

Klantencontactdag
18 september
2008

New Directions in Cataloging at OCLC



Renee Register
Global Product Manager
OCLC Cataloging & Metadata Services

Daniël van Spanje
Global Product Manager
Metadata Services



Trends

- Meer content voor de gebruiker in de catalogus
- Verdere efficiencies in metadatamanagement

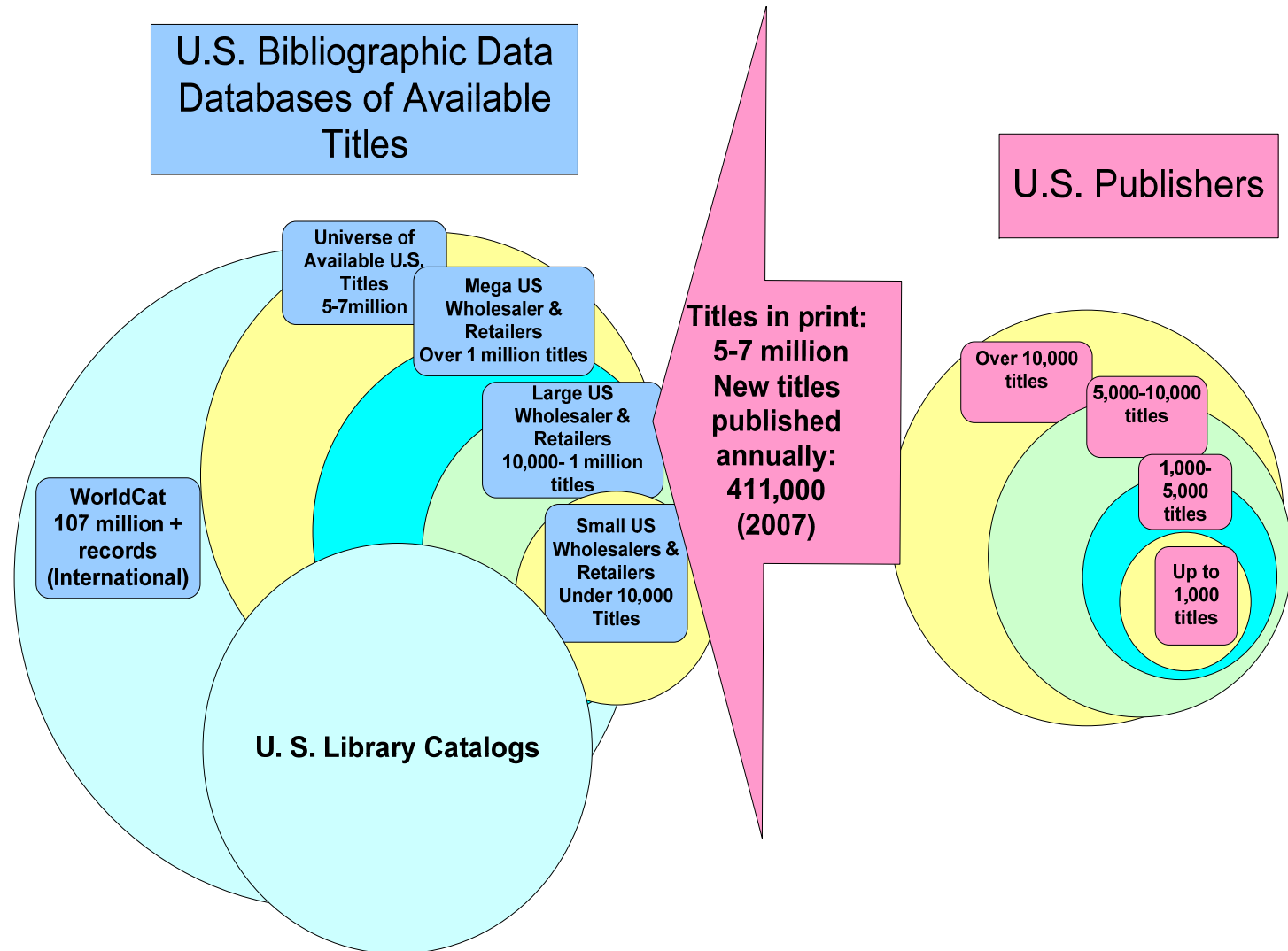
Dat betekent voor metadatamanagement:

- Meer informatie dan alleen een isbd
- Dynamische metadata records
- Naar de bron ("Upstream cataloging")

