Appendices

Appendix A.   **ReCAP Planning and Policy Committee Members**

Appendix B.   **ReCAP Planning Report**

Appendix C.   **ReCAP Holdings Analysis Report**

Appendix D.   **Proposed Revisions to ReCAP Consortium Agreement**

Appendix E.   **ReCAP Shared Collection Plan**

Appendix F.   **ReCAP Cost Estimates**

Appendix G.   **Business models in other shared print programs**

Appendix H.   **Workflow Analysis report**

Appendix I.   **ReCAP Project Technology Report**

Appendix J.   **ReCAP High-Level Architecture 1.0**

Appendix K.   **Discovery to Delivery System Cost Projections**

Appendix A
ReCAP Planning and Policy Committee Members

* Ann Thornton, Andrew W. Mellon Director of the New York Public Libraries

* Jane Aboyoun, Chief Technology Officer, NYPL

* Irene David, Director, Technology Initiatives, NYPL

* Judith Johnson, Director of Enterprise Applications, NYPL

* David Magier, Associate University Librarian for Collection Development, Princeton University Library

* Marvin Bielawski, Deputy University Librarian, Princeton University Library

* Patty Gaspari Bridges, Assistant University Librarian, Collection Development, Princeton University Library

* Richard Schulz, Associate University Librarian for Technical Services, Princeton University Library

* Robert Wolven, Associate University Librarian for Bibliographic Services and Collection Development, Columbia University Libraries

* Robert Cartolano, Director, Libraries Information Technology Office, Columbia University Libraries

* Eileen Henthorne, Executive Director, ReCAP

* Lizanne Payne, Planning Consultant

Appendix B.   ReCAP Planning Report

ReCAP Discovery to Delivery Project
Goals and Priorities

Lizanne Payne
July 13, 2012

This is a report required under planning activity 1d "summarizing goals, issues, and directions. The report will describe agreed priorities and directions, and will include major concerns or constraints reported by individual partners."

Executive Summary

Summary of Issues

Several planning issues were raised during the initial site visits:

1. There is a need to clarify the meanings of "ReCAP" and "shared collection".
2. There is an interdependency between shared collection policies and identification of materials for the shared collection.
3. There is no common resource-sharing consortium among ReCAP partners
4. The controversy over NYPL's Central Library Plan may affect planned transfers to ReCAP.
5. A strict NYPL deaccession policy may complicate deduping.
6. Construction at ReCAP will affect timing of Discovery to Delivery implementation.
7. The role of the Technology Architect needs to be defined and filled.

Summary of Major Decisions

The ReCAP Planning and Policy Committee made the following decisions on May 18, 2012.

The shared ReCAP collection has been defined as those materials housed at the ReCAP facility (or transferred there in the future) that meet certain selection criteria and are placed under a retention agreement or joint ownership agreement. Neither the criteria for selection nor the retention agreement (or joint ownership agreement) have been defined as of July 2012; these will be the subject of further discussions over the next few weeks and months.

The discovery/delivery system will be designed to provide search, display, and request functions for the shared collection at ReCAP. There is no requirement to provide a discovery system across all ReCAP partner general collections or across non-shared items at ReCAP.

Report Details

One of my first deliverables as ReCAP Planning Consultant is to "prepare a report summarizing goals, issues, and directions. The report will describe agreed priorities and directions, and will include major concerns or constraints reported by individual partners." (Activity 1d)

The goals of the ReCAP Discovery to Delivery Project, as stated in the proposal to the Andrew W. Mellon Foundation, are as follows:

1) Reduce the cost-per-title of managing (and expanding) the shared collection;
2) Expand the breadth of material available (without Interlibrary Loan) to the users of the ReCAP partners;
3) Optimize and integrate the discovery experience in regards to the holdings of the ReCAP partnership;
4) Reduce time to delivery by streamlining request and fulfillment mechanisms; and
5) Position the ReCAP partnership to serve the needs of additional institutions.

Planning Consultant Activities (Lizanne Payne)

Between April 16 through 26, 2012, I conducted site visits at each ReCAP partner library and at the ReCAP facility, where I met with key staff to discuss project goals and local priorities or issues.

