



RLG Shared Print Update

Dennis Massie
Program Officer

ALA Midwinter Meetings
Monday, 18 January 2010
OCLC Red Suite

Infrastructure and Implementation

- **Print Archiving & Network Disclosure: MARC 583**
 - Maximize visibility of title-level preservation data
 - Use cases for collection managers
 - Integration in distributed cataloging workflows
- **Decision Tree for De-duplication of Print Journals**
 - Context-appropriate approach to managing redundancy
 - Workflows adapted to different institutional settings
 - Maximize incentives for participation in shared print archiving
- **Toward a 'Cloud' Library**
 - Implementation framework for increasing reliance on shared print & digital repositories, maximizing operational efficiencies
 - Phased approach to rationalization of local print collection
 - Joint effort with HathiTrust, NYU, ReCAP and CLIR

1) MARC 583 for Print Archiving

- Absence of shared infrastructure for disclosing print preservation commitments - a *critical impediment* to achieving 'scale' in distributed print archiving efforts
- MARC 583 proposed as vehicle for sharing preservation data for monographic literature, ca. 2007.
- Now: extend to serials

Goals

- ✓ *Test feasibility of batch updating in local system and WorldCat*
- ✓ *Sample use cases for integration in collection management workflows*

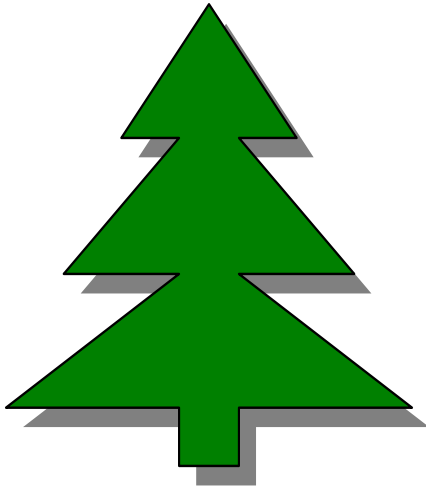
MARC 583 for Print Archiving (cont.)

- Initially explored use of Action note in bibliographic 'master' record
- Proposal reviewed by >125 serials catalogers, preservation officers, collection managers
- Currently exploring use of Action note in local holdings record, CONSER's preferred approach
- Testing against titles in Journals Preservation project

Who's involved:

- ✓ UCLA: John Riemer, Valerie Bross, Jake Nadal
- ✓ Penn State: Christopher Walker
- ✓ NYU: Everett Allgood

2) Deaccessioning Decision Tree



Deaccessioning Print Journals Group

- Concentrated on dual format
- Documented obstacles
- Prioritized data elements
- Noted missing pieces

Assignment from RLG Programs Council

- Create deaccessioning decision tree

Beginning Questions:

- Organizing Principle?
- Level of granularity?
- How long before it's too late?

DAP-J Working Group

✓ ARL

Columbia University

Bob Wolven, Jeff Carroll

Indiana University

Carolyn Walters

New York University

Angela Carreno

University of Arizona

Steve Bosch

University of Michigan

Bryan Skib

✓ Liberal Arts College

Swarthmore College

Amy McColl

✓ Museum

Brooklyn Museum

Deirdre Lawrence

Frick Collection

Debbie Kempe

Metropolitan Museum

Ken Soehner

Museum of Modern Art

Milan Hughston

✓ Special Library

U of Pennsylvania Law

Merle Slyhoff

✓ Legal Depository

Trinity College Dublin

Margaret Flood

Deaccessioning Decision Tree

- Organizing principles
 - Mission of library
 - Risk tolerance

- Types of Research Libraries owning print journals
 - Assume preserving print not part of mission; decisions based on local need
 - Due to risk aversion, policy, politics or personalities, no print will be discarded, aside from weeding out-of-scope items
 - Willing to discard duplicates and selected print journal back files available in e-format
 - Open to discarding titles beyond duplicates; seek to do so sensibly and collaboratively
 - See preserving scholarly record as essential to mission; committed to keeping what print they have

Deaccessioning Decision Tree

- Ithaka paper “What to Withdraw”
 - 4 exemplary scenarios
 - Repeatedly decries lack of centralized information
- Back to first principles
 - Key obstacle: already in storage
 - Storage as a de facto archive?
 - Knowing what is stored currently impossible
 - Is there a moderate-effort way to approximate a snapshot of what’s held in storage across the entire network?
 - Why not assign items in storage a different WC symbol?
 - With that as a first step, what could be done on top of that?

