

Envision Tomorrow's Catalog: the evolution of cataloging in the e-commerce era

Cindy Cunningham
U.S. Catalog Librarian, Amazon.com
July 14, 2003, Seattle, WA
American Association of Law Librarians

Overview

- How cataloging has evolved,
- Collaboration and multiple sources has shifted the
nexus of authority
- Creating records using normalization and
reconciliation
- New sources of authority and new definitions for
content
- Retail developments:DOI, Onix, and streaming
video content—new formats and content
definitions
- What does the catalog of the future look like?

Evolution of Cataloging

- Limited number of authoritative sources
- Collaborative rule-making by standards bodies
- International standard for data format and subsequent cataloging practices
- OCLC and shared cataloging started a trend towards multiple authorities, but everyone was still following the same rules

E-Commerce brought changes:

- Amazon started July 16, 1995 with a single catalog source: Baker & Taylor
- Soon, Ingram was added, and our process of reconciling multiple data sources began
- Because Amazon encouraged catalog contributions, data came in from everyone—the independent publisher, to the large publishing houses who wanted to influence how their data was presented.
- Data quality became more and more important as its relationship to sales became obvious

Multiple Data Sources needed a new approach

- Amazon recognized some data sources were more reliable than others.
- Onix was born, in 2000, to give publishers a single XML format to transmit all data—the idea was that their data was the most authoritative and would trump other sources.
- Amazon moved into other product lines and also solicited multiple item records by encouraging any merchant to sell online.

Creating The Authoritative Item

What the Vendors Provide...

Somedata Music

Title: Leaving Las Vegas - Soundtrack
Artist: Don Henley; Mike Figs;
The Palladinos;
Format: CD
UPC: 61868100552

Stellar Music

Title: Leaving Las Vegas Motion
Picture Soundtrack
Artist1: Henley, Don
Artist2: Sting,
Director: Figs, Mike
Media Type: Compact Disk
UPC: 9699857932

Krapdata Record Store

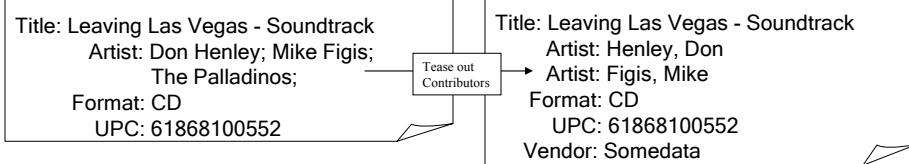
Title: LEAVING VEGAS
Artist: VA
UPC: 61868100552



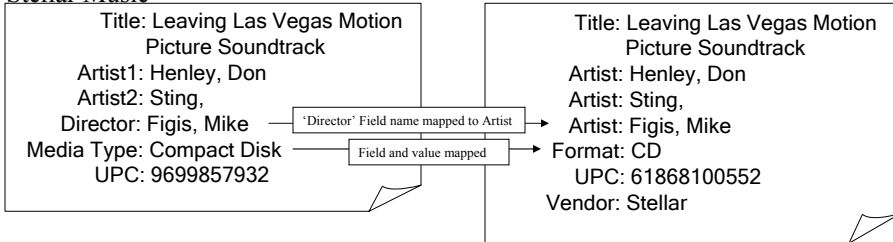
What Happens When This Process Doesn't Work

Normalization

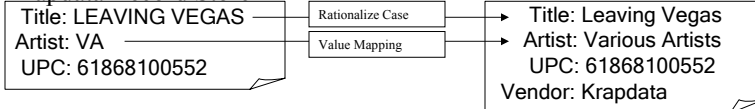
Somedata Music



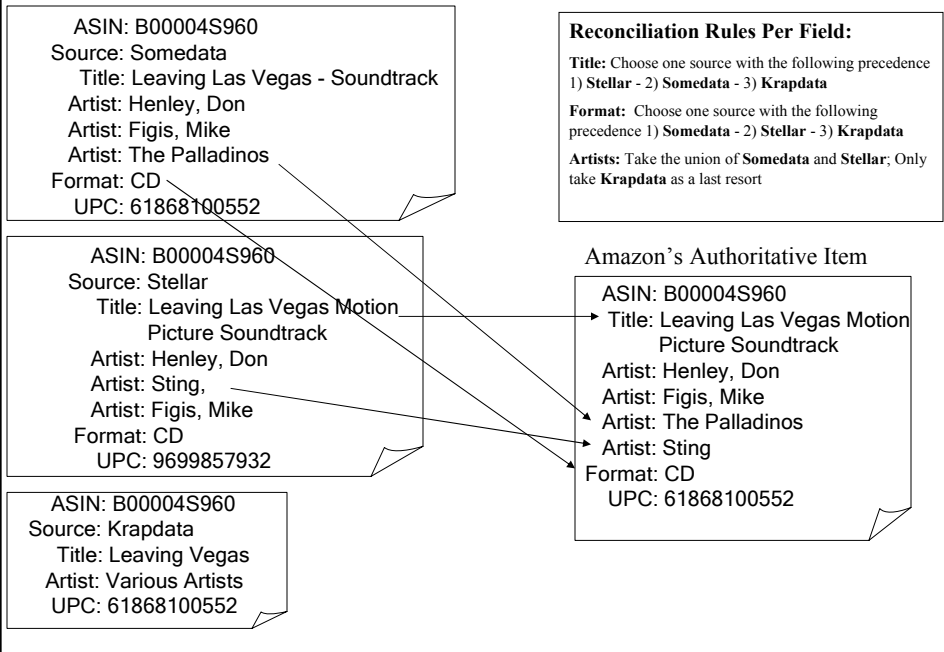
Stellar Music



Krapdata Record Store



Reconciliation – All Three Sources Exist



Before and After

Somedata Music

Title: Leaving Las Vegas - Soundtrack
Artist: Don Henley; Mike Figis;
The Palladinos;
Format: CD
UPC: 61868100552