After those onsite meetings, I conducted a joint meeting of the ReCAP Planning and Policy Committee and many key staff from ReCAP partners, held at the New York Public Library on May 18, 2012. The goal of that meeting was to:

- synthesize the results of the individual meetings
- establish consensus about priorities
- identify areas of difference
- propose directions for future activities.

The report which follows incorporates the results of the original site visits and the May 18 joint meeting.

Summary of Related Consultant Activities

During this initial planning period, the other project consultants were also gathering information and performing preliminary analyses, summarized below:

1. Collection Analysis (OCLC Research)

OCLC Research performed a preliminary analysis during the pre-grant phase (February-March 2012), as stipulated by The Andrew W. Mellon Foundation, based on extracting a new set of ReCAP partner

holdings from OCLC WorldCAT, and comparing additional records identifying holdings in ReCAP and items to exclude from sharing. A Holdings Analysis Subcommittee was identified, consisting of Bob Wolven and Zach Lane (Columbia), Denise Hibay, Daphna Blatt, and Jason Baumann (NYPL), and David Magier and Patricia Gaspari-Bridges (Princeton). This group requested that OCLC perform a second round of analysis to show holdings counts by serials vs monographs, campus holdings vs ReCAP holdings, holdings in Hathi Trust (all) and in Hathi Trust public domain. This second analysis is still in process as of mid-July 2012.

2. Technology Consultant (Marshall Breeding)

Marshall Breeding conducted site visits between April 23 – 30, 2012. Key points identified during those visits were:
- There is no common ILS or discovery system among the partners
- There is no ReCAP catalog
- ReCAP uses a batch-mode GFA inventory control system with no APIs for interoperability or online status query
- Planning for new systems is underway at each partner institution, major systems are in transition

3. Workflow Consultant (Sustainable Collections Services)

SCS conducted site visits between May 8 – 14, 2012, to document current workflows for materials into and out of ReCAP and to explore what de-duplication workflows (both pre-facility and post-facility) would require.

Planning Issues

Several important issues arose during the planning discussions and with the related consultants that could affect the planning project and/or a future implementation project.

1. There is a need to clarify the meanings of "ReCAP" and "shared collection".

    It became clear that different people meant different things by the terms "ReCAP" and "shared collection". The proposal language stating a goal to "Optimize and integrate the discovery experience in regards to the holdings of the ReCAP partnership" was interpreted differently by various people.

    - What is ReCAP? Is it the shared facility in Princeton? The consortium of partner libraries? The consortium only in the context of the facility?
    - What is the "shared collection"? Is it only items held at the ReCAP facility? General collections at partner libraries? A "shared print" archival agreement? A resource-sharing agreement?

2. There is an interdependency between shared collection policies and selection of materials for the shared collection.

Several people mentioned that it would be important to know what the retention and access policies are before declaring certain materials as eligible for the shared collection. Conversely, the characteristics of shared materials could influence the policies defined for the shared collection. According to the proposed schedule, the content (identification of materials) for the shared collection should be completed by July 2012 while the definition of policies is part of the policy planning effort to be completed by December 2012 (although discussed in detail throughout the planning project).

3. There is no common resource-sharing consortium among ReCAP partners

Ideally it would be preferable to support lending from the shared ReCAP collection using an existing resource-sharing mechanism across the three partners. However, there is no common resource-sharing agreement in place. Columbia and Princeton belong to BorrowDirect, but NYPL does not. Columbia and NYPL belong to MARLI, but Princeton does not. It does not seem likely that these parties would join one of the existing agreements and systems.

4. The controversy over NYPL's Central Library Plan may affect planned transfers to ReCAP.

While NYPL had made a preliminary identification of the type and quantity of materials to transfer to ReCAP in the near future, it seems likely that these planning targets may change in the wake of the public controversy over the Central Library Plan. NYPL may also face increased scrutiny of any sharing or retention agreements that may come about through the Discovery to Delivery project.

5. A strict NYPL deaccession policy may complicate deduping.

NYPL is bound by a policy that requires review and approval at a detailed level of any deaccessioning decisions. This could limit NYPL's ability to deduplicate its collections in reliance on other ReCAP partner holdings, and could create pressures to use the NYPL copy as the shared copy of any duplicated holdings.