3) 'Toward a Cloud Library'

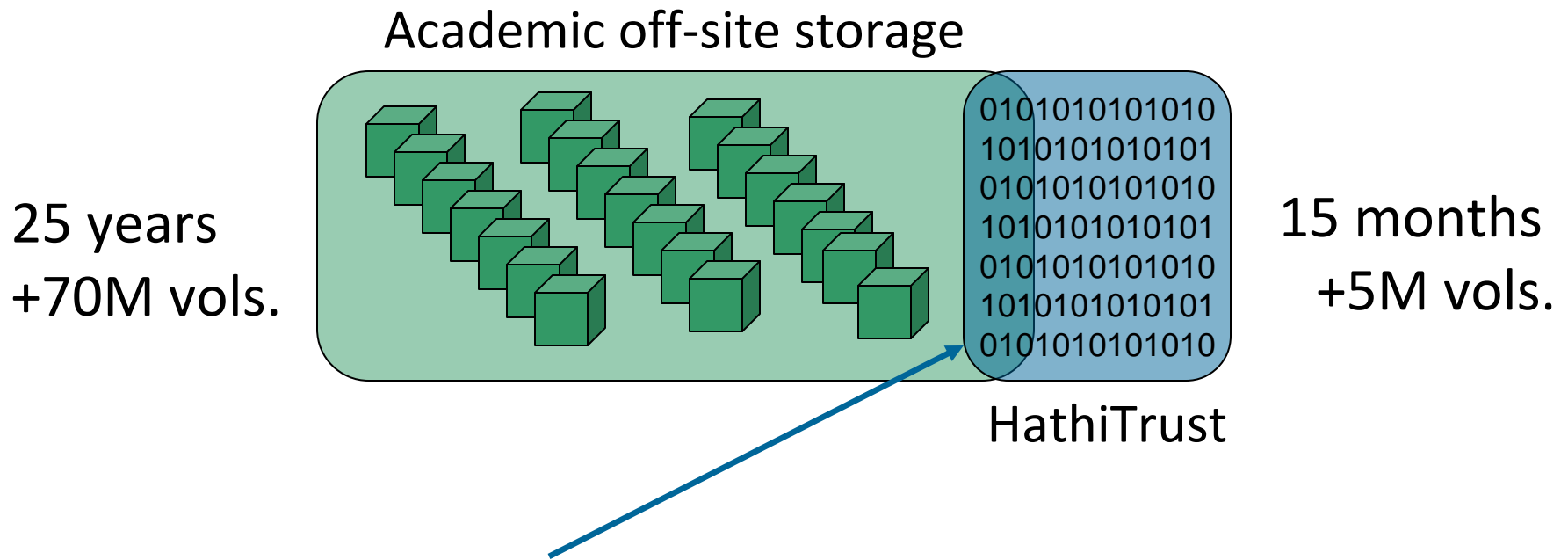
Objective: Characterize the *near-term opportunity for externalizing management of academic research collections*, leveraging capacity of *large-scale shared print and digital repositories*

Outcomes: *opportunity and risk assessment* based on aggregate collection analysis; *draft service agreement* enabling generic consumer library to selectively outsource preservation and access of low-use research collections to large-scale print and digital repositories

Who's involved: NYU, Hathi Trust, ReCAP libraries
from OCLC Research: Constance, Roy, Shailen
with support from: CLIR, Mellon Foundation

Timeline: July - December 2009
draft report anticipated Jan/Feb '10

Shared Infrastructure: Books & Bits



Will this intersection create new operational efficiencies?

