

# Partnership for Shared Book Collections

## Infrastructure Working Group Annual Report

### September 2021

---

## Purpose

The Infrastructure Working Group of the Partnership was established to develop an understanding of Shared Print program data needs, to explore the existing infrastructure capabilities, and to define recommendations and advocate for improvements. To facilitate those aims, the following Task Forces have been created: Resource Sharing, Unique Materials, and Discovery / Metadata.

## Membership

Linda Wobbe (SCELC) – Chair  
David Almodovar (Pace Univ.)  
Judy Dobry (CDL)  
Maggie Dull (Univ. of Rochester)  
Melissa Gasparotto (NYPL)  
Tracy Gilmore (CSU, Long Beach)  
Maggie Halterman-Dess (University of Iowa)  
Shari Laster (Arizona State University)  
Jennifer Martin (Salisbury Univ.)  
Tom Teper (Univ. of Illinois)  
Amy Wood (CRL)  
Sara Amato (Partnership)

## Process and Reports

Each Task Force investigated developed use cases and vendor questions prior to holding a series of meetings with developers and vendors. Membership, vendor questions, and findings are summarized in the annual reports of each group:

- [Resource Sharing and Controlled Digital Lending Report](#) (page 3)
- [Discovery and Metadata](#) (page 7)
- [Unique Materials and Metadata](#) (page 11)

## Conclusions and Recommendations

We were interested to learn about the many exciting options for accomplishing the infrastructure needs of the Partnership. While other vendors and options have yet to be explored, our reports include some specific recommendations for further work. In most instances development and testing will be required to make progress. We hope that the Partnership OC/EC will set a framework for deciding which options recommended in the reports to pursue, perhaps through test pilot projects, and provide direction for the future work of the Infrastructure group.

# Partnership for Shared Book Collections

## Resource Sharing Task Force Report

### September 2021

---

## Purpose

The purpose of the Resource Sharing Task Force is to investigate the resource sharing infrastructure that can be used to support resource sharing within and among Partnership programs.

## Context

This report focuses on essential services or functions supported by various tools, but the Task Force acknowledges that resource sharing is a service that is continually responding to changing needs and expectations of global researchers, and continually adapting to newly developed tools and workflow designed to meet these needs. These global researchers, the infrastructure that supports them, and each library's relationship with its technological services provider will influence the shared print community's strategic use of these tools and the process for improving them.

Resource sharing is a multi-faceted service that exists in a complex ecosystem. It includes ILL (Interlibrary loan), scanning, printing, and photocopying services of articles, book chapters, full volumes and in some cases, multiple volumes and providing access to vendor created digital copies. Access, whether physical or electronic, is governed by copyright, licensing agreements, consortial and partnership agreements, library determinations of condition and agreements and inter-institutional relationships among ILL staff.

## Membership

Linda Wobbe (SCELC), Chair  
Megan Gaffney (Univ. of Delaware)  
Melissa Gasparotto (NYPL)  
Tracy Gilmore (CSU, Long Beach)  
Rachel Manning (Middlebury College)  
Amy Wood (CRL)

## Process

We met with individuals and groups that sell or develop resource sharing infrastructure in order to understand the existing capabilities and future developments on the horizon. We advocated for enhancements, including support of Controlled Digital Lending.

Discussions were held with:

- ReCAP - Melissa Gasparotto (NYPL, ReCAP Product Steering Committee member)
- OCLC - Stacy Brunner, Jenny Rosenfeld, and Cynthia du Chane
- ExLibris Rapido - Mike Richins, Katy Aronoff, Tate Nunley
- Project ReShare - Sebastian Hammer, Index Data; Boaz Manes (Lehigh University)
- Auto-Graphics - Paul Cope CEO, Albert Flores VP Sales, Debbie Hensler SHAREit Product Manager, Cheryl Slinkard, CTO

Because Resource Sharing and Discovery are so interdependent, discovery of shared print retention commitments was also discussed in some cases, and Discovery Task Force and Infrastructure Working Group members were invited to attend some of these discussions. Our Task Force invited representatives from other Partnership Working Groups and Task Forces involved in Resource Sharing to report on their work at our monthly meetings to keep us informed of their work. Megan Gaffney reported the work of the Research & Network Level Resource Sharing Activities, and we were invited to give input to their survey to program managers. Marie Waltz met separately with Linda Wobbe monthly to report on the work of the Partnership + Rosemont subcommittee on Marketing shared print to Resource Sharing librarians.

To frame our discussions we explored the Open Data Working Group's Use Cases, and used a template to flesh out such questions as who would be performing the task and how success would be defined. These templates can be viewed [here](#) and were shared with the vendors and developers. As the year progressed, the CCH Collaborative developed a more expansive set of use cases "[Shared Print in the Collections Lifecycle](#)", which further helped inform vendors about our aims.

### **We asked vendors and developers to discuss whether their tools can:**

1. Identify the kind of access provided for retained holdings
2. Enable analysis of resource sharing activity for items committed for retention/shared print
3. Support interoperability with resource sharing systems to allow retained holdings to be preferred or deprecated for lending
4. Enable identification of highly accessed copies as candidates for "digitization FOR access"
5. Support controlled digital lending initiatives

## Findings

Each vendor/developer discussion was captured in a Responses table.

**ReCAP** utilizes a Shared Collections Service Bus Open Source software to create a centralized index of participating libraries holdings which is used to facilitate rapid resource sharing. The following ILS' are supported: Sierra, Alma, Voyager. Discovery options in use include Blacklight, Summon, Primo, and locally developed options. With development, this solution could be used by other programs to facilitate resource sharing and/or to create a shared index of retention commitments. Analytics, Controlled Digital Lending, and support of non-NCIP-supporting ILS' such as WMS would require more development.