6. Construction at ReCAP will affect timing of Discovery to Delivery implementation.

Construction of two additional storage modules is underway at the ReCAP facility in Princeton. The timing of this construction – or, more importantly, the availability of these new modules – will affect the timing of significant aspects of the planned Discovery to Delivery systems. ReCAP partners will need an efficient mechanism to identify and prevent transferring duplicates into ReCAP by summer/fall 2013, but implementation of a new ReCAP discovery system is not likely to occur before January 2014 at the earliest, and more likely later.

Details of the timing issues are:

- Complete current planning grant: March 2013
- Complete ReCAP modules 8 & 9: June 2013
- Begin load (ReCAP ingest with significant pent-up demand): summer/fall 2013: *therefore need interim discovery tool to avoid transferring duplicates by summer/fall 2013*
- Acquire implementation funding:
  - Target for proposal: March 2013 or Sept 2013
  - Target for funding available: Sept 2013 or January 2014
- Go live with ReCAP discovery system: earliest between January – June 2014

7. The role of the Technology Architect needs to be defined and filled.

As of July 2012, the Technology Architect (included in the project proposal) had not yet been hired. Discussions among project partners indicated different understanding of the role of the Technology Architect compared to the Technology Consultant (Marshall Breeding) and lack of consensus about the level of detail that is necessary during the planning phase. The Planning and Policy Committee determined that the Technology Architect work could be postponed until after completion of the technology environmental scan, and that the project could consider engaging Marshall Breeding to perform that role as well, while continuing to recruit for the position externally.

Scope of the Shared Collection

An important question facing ReCAP planners in spring 2012 was to determine the scope of the "shared collection". Note that a number of other planning decisions are pending and will be made during the course of the planning project.

Questions about the meaning and scope of the "shared collection" were posed and answered by the ReCAP Planning and Policy Committee during the meeting on May 18.

1. Does "Shared ReCAP Collection" = designated materials that are physically housed at the ReCAP facility (present and future) and covered by a ReCAP ownership/retention agreement? **YES.**

2. Is "discovery" required for
    - The Shared ReCAP collection as defined above? **YES.**
    - Union catalog of all partner collections? **NO.**

3. How will improved access to partner general collections be administered:
    - Through a ReCAP partners resource-sharing program? **NO.**
    - Through another existing program e.g. Borrow Direct? **MAYBE. Not a goal of this project.**

4. Will cooperative collection development (future acquisitions) of <u>general collections material</u> be defined through this project? **NO. Not a goal of this project.**

Questions were also raised about the importance of deduplicating holdings in various contexts:

1. Should future transfers to ReCAP be deduplicated (i.e. no new duplicates in ReCAP)? **YES.**

Items in the shared ReCAP collection should not be duplicated by future transfers to ReCAP. Further discussion will be necessary to determine the duplication policy with respect to any <u>non-shared</u> ReCAP items.

2. Should current ReCAP holdings be deduplicated? **NO, EXCEPT POSSIBLY FOR CONTIGUOUS JOURNALS.**

Deduplication of volumes in a high-density Harvard-model facility is a very labor-intensive and expensive process (need to retrieve, compare, reshelve, modify metadata and barcodes, dispose, and consolidate trays). It may be worth considering only for bound journals where a significant portion of a given run is shelved contiguously and therefore could release significant amounts of shelf space.

Summary

The shared ReCAP collection has been defined as those materials housed at the ReCAP facility (or transferred there in the future) that meet certain selection criteria and are placed under a retention agreement or joint ownership agreement. Neither the criteria for selection nor the retention agreement (or joint ownership agreement) have been defined as of July 2012; these will be the subject of further discussions over the next few weeks and months.

The discovery/delivery system will be designed to provide search, display, and request functions for the shared collection at ReCAP. There is no requirement to provide a discovery system across all ReCAP partner general collections or across non-shared items at ReCAP.

Appendix C.   ReCAP Holdings Analysis Report

ReCAP Holdings Analysis Report

The ReCAP partner libraries engaged OCLC Research to analyze partner collections and current holdings in the ReCAP facility during 2012. This collection analysis provided a number of reports that supported subsequent decisions about the shared collections.