For which libraries?

Under what conditions?

How soon and with what impact?

Value of partnership increases as number of participants grows

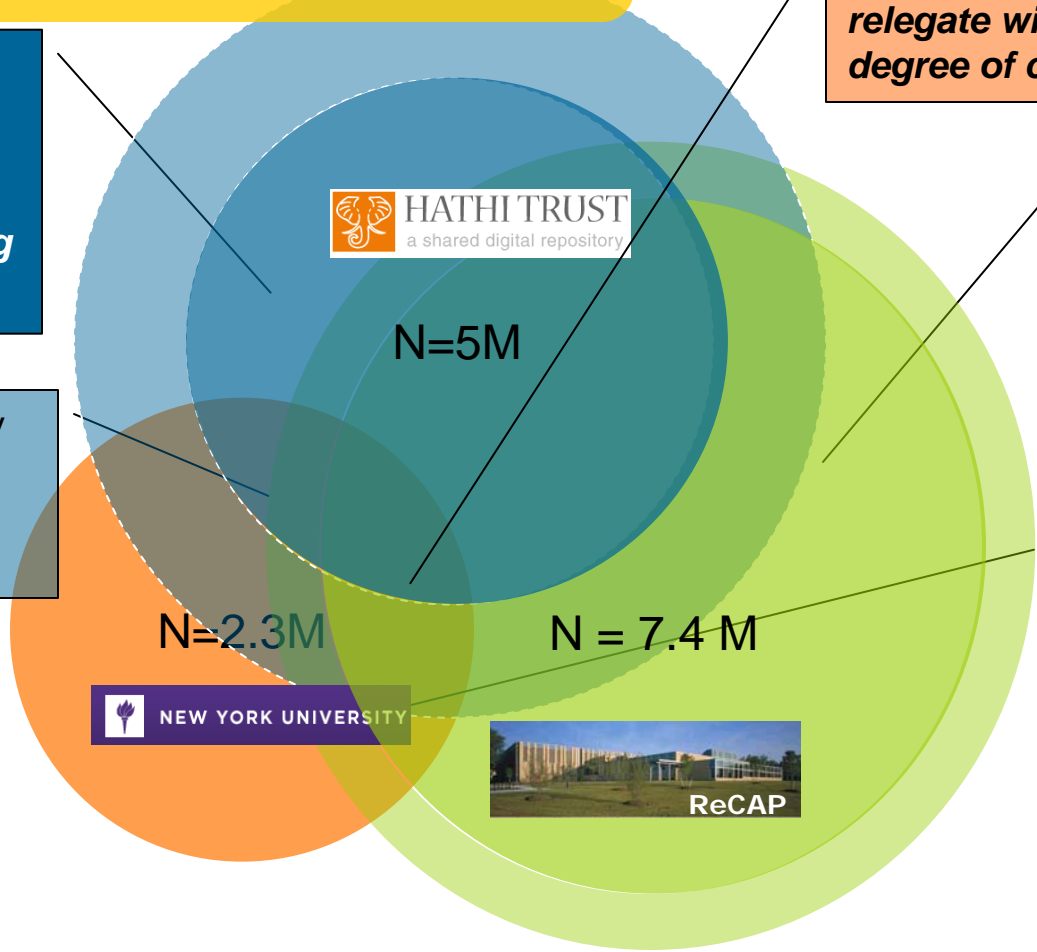
Material that NYU can obtain through HT dependent on copyright status – means of **enhancing 'local' collection**

Material that NYU may choose to **relegate based on copyright/availability**

Material that NYU can **relegate with a high degree of confidence**

Material that NYU can already source through existing ILL – **enhance local collection**

Material that NYU may choose to **relegate with appropriate service level agreement**