**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**

Bibliographische metadata voor titels in de VS

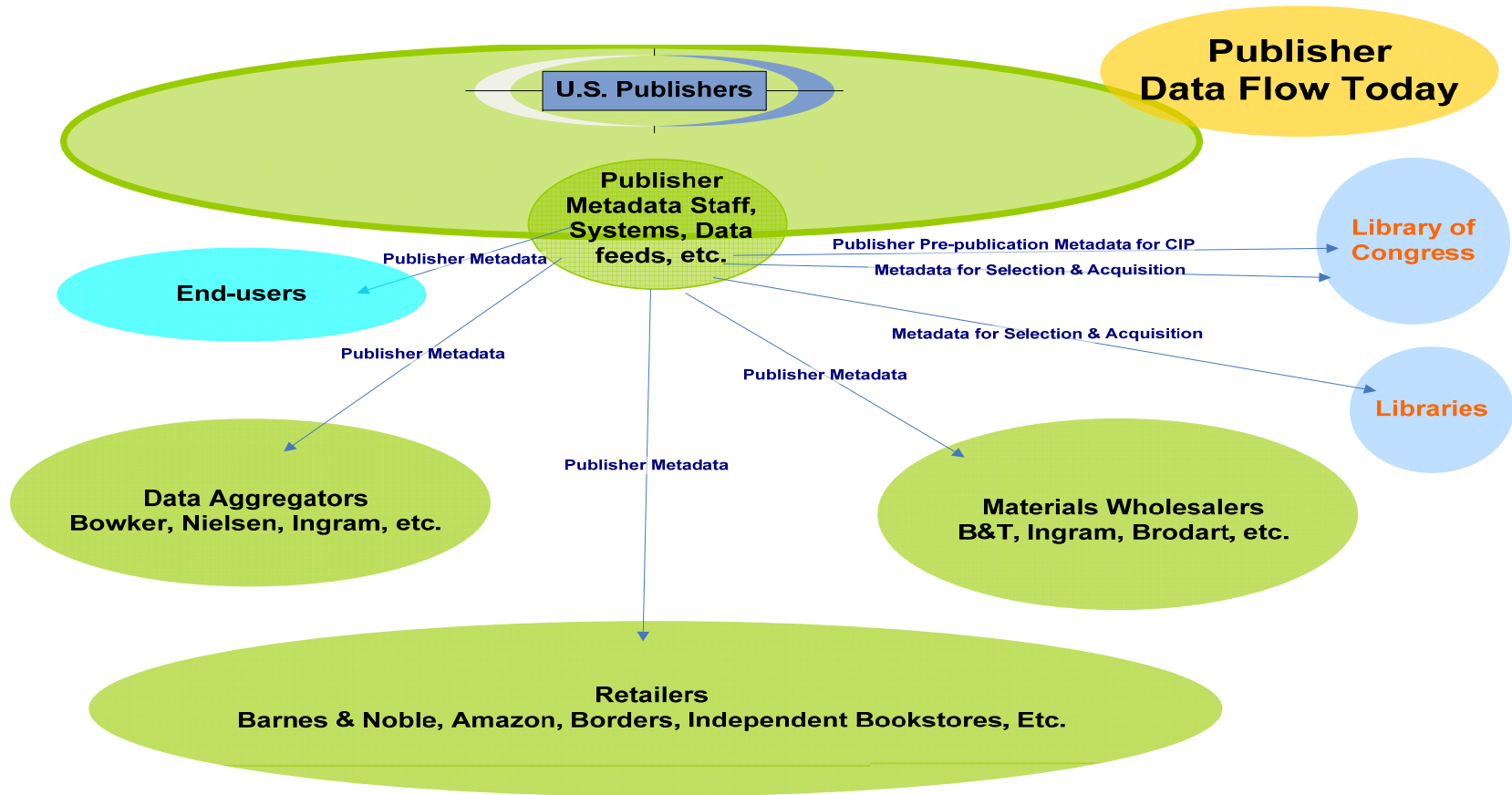


Metadata ontstaan bij de uitgever of de leverancier van de content

Die metadata worden op verschillende manieren gebruikt:

- ✓ Gedrukte catalogi voor marketing en advertenties
- ✓ Websites van uitgevers
- ✓ Voorraadssystemen van uitgevers
- ✓ Uitgevers managementinformatiesystemen
- ✓ Uitgeversgegevens voor leveranciers
- ✓ Uitgevers investeren in medewerkers en systemen om te kunnen voorzien in hun behoefte aan metadata