Collection Analysis Methodology

OCLC Research provided three sets of reports including two major rounds of new analysis for the ReCAP shared collections project:

- Review of earlier data from the Cloud Library project
- "Pre-grant analysis" in March 2012, designed to be completed before the current planning project began in April 2012
- Detailed analysis using updated data in July-October 2012 (see Attachment 1: ReCAP Partner Shareable Holdings Summary).

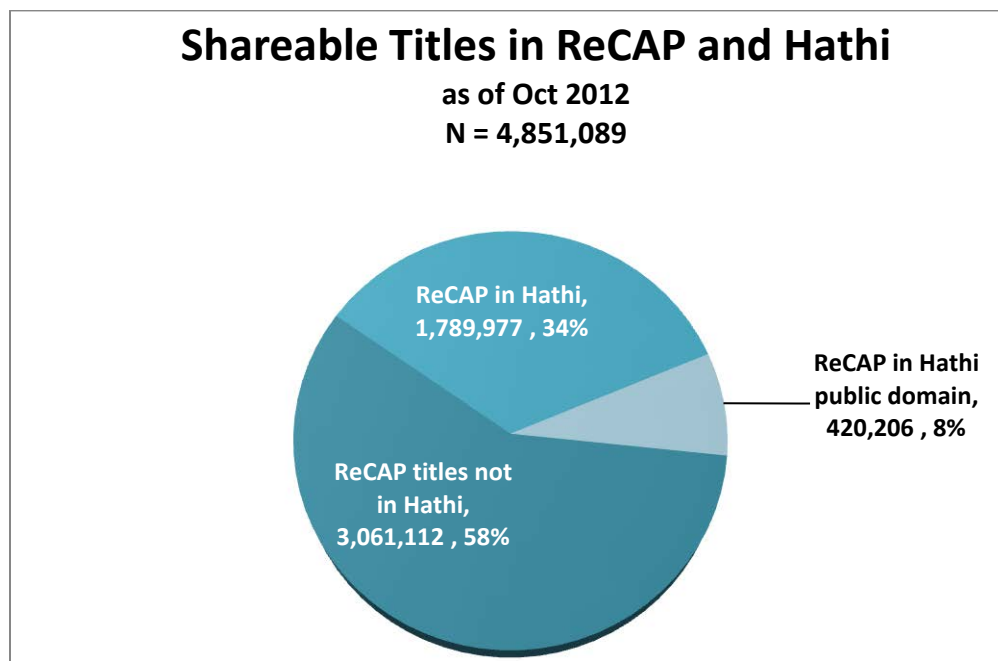In the pre-grant analysis (March 2012), OCLC Research compared the following data sets:
- ReCAP holdings (title records) provided by each partner
- ReCAP holdings (titles, OCLC numbers) to exclude from consideration as shared collection (e.g. rare books)
- NYPL and Columbia items (title records) planned for near-term transfer to ReCAP
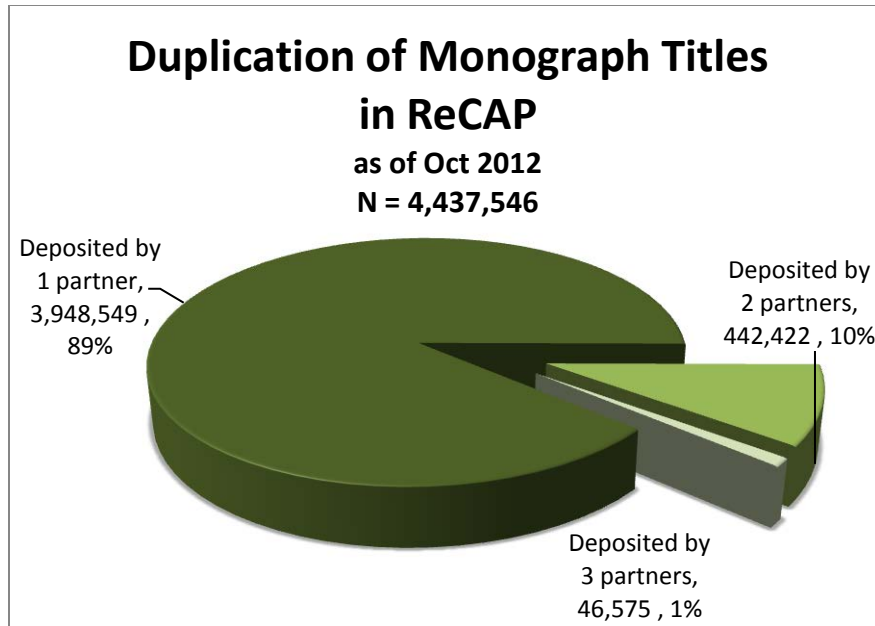- Hathi Trust holdings
- WorldCat holdings