Intersections

- Opportunities for Institutional Cooperation*
- Shared Policy Frameworks*
- Joint Service Agreements*
- Increased Operational Efficiencies*

Methodology

Data processing and analysis (Roy, Constance)

Harvest Hathi metadata

... Enhance

... Map to WorldCat bibliographic records

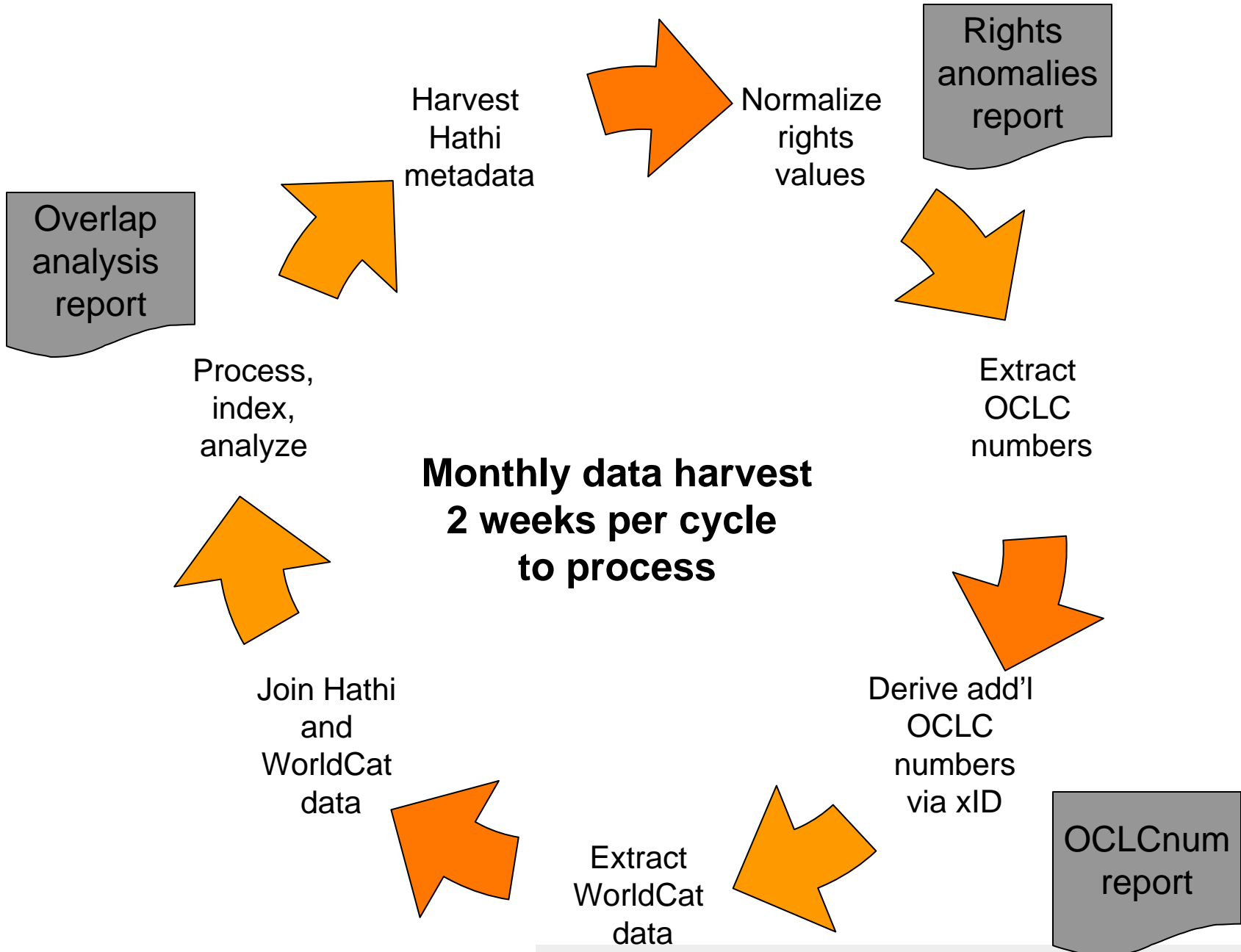
... Merge WorldCat, Hathi, ReCAP (sample) data

... Analyze *rinse, repeat* = 17M records

Interpretation (OCLC Research, NYU, Hathi, ReCAP)

Facilitated dialogue: service expectations - Sept

Collaborative authoring: service agreement - Dec



What's in the Cloud?