Stellar Music

Title: Leaving Las Vegas Motion
Picture Soundtrack
Artist1: Henley, Don
Artist2: Sting,
Director: Figis, Mike
Media Type: Compact Disk
UPC: 9699857932

Krapdata Record Store

Title: LEAVING VEGAS
Artist: VA
UPC: 61868100552

Amazon's Authoritative Item

ASIN: B00004S960
Title: Leaving Las Vegas Motion
Picture Soundtrack
Artist: Henley, Don
Artist: Figis, Mike
Artist: The Palladinos
Artist: Sting
Format: CD
UPC: 61868100552

Subjectivity replaces objectivity in data source authority:

- The Internet started this trend towards creating more sources of authority, and e-commerce cataloging reflects this, too.
- Used to be listprice was a fixed value—turns out various merchants can have different views of that; the legal definition states a price that has been in effect for the past 30 days somewhere.
- Classification can be driven by customers' buying habits—"people who bought this book, also bought this.." a new way to bring related items together in less tangible ways than conventional cataloging could address.
- Categories are seen as related to marketing, not intrinsic to the item—Amazon's browse categories are now owned by buyers and merchants and can change.

Everyone can be an authority:

- Publishing houses determine subject classification system that drives how all bookstores, including online, classify their items. The BISG has a BISAC committee that determines subject headings.
- Multiple merchants can vie for being the best data contributor: Amazon can rely upon those doing the best job at data quality.
- Customers create their own pages, are rated by their peers, and can influence customer opinion and buying patterns. Customer reviews are key to product selling.

Leveraging these authority sources-- changing cataloging assumptions:

- Amazon's new catalog pipeline takes in all data but has a rating system that relies on confidence levels to determine reconciliation rules.
- Merchants will soon be able to contribute their own online corrections—but their contributions will be weighed according to their reliability as a data source.
- Customers have an online form where they, too, can contribute to data corrections; their reliability is also rated by past contributions.

A Catalog is more than simple bibliographic data:

- E-commerce shifted user's expectations, as the Internet has made a more virtual experience possible.
- People expect covers, inside graphics, tables of content, but also warranty information, color and size, accessories, required additional components, industry ratings and 3-D images.

Retail Cataloging and Product Development:

- bisg.org
- editeur.org, bic.org.uk
- doi.org, crossref.org, contentdirections.org
- Bookstream.com

DOI, the digital object identifier

- Gives the owner of the product the ability to create and control all content related to the item
- Gives multiple online sources the ability to display the item content easily and consistently
- Provides a way to link all digital content and to recombine and create new packages and new products, like chapter excerpts, edocs and packages tailored for specific uses.
- Leaves the cataloging—including establishing the unique code—in the hands of individual publishers.

Onix, the XML standard for online selling

- <Product>
 - <RecordReference>1234567890</RecordReference>
 - <NotificationType>03</NotificationType>
 - <ISBN>0816016356</ISBN>
 - <ProductForm>BB</ProductForm>
 - <DistinctiveTitle>British English, A to Zed</DistinctiveTitle>

Sample Onix file using short tags

<product>

<a001>1234567890</a001>

<a002>03</a002>

<b004>0816016356</b004>

<b012>BB</b012>

<b028>British English, A to Zed</b028>

Onix Tag Definitions

- Product: refers to a single product record
- Record Reference Number: a unique number always associated with this product; can be internal to your organization—not the ISBN!
- Notification Code: type of notification: early, advance book in hand, update, delete.
- Product Form: BB means hardback, but there are also codes for variations on book bindings, audio formats, maps, films and microfilms. \

Some Onix elements more common in library catalogs

- Edition Type—reprint, abridged, annotated, student edition, etc.
- Contributor Role—main author, illustrator, translator, etc.
- Biographical Note—free text information
- Audience Code—reader level

Some Onix codes specific to marketing

- Expected Availability Date—date on which shipping to retailers is expected to begin
- On Sale Date—date a new product can be placed on sale
- Audience Restriction Flag
- Promotion Contact Details
- Initial Print Run

So, where are we going, and what does tomorrow's catalog look like?

- As our world becomes more virtual, more and more types of content will develop to bridge the physical and online worlds.
- The ability to personalize and tailor information, as well as to contribute and enhance information, will continue to be very important, and this activity will occur on various levels and by various groups—buyers, sellers, creators, marketers.
- The glut of information sources will make cataloging and its traditional processes impossible. More and more we will rely on on-the-fly categorization and search retrieval, rather than laborious and proprietary data record construction.

Fundamental cataloging concepts, will endure:

- Data quality must continue to be good, and outsourcing this work will continue and become even more commercially viable.
- More and more types of items and objects will become cataloged and tracked—much of it will never live in our physical libraries, but will be available to customers in various ways—and library customers will continue to rely on libraries to know about and find everything.
- While multiple sources and authorities may contribute to cataloging, the ability to build and maintain reliable catalogs for the future will still need to rest with those who have a deeper understanding of the challenges and opportunities this new world has brought.