In the updated analysis (July – October 2012), OCLC Research generated new comparisons using updated ReCAP holdings (title records) provided by NYPL and Princeton. These new record sets were needed because data problems in some of the files originally provided in March 2012 adversely affected the holdings comparisons.

OCLC Research provided a variety of comparisons between ReCAP partner collections, ReCAP holdings, Hathi Trust holdings, and WorldCat holdings. Attachment 2 lists the different analyses that were performed (and files provided) in support of the ReCAP planning project. It would be very useful and not very difficult to create a separate database from the spreadsheet files provided by OCLC Research, to support further analysis of this snapshot of ReCAP holdings by subject, rights status, and other attributes.

Major collection overlap characteristics

ReCAP partners provided information on more than 4.8 million titles (OCLC records) representing items currently held in ReCAP that are potentially shareable.

All Shareable Titles in ReCAP
as of Oct 2012
N=4,851,089

Princeton
1,308,080
27%

Columbia
2,001,434
41%

NYPL
1,541,575
32%

Shareable Titles in ReCAP and Hathi
as of Oct 2012
N = 4,851,089

ReCAP in Hathi,
1,789,977 , 34%

ReCAP in Hathi
public domain,
420,206 , 8%

ReCAP titles not
in Hathi,
3,061,112 , 58%

Duplication of Monograph Titles in ReCAP
as of Oct 2012
N = 4,437,546

Deposited by 1 partner, 3,948,549 , 89%

Deposited by 2 partners, 442,422 , 10%

Deposited by 3 partners, 46,575 , 1%

Note: To get a sense of the quantity of <u>volumes</u> corresponding to these titles, a chart developed by Sustainable Collections Services (SCS) shows ReCAP holdings (items) for the potentially shareable customer codes:

Customer Code	Customer Name	Total items	Retrieved during the last fiscal year
CJ	Journalism Library (Columbia)	0	0
CU	Columbia Standard (Columbia)	2,576,788	45,196
EV	East Asian Vernacular (Columbia)	336,581	7,919
GC	Government Documents (Columbia)	33,422	264
HS	Health Science Library (Columbia)	48,447	700
JC	JSTOR Standard (Columbia)	0	0
SW	Social Work Library (Columbia)	0	0
PA	Unrestricted (Princeton)	1,807,683	19,479
QK	Mendel Sound & Video (Princeton)	32,093	219
GP	Government Documents (Princeton)	18,183	273 (?)
JP	JSTOR Standard (Princeton)	0	0
NA	NYPL Standard	2,677,484	27,736
	Totals	**7,512,498**	**102,665**

Observations regarding ReCAP holdings

Upon reviewing results of the holdings analysis, Constance Malpas of OCLC Research made the following observations with respect to partner holdings in ReCAP. (These observations originally were made with respect to the March 2012 data and have been updated with data from summer/fall 2012.)

- There is *significant overlap (42%) between current ReCAP storage inventory and HathiTrust*. By effectively surfacing access to digital surrogates -- including search-only access for in copyright volumes -- ReCAP partners may be able to reduce retrievals and associated operating costs.