Based on analysis of titles in Hathi archive:

- 2.8 million digitized *books* (97%); 97K serials (3%)
- *Humanities* constitute >50% of collection
- ~350K titles (12%) in the *public domain*

- 2 million titles archived in Hathi (70%) are *also* held in print form by at least ~~one~~ ~~large-scale~~ *shared print* repository
- 1.4 million archived titles (48%) are held by *fewer than 25 libraries*

Implications for Collection Management

- If a *guarantee of digital preservation alone* were sufficient to justify de-duplication of print books, academic libraries in North America could achieve a *20-40% reduction* in local inventory in the near term
- If a *supplemental guarantee of preservation in a shared print repository* is required (secure digital copy + secure print copy), a *5-15% reduction* is possible
- Levels of duplication in *'unsecured' inventory* may influence adoption of either strategy

Impact on Library Operations

- Selective reduction in low-use print inventory enables *reallocation of library resources* toward more distinctive service profile
- Storage transfer and weeding decisions informed by *system-wide view* of preservation infrastructure
- Significant *long-term cost avoidance*: print book inventory drives 95% of ARL lifecycle expenditures
- Shared service agreements *limit institutional exposure to risk*

Medium	Discounted Life Cycle Cost (per unit)	Total Life Cycle Cost (per unit)	Purchase Cost (per unit)	Total Cost / Purchase Cost (per unit)
Monographs	\$ 119.56	\$ 343.03	\$ 47.78	718%
Current serials	\$ 634.91	801.87	590.97	134
Microforms	\$ 0.27	0.45	0.11	256
CD-ROMs				
Magnetic tapes				
Magnetic disks				
Computer files				
Video & Film	\$ 128.95	107.50	15.70	307
Computer files	\$ 0.17	0.07	0.01	331

**“monographs are overwhelmingly the largest source or driver of library costs . . .
If research libraries want to control their costs, they must work to control and reduce the life cycle costs of maintaining their monograph collections” S. Lawrence et al (2001)**

NYU and Hathi Collections

As of December 2009,

- NYU: 2.3 million titles in WorldCat

600K titles (28%) duplicated in Hathi

or 38,000 linear feet of shelf space

~36,000 (6%) in the public domain

- What is NYU's risk tolerance for weeding redundant holdings?
- Which subject areas and imprint ranges are off limits?