De metadata supply chain: leveranciers



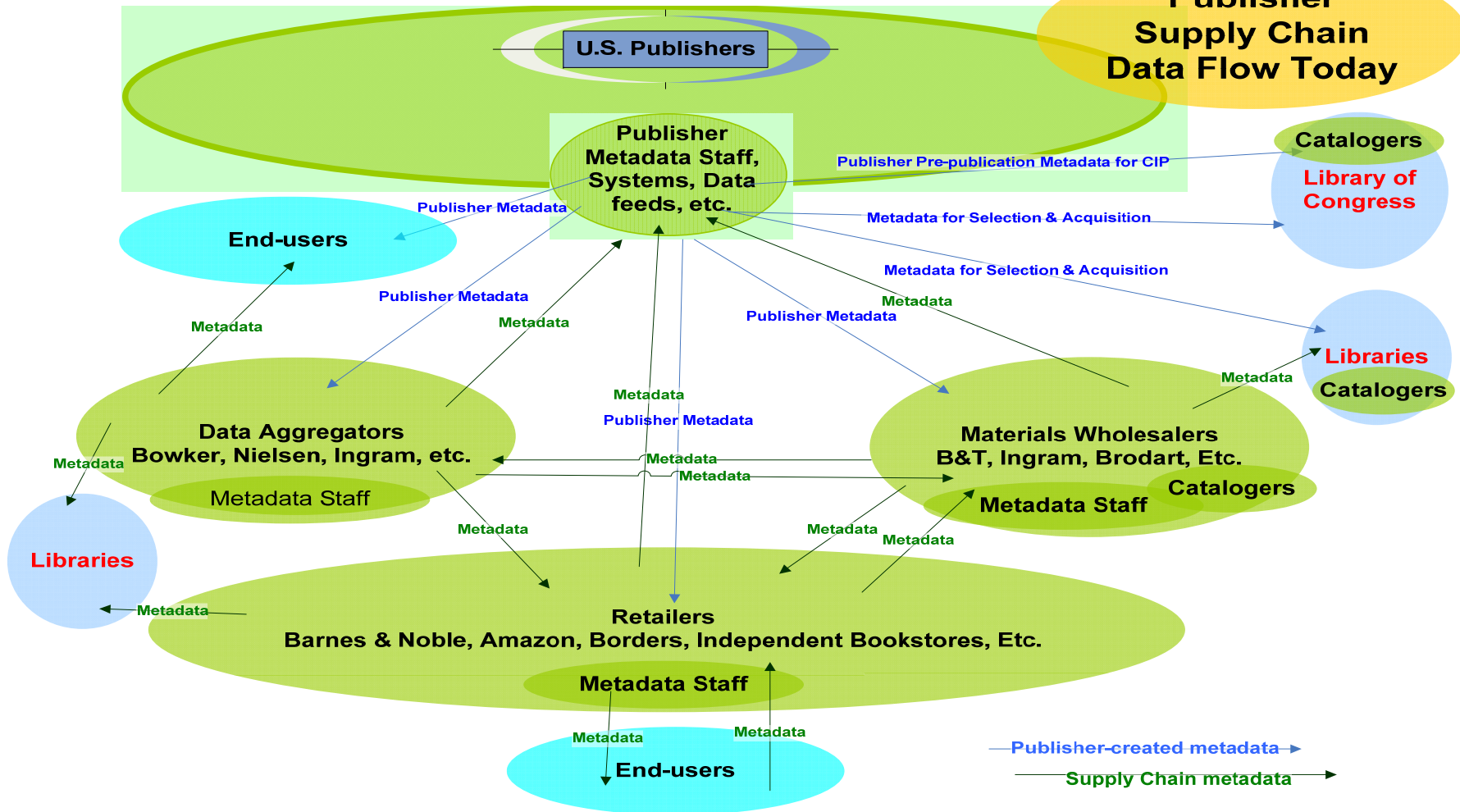
OCLC™

Niet alleen uitgevers ook andere organisaties in de uitgeversketen investeren in medewerkers en systemen voor het maken, verrijken en bewerken van metadata.

Leveranciers verzamelen metadata van verschillende uitgevers en bewerken het zo, dat ze bruikbaar worden voor boekhandels, bibliotheken en eindgebruikers.

- ✓ Doorzoekbare websites met extra features (bestelopties, selecties voor notification en approval plans)
- ✓ Extra gegevens die meegeleverd kunnen worden (ToC, recensies)
- ✓ Ook daar medewerkers die werken met metadata die van de uitgevers komen