- Based on data supplied, *title-level duplication in current ReCAP inventory is calculated at 11%.* Rather than de-duplicating this stock, ReCAP might instead view it as a *seed-collection for a shared print offering beyond the ReCAP partnership*. Especially in the context of Hathi shared print plans, it seems that this additional buffer of duplication may have business value. Practically speaking, duplication in the print storage collection provides both valuable redundancy from a preservation perspective and constitutes a more reliable source of supply. This is not to say that additional duplication should be built into prospective transfers; unless the 'business value' of the redundancy is made real, every additional copy in ReCAP will effectively represent an opportunity cost. It's possible that *different strategies will be needed to manage retrospective redundancy in ReCAP and prospective development of the shared collections.*

- A *substantial portion of the near-term transfers (~20%) are already present in the ReCAP inventory.* This suggests that without additional coordination in selection protocols, duplication in ReCAP inventory will continue to grow.

- An even greater part *(~45%) of the near-term transfers are duplicated in HathiTrust*. While relatively few of these titles are in the public domain, the aggregate library holdings per digitized title are relatively high, which may suggest that ReCAP partners can apply stricter non-duplication rules for storage transfers, maximizing the value of current investment in HathiTrust.

- The relatively high 'average library holdings per digitized title' in ReCAP (ranging from 131 for all digitized titles in ReCAP to 177 for digitized titles deposited by all three libraries) suggests a *potentially broad market for service, should ReCAP opt to make the shared collection available to other libraries.*

- Overall, the high rates of duplication with HathiTrust in extant storage inventory and near-term transfers suggests *that ReCAP partner libraries will benefit from factoring the additional redundancy of HathiTrust into shared print management plans*.

These totals are based on OCLC Research Library Partner reports and additional analysis prepared by OCLC Research in July - October 2012 using data on potentially shareable holdings provided by ReCAP partners.

Please note that these data were derived at different times and thus are not precisely comparable, but should be sufficient for general analysis.

	Data as of	Title Records (OCLC numbers)			
		Columbia	NYPL	Princeton	Total
Total Partner Collections					
Partner collections	Jan 2012	4,959,575	7,264,296	4,495,371	16,719,242
Titles in Hathi	Jan 2012	1,669,459	1,903,931	1,554,404	5,127,794
Titles in Hathi public domain	Jan 2012	246,861	376,918	316,554	940,333
Monographs					
Partner monographs	Sep 2012	4,677,013	6,338,896	4,204,591	15,220,500
Monos in Hathi	Jul 2012	1,625,868	1,822,665	1514966	4,963,499
Monos in Hathi public domain	Jul 2012	254,135	381,381	332,604	968,120
Mono triplicates	Sep 2012	1,124,643	1,124,643	1,124,643	1,124,643
Mono triplicates in Hathi	Jul 2012	634,835	634,835	634,835	634,835
Mono triplicates in Hathi public domain	Jul 2012	71,240	71,240	71,240	71,240
Serial Titles					
Partner serials	Sep 2012	146,874	395,955	170,387	
Serials in Hathi	Jul 2012	47,401	77,603	60,950	
Serial triplicates	Sep 2012	26,595	26,595	26,595	26,595
Serial triplicates in Hathi	Jul 2012	18,501	18,501	18,501	18,501
ReCAP Titles (Shareable)					
ReCAP titles	Oct 2012	2,001,434	1,541,575	1,308,080	4,851,089
ReCAP in Hathi	Sep 2012	748,375	512,697	528,905	1,789,977
ReCAP in Hathi public domain	Jul 2012	147,095	96,132	176,979	420,206
Monograph titles in ReCAP	Oct 2012	1,866,851	1,349,041	1,221,654	4,437,546
ReCAP monos in Hathi	Jul 2012	718,992	470,238	505,873	1,695,103
ReCAP monos in Hathi public domain	Jul 2012	140,096	86,169	167,131	393,396
Serial Titles in ReCAP		62,177	132,432	51,287	245,896
ReCAP serials in Hathi	Sep 2012	20,332	29,980	22,173	72,485