NYU and ReCAP Partners

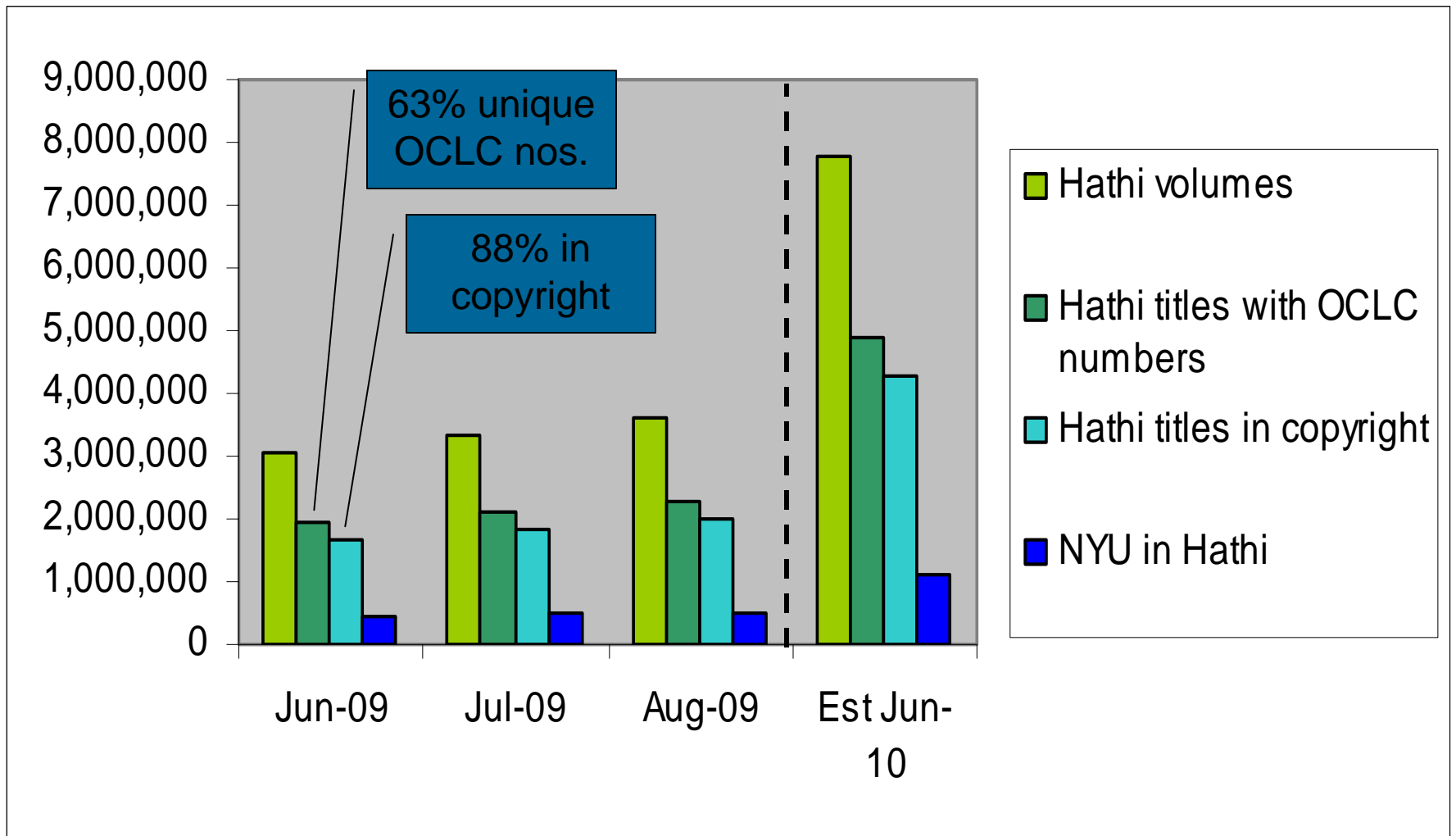
As of December 2009,

- NYU: 2.3 million titles in WorldCat
1.45 million titles (63%) duplicated in aggregate ReCAP partner collections +200K (10%) duplicated by ALL ReCAP libraries and Hathi