Publisher Supply Chain Data Flow Today



Metadata in de bibliotheek



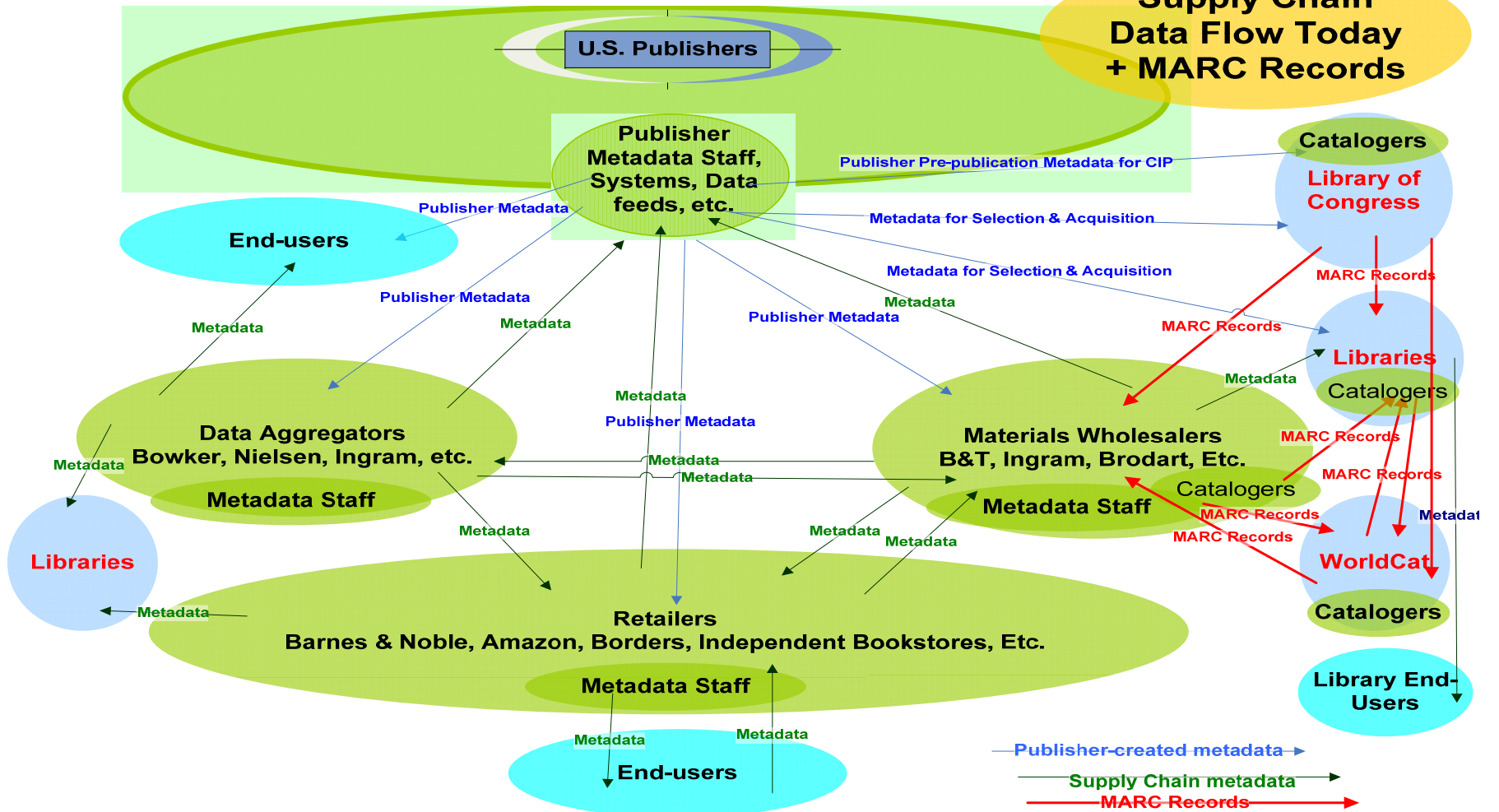
- Bibliotheken gebruiken metadata uit de uitgeversketen door het bestellen, gebruik van andere catalogi, collectieprofielen of approvalplans, e.d.
- De meeste bibliotheken willen de data in een bepaald formaat dat de uitgever niet altijd kan leveren. In de US willen bibliotheken vooral MARC
- Bibliotheken hebben daarnaast behoefte aan classificaties en andere onderwerpsgegevens die de uitgeversketen niet kan leveren.

Metadata in de bibliotheek



- De Library of Congress werkt de “uitgevers”-metadata om tot MARC records en voegt bibliotheekspecifieke gegevens toe en maakt er CIP records van
- In de bibliotheken ontlennen catalogiseerders MARC records van LC en andere bronnen en maken die MARC records verder af bij ontvangst van de bestellingen
- Leveranciers hebben vaak catalogiseerders in dienst om MARC records te maken en ander personeel om boeken “kastklaar” te kunnen leveren