	Data as of	Title Records (OCLC numbers)			
		Columbia	NYPL	Princeton	Total
ReCAP Duplicate Titles		(All, not by owners, see owner details below)			
ReCAP Duplicate Titles	Sep 2012				471,593
ReCAP duplicates in Hathi	Sep 2012				265,483
ReCAP duplicates in Hathi public domain	Sep 2012				53,097
	Sep 2012				
ReCAP mono duplicates	Sep 2012				442,422
ReCAP mono duplicates in Hathi	Sep 2012				251,656
ReCAP mono duplicates in Hathi public domain	Sep 2012				48,940
ReCAP serial duplicates	Sep 2012				18,958
ReCAP serial duplicates in Hathi	Sep 2012				11,004
ReCAP serial duplicates in Hathi public domain	Sep 2012				4,135
ReCAP Triplicate Titles					
ReCAP triplicates	Oct 2012	48,880	48,880	48,880	48,880
ReCAP triplicates in Hathi	Sep 2012	32,370	32,370	32,370	32,370
ReCAP triplicates/Hathi/public domain	Sep 2012	6,480	6,480	6,480	6,480
ReCAP mono triplicates	Oct 2012	46,575	46,575	46,575	46,575
ReCAP mono triplicates in Hathi	Sep 2012	30,858	30,858	30,858	30,858
ReCAP mono triplicates in Hathi public domain	Sep 2012	5,933	5,933	5,933	5,933
ReCAP serials triplicates	Oct 2012	2,254	2,254	2,254	2,254
ReCAP serials triplicates in Hathi	Sep 2012	1,475	1,475	1,475	1,475
ReCAP Duplicate Monographs by own					
Columbia/Princeton duplicate monos	Oct-12	189,321	n/a	189,321	189,321
ReCAP duplicates in Hathi	Oct-12	96,409	n/a	96,409	96,409
ReCAP duplicates in Hathi public domain	Oct-12	26,400	n/a	26,400	26,400
Columbia/NYPL duplicate monos	Oct-12	174,177	174,177	n/a	174,177
ReCAP duplicates in Hathi	Oct-12	66,868	66,868	n/a	66,868
ReCAP duplicates in Hathi public domain	Oct-12	2,608	2,608	n/a	2,608
NYPL/Princeton duplicate monos	Oct-12	n/a	100,172	100,172	100,172
ReCAP duplicates in Hathi	Oct-12	n/a	15,930	15,930	15,930
ReCAP duplicates in Hathi public domain	Oct-12	n/a	1,609	1,609	1,609

ReCAP Holdings Analysis and Data Sets Provided by OCLC Research

These data sets were created through collection analysis by OCLC Research and were provided by OCLC Research. The individual data sets are posted on the ReCAP wiki or in a shared DropBox account as identified below.

Analysis	*File Name*	*Data As Of Date*	*Location*
Partner Library Collections compared to Hathi Trust (Reports to OCLC Research Library Partners)	Columbia University Report.xlsx New York Public Library Report.xlsx Princeton University Report.xlsx	January 2012	ReCAP Project Wiki
Monographs in Partner Collections and also in Hathi Trust public domain	testNYPpdBKSJul2012 testPULpdBKSJul2012 testPULEApdBKSJul2012 testZCUpdBKSJul2012 testZCHpdBKSJul2012 testZCLpdBKSJul2012 (these are labeled "test" because of the format; the data is complete)	July 2012	DropBox ReCAP data
Monographs held by all (triplicates) in Hathi public domain	PULZCUNYPpdBksincommonJul2012)	July 2012	DropBox ReCAP data
Serials held by all (triplicates) in Hathi public domain	testPULZCUNYPHathiTrustSERincommonJul2012	July 2012	DropBox ReCAP data
Partner Library holdings in ReCAP and also in Hathi	Columbia_recap_HT_Jul2012.xlsx NYPL_recapHTJul2012.xlsx Princeton_recap20120815_HTJul2012	July/August 2012	ReCAP Project Wiki
Duplicates in ReCAP	2ReCAPOCNSep2012WCdetailscomplete	September 2012	ReCAP Project Wiki
Duplicates in ReCAP and Hathi	2copies inReCAPSep2012 and HT Jul 2012	July/Sep 2012	ReCAP Project Wiki
Triplicates in ReCAP	3ReCAPOCNSep2012WCdetailscomplete	Sep 2012	ReCAP Project Wiki
Triplicates in ReCAP and Hathi	3copies inReCAPSep2012 and HT Jul 2012		ReCAP Project Wiki