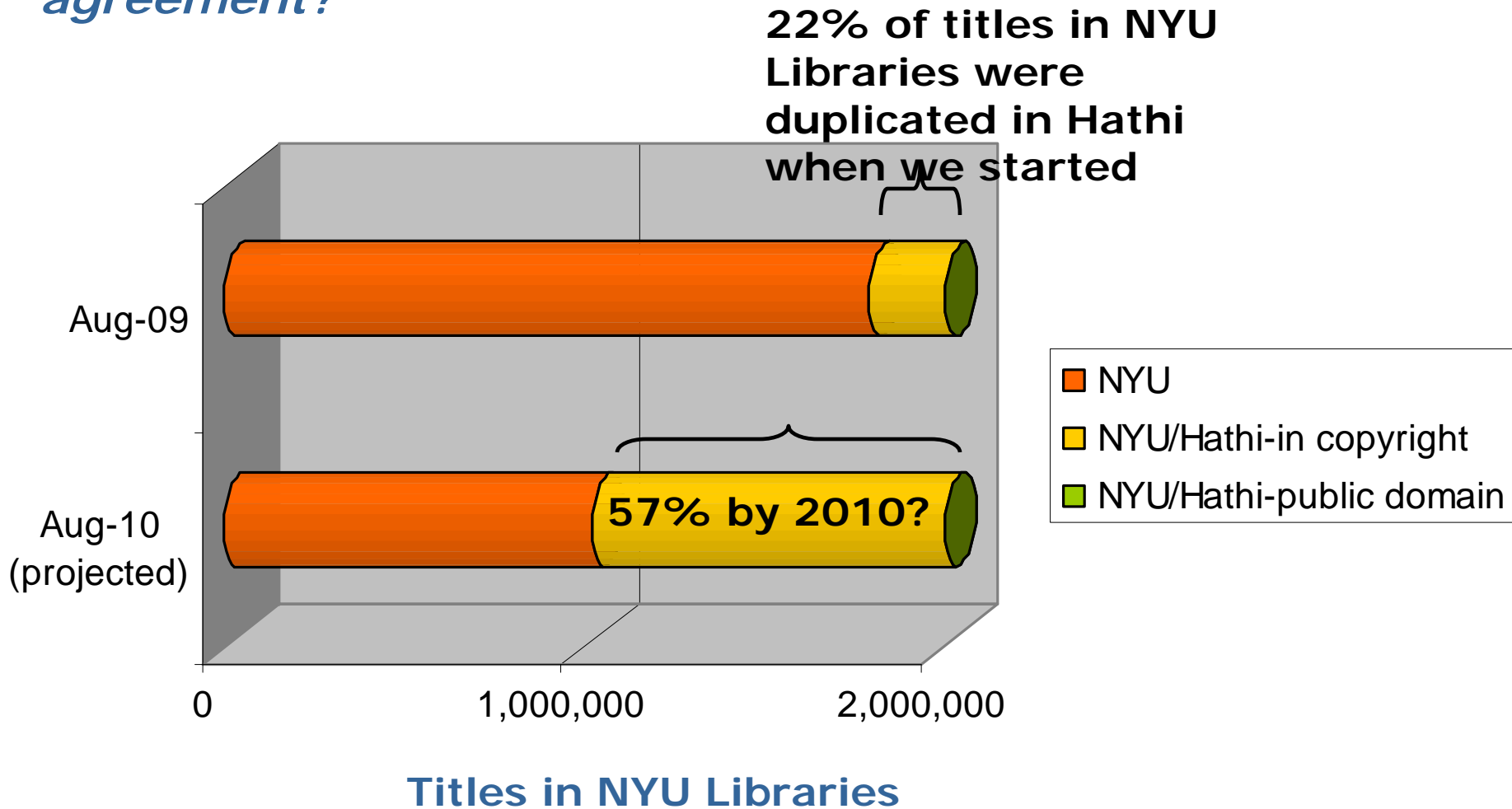
- How many of these titles are in ReCAP facility?
- How many are unrestricted?
- How many are already in NYU storage?

Compared with Columbia snapshot of 2.1 million titles, 160K are in all 3 (7%)

Hathi Growth Trajectory: Volumes, Titles, Coverage



How much coverage is 'enough' to warrant shared print/digital service agreement?



Food for thought . . .

Institution	Title overlap with HathiTrust	As % of holdings in WorldCat	Titles in the public domain
University of Pennsylvania	647,431	20%	50,823
University of Arizona	511,614	17%	30,539
Swarthmore College	129,661	25%	14,503
UC Southern Regional Library Facility	524,013	21%	50,692
CRL	82,651	6%	8,704

Questions, Comments?

massied@oclc.org