Supply Chain Data Flow Today + MARC Records

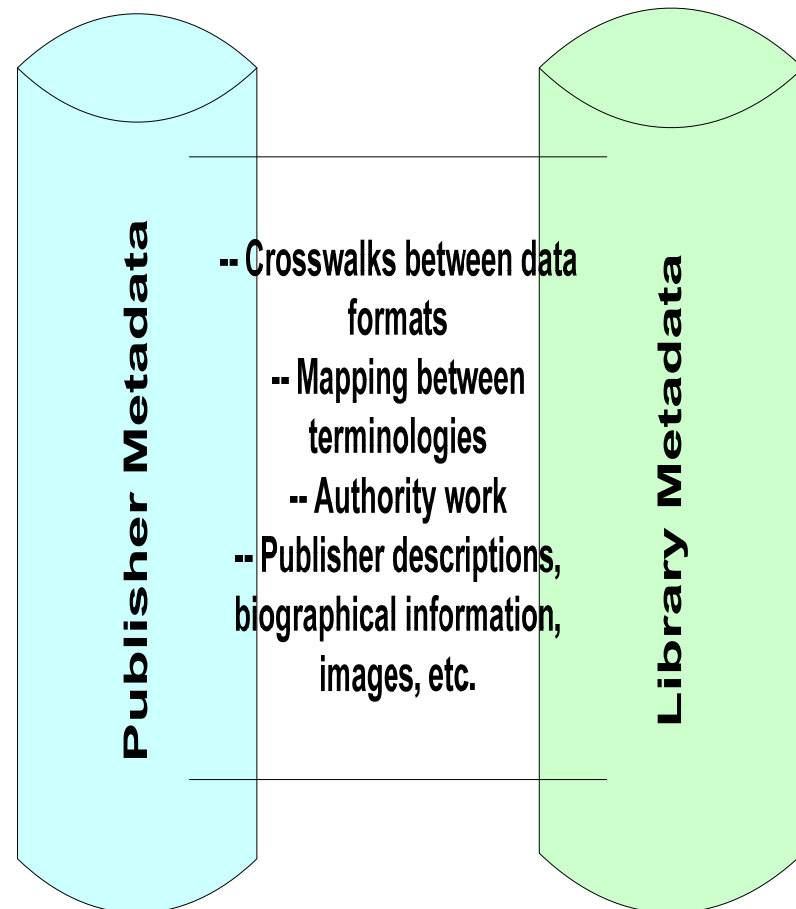


- Wij moeten toegeven dat de bestaande modellen niet in stand gehouden kunnen worden
- Wij kunnen de metadatapraktijken niet langer blijven “verzuilen”



- Re-mix en hergebruik van bestaande metadata
- Toenemen van collaboratie en cooperatie tussen bibliotheek en uitgevers supply-chain gemeenschappen
- Afbreken van scheidingen tussen metadata die je gebruikt om te bestellen en metadata die je gebruikt voor discovery, business intelligence en collectie management
- Meer betrokken raken in het “upstream metadata creation” proces, en integreren van beschikbare metadata in workprocessen upstream en accepteren dat metadata zich in de tijd ontwikkelen

- toepassingen moeten op elkaar aansluiten en gemakkelijk kunnen worden gedeeld - binnen en buiten de bibliotheekgemeenschap
- De bibliotheekgemeenschap moet zijn deskundigheid verbreden, en uitgevers en leveranciers in haar samenwerkingsverbanden betrekken



