation Pilot

• How the pilot works

  • Publisher and vendor pilot partners provide title data in ONIX (ONline Information EXchange)
    • XML standard used by the publishing industry to share metadata among players (e.g. publisher → Amazon or Barnes & Noble)
  • OCLC converts the data to MARC
  • The metadata is enriched through data mining of WorldCat and data mapping from existing data elements
  • The resulting MARC record is added to WorldCat
  • Library pilot partners give feedback on the records added to WorldCat
  • The enhanced metadata is converted back to ONIX and returned to publisher/vendor pilot partners for review and feedback
“Next Generation” Cataloging and Metadata Creation Pilot

Methodology

- The OCLC crosswalk from ONIX to MARC has been enhanced to capture as many data elements as possible
- ONIX records are converted to MARC
- The system attempts to find an exact match to a WorldCat Record
- If an exact match is found, the existing record is enhanced by adding appropriate publisher/vendor metadata from ONIX to the record and by mining WorldCat using processes described below
- If no exact match is found, a record is created using the incoming ONIX metadata and enhanced using processes outlined below
“Next Generation” Cataloging and Metadata Creation Pilot

- Methodology
  - After a MARC record is identified or created, we mine WorldCat to retrieve the FRBR work-set
  - Using hierarchies and filters to determine the best records in the work-set for various data elements, we add or replace data in the new record or enhance an existing record
    - Contributor names
    - Dewey and LC class numbers
    - LCSH
    - Notes
    - Etc
“Next Generation” Cataloging and Metadata Creation Pilot

- **Methodology**
  - As possible, we map between classification and terminologies to add additional subject metadata
    - Mapping between DDC and BISAC subject headings is underway
    - We will continue to add additional mappings
  - The resulting MARC record is available in WorldCat for library use
  - An OCLC crosswalk from MARC to ONIX has been developed
  - The enhanced MARC record is crosswalked to ONIX and returned to the publisher or vendor in ONIX
“Next Generation” Cataloging

Vendor ONIX Metadata for All Resources
- ISBN
- Previous ISBN (if known or applicable)
- Author
- Title
- Imprint
- Format: audio, e-resource, hardcover, paperback, etc.
- Other metadata carried in ONIX format

WorldCat Record for Previous version, format or edition
- Fixed Field
- LC & Dewey call numbers
- Authorized headings
- Title
- Imprint
- Notes (e.g. summaries, track listings, etc.)
- Subject headings

Template System
- Control fields (006/007)
- GMD
- Applicable data elements based on format: templates for e-resource, audio, video, etc.

New record added to WorldCat
- Fixed Field
- ISBN
- Vendor ID number
- LC & Dewey call numbers
- Author (authorized)
- Title & Subtitle
- GMD
- Imprint
- Format specific information
- Series (authorized)
- Notes
- LCSH
- Other terminologies including BISAC
- Added entries (authorized)

Vendor data returned in ONIX enhanced with:
- OCLC control number
- Original vendor data if not replaced
- Authorized names
- Authorized series
- Authorized LCSH
- Other subject terminologies including BISAC
- Call numbers
- Notes (as applicable)

Results in ...

Attempt match to

Go thru

Return
“Next Generation” Cataloging and Metadata Creation Pilot

- Measuring and Reporting Pilot Results
  - Statistical analysis of record ingest, creation and enhancement activities
  - Pilot Advisory Board provides input into development of evaluative tools and measures
  - Pilot partners use evaluative tools to provide feedback on record quality and usefulness in metadata creation and management processes
  - Case studies are created for each pilot partner
  - OCLC, pilot partners and Advisory Board recommend next steps
  - Report on pilot results at ALA Annual 2008
“Next Generation” Cataloging and Metadata Creation Pilot

- Pilot Progress January-May 2008
- ONIX to MARC crosswalk enhanced
- MARC to ONIX crosswalk created and enhanced
- Rules and hierarchies defined for data mining using FRBR work sets
- Software developed to perform record creation and enhancement actions
- Began receiving ONIX metadata from partners
“Next Generation” Cataloging and Metadata Creation Pilot

- Pilot Progress January-May 2008
- Developed evaluative tools and case study templates
- Collected statistical information on metadata received and enhanced
- Refined algorithms, crosswalks and software based on live metadata results
- Began mapping between DDC and BISAC subject headings
“Next Generation” Cataloging and Metadata Creation Pilot

- Pilot Wrap-up -- May-June 2008
- Return metadata for library and publisher/vendor pilot partner review
- Complete statistical analysis and compile pilot partner evaluation results
- Complete case studies of pilot partners
- Share results with pilot partners, advisory board
- Share results with library and publisher supply-chain communities
Pilot update at ALA Annual

“New Directions in Cataloging at OCLC”

Sunday, June 29

1:30-3:30PM

Anaheim Convention Center, Room 208B

Registration is required

www.oclc.org/info/ala/
Contact Information for “Next Generation” Cataloging & Metadata Pilot

Renee Register

register@oclc.org
614-764-6107

Maureen Huss

hussm@oclc.org
614-764-4327
Thank You!

Questions?