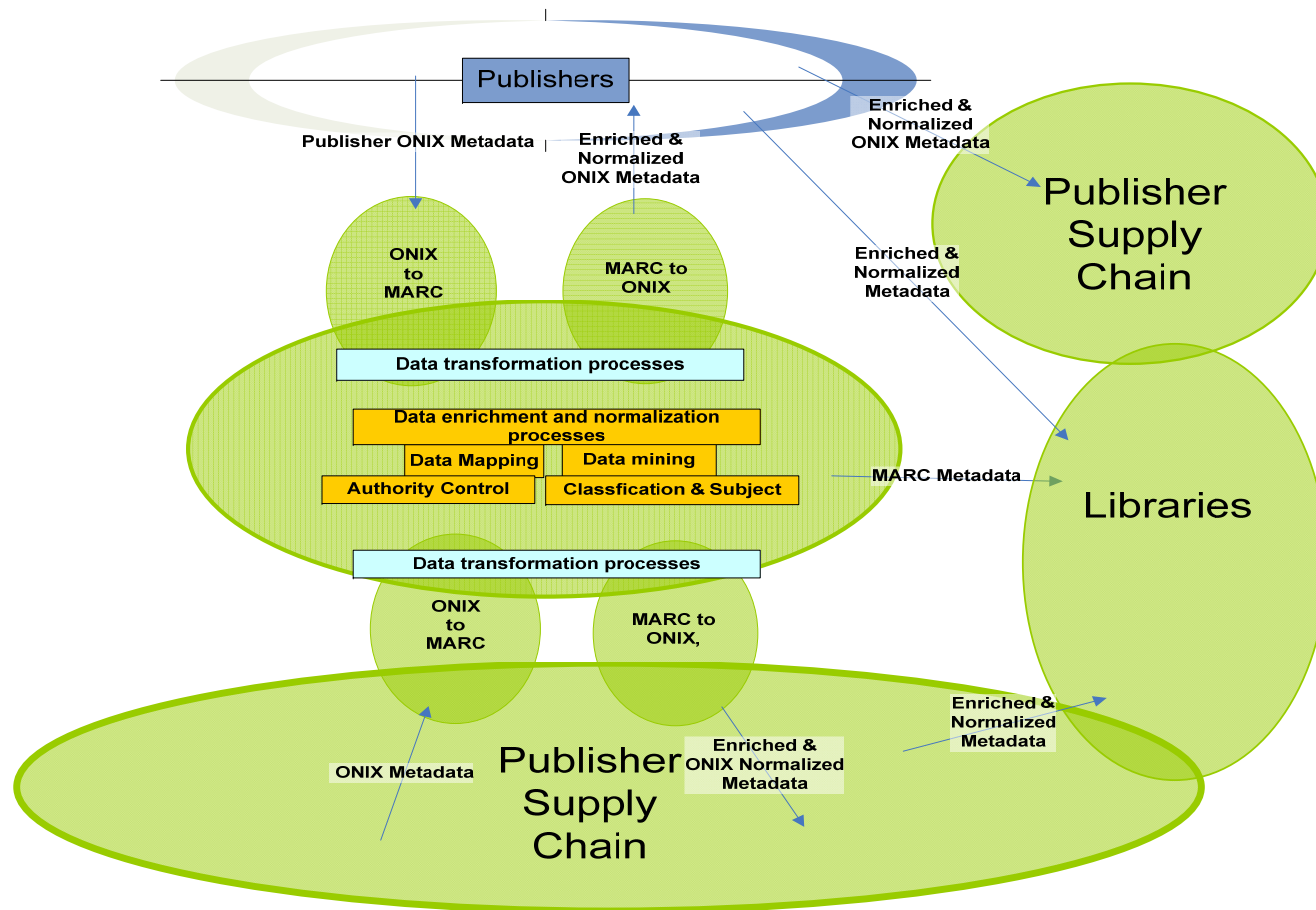
"Next Generation" Pilot



Geautomatiseerd proces voor het ophalen,
crosswalk-en en verrijken van de ONIX metadata
van de uitgevers

- Output in MARC en ONIX voor bibliotheken en uitgevers
- OCLC pilot programma met uitgevers, leveranciers en bibliotheken januari - oktober 2008

"Next Gen" Data Flow



Deelnemende bibliotheken

- ❖ Phoenix Public Library
- ❖ The Ohio State University Libraries
- ❖ Chicago Public Library
- ❖ MIT Libraries

Deelnemende uitgevers / leveranciers

- ❖ Princeton University Press
- ❖ Hachette Book Group
- ❖ Ingram Book Group
- ❖ Taylor & Francis

"Next Generation" Pilot



Adviesraad

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

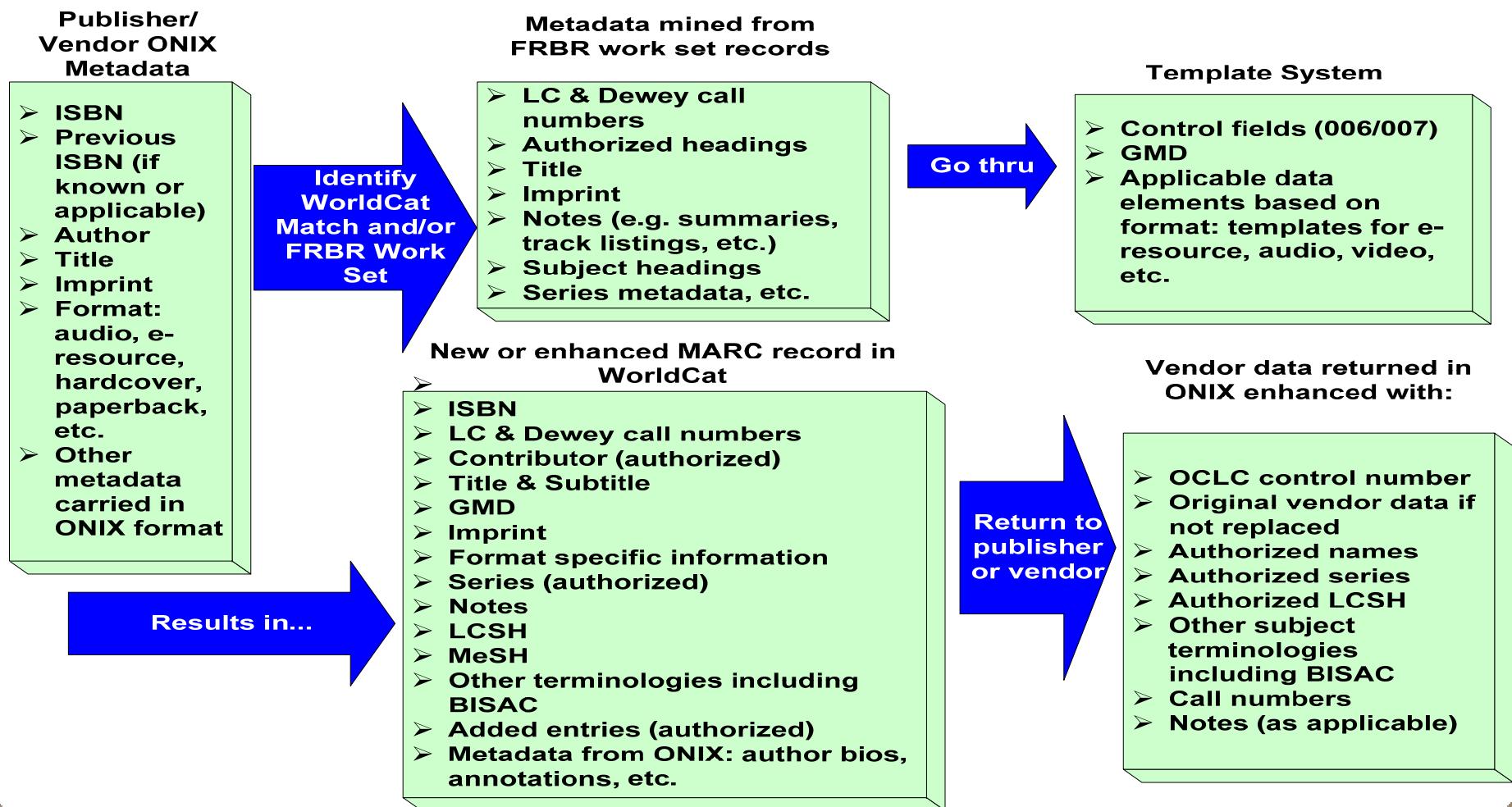
Nora Rawlinson, Publishing industry veteran and consultant to libraries on collection development

Kevin Clair, Penn State University Libraries

Richard Stark, Barnes & Noble

Marlene Harris, Chicago Public Library

"Next Generation" Catalogisieren



3 pilot partners hebben ieder ongeveer 1,000 ONIX records aangeleverd

Totaal aantal verwerkte records: 2,935

Totaal aantal records gematched in WorldCat: 2,724

Totaal aantal records zonder exacte match: 211

Totaal aantal verrijkte records: 2,706

“Next Generation” Pilot - Statistiek - Data Verrijking



Field Type	# Added or Enhanced During Enrichment	Field Type	# Added or Enhanced During Enrichment
0XX Classification		5XX General notes	
050 LC Classification	2,068	504 Bibliography	1,634
060 NLM Classification	163	505 Contents	75
082 Dewey Classification	2,139	511 Participants/Performers	93
		521 Target Audience	64
1XX Contributors		Other 5XX general notes	1,794
1XX Main Contributor	1,826		
130 Uniform Title	6	6XX Subjects	
		600 Personal Name	716
2XX Title		610 Corporate Name	198
24X Title and Subtitle	2,997	650 Topical Term	14,470
25X Edition	7	651 Geographic Name	1,030
26X Publication Info	15	Other 6XX added	1,265
3XX Physical Description	364	7XX Added Entry/Access Points	
		700 Personal Name	373
4XX/8XX Series	4,649	710 Corporate Name	373
		711 Meeting Name	7
		730 Uniform Title	118
		Other 7XX various access points	416

1 deelnemer leverde een grotere data set

Totaal aantal verwerkte records: 7,649

Totaal aantal records gematched in WorldCat: 6,584

Totaal aantal records zonder exacte match: 1,065

Totaal aantal verrijkte records: 6,565

“Next Generation” Pilot - Statistiek - Data Verrijking



Field Type	# Added or Enhanced During Enrichment	Field Type	# Added or Enhanced During Enrichment
0XX Classification		5XX General notes	
050 LC Classification	5,811	504 Bibliography	1,213
060 NLM Classification	72	505 Contents	1,047
082 Dewey Classification	6,079	511 Participants/Performers	34
		521 Target Audience	253
1XX Contributors		Other 5XX general notes	4,528
1XX Main Contributor	6,004		
130 Uniform Title	6	6XX Subjects	
		600 Personal Name	1,768
2XX Title		610 Corporate Name	520
24X Title	6,754	650 Topical Term	34,109
25X Edition	14	651 Geographic Name	4,135
26X Publication Info	15	Other 6XX added	9,868
Other 2XX added entries	6		
		7XX Added Entries/Access Points	
3XX Physical Description	692	700 Personal Name	2,806
		710 Corporate Name	430
4XX/8XX Series	8,361	711 Meeting Name	7
		730 Uniform Title	260
		Other 7XX various access points	1,886

Voorbeeld van verrijking op basis van een aanwezig record in WorldCAT (exacte match)



Major ONIX Data elements	ONIX Metadata from Pilot Partner	MARC Data Elements	Metadata added to ONIX record from WorldCat matching record
Numeric identifiers	UPC	OCLC Number	57599515
		Publisher Number	B0003007-02 Decca
		LC Classification (050)	SDB 10822
Title	#1 Puccini Album	Title (245)	The #1 Puccini album
		Alternate Title (246)	Number one Puccini album
		Alternate Title (246)	Puccini album
Contributor	Various Artists	Main Entry (100)	Puccini, Giacomo,
		Birth & Death Dates	1858-1924
Publication date	20040810		
Imprint name	Decca		
Physical description	Various ONIX codes	Physical description	2 sound discs (142 minutes) – digital – 4 ¾ in.
		Performers (511)	Roberto Alagna, Carlo Bergonzi, Jussi Bjorling, etc.
		Topical Subjects (650)	Operas -- Excerpts
Price	\$17.98		

Voorbeeld van een nieuw record met verrijking op basis van de FRBR werk-set (Geen hit in WorldCat)



ONIX data elements	ONIX Metadata from Pilot Partner	MARC data elements	Metadata mined from FRBR work set and ONIX metadata combined to create new record
Numeric identifiers	ISBN 0898594707		(Map ONIX data)
		OCLC Number	New OCLC #
		LC Classification (050)	HM132
		DDC (082)	301.11
Title	Personal Relationships	Title	Personal relationships
Subtitle	Their Structures and Processes	Subtitle	their structures and processes
Contributor	Kelly, Harold H.	Main Entry (100)	Kelly, Harold H.
Publication date	1982		(Map ONIX data)
Imprint name	Lawrence Erlbaum		(Map ONIX data)
Physical description	Various ONIX codes that describe the item would drive fixed field info, etc.	Physical description	(Map ONIX data to drive material type, etc.)
Pagination	192		(Map ONIX data)
		Series (440)	John M. MacEachran memoria
BISAC Subject Heading	PSY031000	Topical Subjects (650)	Interpersonal relationships
		Non-English Subjects	Psicologia social Relations humaines. Interpersoonlijke relaties.
Price	\$39.95		(Map ONIX data)

Voorbeeld van verrijking op basis van ONIX data, WorldCat record en de FRBR werk-set



Major ONIX Data elements	ONIX Metadata from Pilot Partner	MARC Data Elements	Metadata added to ONIX record from WorldCat matching record and FRBR work set
Numeric identifiers ISBN	0892962844	OCLC Number	16867603
		LC Classification (050)	PS3515.U585
		DDC (082)	813.54
Title	McBain's Ladies	Title (245)	McBain's ladies
		Subtitle (245b)	the women of the 87th Precinct
Contributor	McBain, Ed	Main Entry (100)	McBain, Ed
		Birth and Death Dates	1926-2005
Publication date	1988		
Imprint name	Grand Central Publishing		
Pagination	320	Physical description	320 p. ; 24 cm.
BISAC	FIC000000	Topical Subjects (650)	87th Precinct (Imaginary place) -- Fiction
	(Fiction – General)		Police -- Fiction
			Women – Fiction
			Police spouses -- Fiction
			City and town life – United States -- Fiction
			Policewomen -- Fiction

“Next Generation” Pilot - Evaluatie door de bibliotheken van de 1e dataset



Feedback on specific data elements:

Data Element	Present	Correct	Missing key info	Format Accuracy	Staff Attention Needed Consistently Before Usable
Title and subtitle	XXXXX	XXX	X		XX
Alternate title	XXX	XX	X	XXX	
Author and Contributor data	XXXX	X	XXX	X	
LC Classification	XXXX	XX		XX	X
Dewey Classification	XX		X	XX	X
NLM Classification					
Imprint and publication data	XXXXX	XX	XXX	XX	XX
Format specific data	XXX	XX	X	X	
Series data	XXX	X	X	X	
Audience data	XX	X		X	
Library subject headings:					
LCSH	XXXXX	XX		X	X
MeSH					
SEARS					
BISAC subject headings	XXX		X	XX	X
Physical dimensions	XXX	X	XX		X
Language(s) of item	XXXX	XX		X	
General notes	XXXX	X		X	X
Summary notes	XXXX	X		X	X
Contents notes	XXX	X		X	
Subject information	XX	X			
Key words or internal subjects	XX	X			

“Next Generation” Pilot - Evaluatie door de bibliotheken van de 1e data set



Rate the value of the enhanced records for the following library processes:

Library processes	Ranking by responding libraries
Selection and acquisition	100%
Cataloging and physical processing	90%
End-user retrieval and circulation	85%
Overall rating by libraries of enhanced records:	64%
% of Libraries comfortable loading all of the records received into ILS for immediate use	48%
If no, estimated time needed on average per record fby staff to make a record "loadable"?	
<15-mins	80%
15-30 mins	20%

"Next Generation" Pilot: acties



Pilot Wrap-up - Mei - Oktober 2008

- Doorgaan met het leveren van metadata voor evaluatie door deelnemers
- Verder verfijnen van software algoritme's, m.n. voor het maken van records
- Statistische analyses afmaken en samenvatten van de resultaten van de evaluaties met de deelnemers
- Afzonderlijke case studies van deelnemers afmaken
- Resultaten doorgeven aan de deelnemers en de adviesraad
- Resultaten doorgeven aan bibliotheken, uitgevers en leveranciers
- In productie gaan: najaar 2008

Bedankt