**OCLC** focused on the potential for extending visibility and actionability of the shared print retention flag in OCLC's resource sharing tools, including WorldShare ILL, ILLiad, Tipasa and D2D. OCLC has provided a summary of possible future developments, and believes it will be possible to add the Shared Print retention flag in statistics, potentially in the 3rd quarter of next year, as part of updating Group Analytics. OCLC would like to meet with us again to talk about ways resource sharing can be enhanced within or among partnership programs.

**ExLibris Rapido** is a new resource sharing tool. A shared index of shared print retention commitments for the Partnership could be created and used by participating libraries as a discovery and resource sharing tool. Grouped together in a "pod", participating libraries agree to response and delivery protocols. Integrations with ILS systems other than Alma are on the development calendar. Controlled Digital Lending is on the development calendar. Analytics would be difficult because the system does not know which pod an item is borrowed from.

**Project ReShare** is a community-owned and developed Open Source resource sharing system which has co-investment from Index Data, the development partner. PALCI and ConnectNY will begin utilizing the system this year. ILS integration with almost all systems including WMS, Innovative, Alma, Sirsi Dynex and many more. Can be used as a Discovery target through export to EDS or via VuFind. Open Source shared index is built using a custom Harvester which can utilize OAI-PMH or harvest from an FTP site populated by the library on a schedule. De-duplication merge/match logic based on GoldRush MatchKey. Controlled Digital Lending needs future development. Possibility of leveraging the project or community to build infrastructure to support shared print; currently the 583 shared print information is not actionable.

**Auto-Graphics Shareit** is a patron or staff-initiated resource sharing system currently in use by 13 state-wide multi-type library networks. Interoperates with many resource sharing tools, such as OCLC's WorldShare ILL, Tipasa, ILLiad. Supports most ILS including Innovative, Polaris, Alma, Ex Libris, TLC, Koha, Evergreen, Syrsi Dynix, WMS partly (doesn't fully support NCIP) , Z39.50 or FTP updates; real-time status through NCIP real-time lookup. SHAREit could be used as a discovery system for Shared Print retention commitments across a group using the Union Catalog feature of SHAREit. They would create links back to each individual library's ILS,

providing real-time availability. The records could be downloaded for inclusion in a discovery layer, with the inclusion of links back to the SHAREit system to enable statistics reporting. Individual libraries or programs can FTP records on the schedule they choose. Registered retention commitments could be downloaded from OCLC and uploaded to Auto-Graphics' FTP site. Company Policy: the records always belong to the library. Some Controlled Digital Lending functionality is built-in, such as the ability to send a digital equivalent for an item in the returnable request workflow.

## Conclusions and Recommendations

Several resource sharing tools show promise in facilitating resource sharing and discovery within and among partnership programs. Development would be needed to fulfill all of the use cases and to gather comprehensive statistics. OCLC has not yet made the shared print retention flag actionable or visible in their resource sharing tools, but is exploring these developments. Full controlled digital lending has not yet been developed. Additional tools such as D2D and INNReach have not yet been investigated.

Recent technological trends have made most discovery and resource sharing tools cloud-based that enable work and growth at a scale previously unknown. Data is shared, reports are created and systems interact in more automated, IT driven approaches, such as APIs, than previously used by resource sharing staff. Anecdotal information indicates that while helping individual libraries meet reporting and data sharing goals, these new systems require more intervention from IT staff to set up reports or extract data. This is a possible barrier that is not addressed in this review, but is one the Partnership needs to understand in more detail before moving forward with a full recommendation.

For that reason, the Task Force will consider if one or more pilot projects should be supported by the Partnership to further understand operational barriers to developing resource sharing opportunities within and among Shared Print Programs. We also hope to have a deeper understanding of provider initiatives and resource sharing trends through our own investigations and from the outcomes of the CCH Collaborative Summit.

The Task Force hopes our investigation proves helpful to the CCH Collaborative and the Partnership, and looks forward to continuing the investigation as directed by the Executive and Operating Committees of the Partnership.

# Partnership for Shared Book Collections

## Discovery Task Force Report

### September 2021

---

## Purpose

To investigate feasible approaches to integrating retention commitments within a program or across programs for discovery by individual participating libraries.

## Membership

Susan Stearns (EAST), Chair  
David Almodovar (Pace Univ.)  
Maggie Dull (Univ. of Rochester)  
Amy Wood (CRL)

## Process

The Task Force explored the use cases defined by the Open Data Working Group and the CCH Collaborative Shared Print in the Collections Lifecycle to frame a set of questions to explore with vendors. To bring this discussion to a wider group, a few vendor conversations were held during the Infrastructure Working Group meetings. Because of the interrelationship between Resource Sharing and Discovery, several tools investigated by that group could be used for discovery of shared print retention commitments.

## Vendor Questions:

- What tools are available to expose retention commitments (defined by 583)?
- What Data is available in your system to expose retention commitments held by other SP program libraries?
- Is there an ability to pull in data to expose retention commitments held by other SP program libraries?
- Do you have ways to categorize data to expose SP titles but not in the home library and prioritize levels of access and/or speed of delivery.

## Findings

### **WEST member Discovery options / Anna Striker**

- WEST ingests holdings data from libraries, and has used that information to create MARC records
- WEST is making the records available to participating libraries to add to their catalogs or Discovery layers
- Problem: How can library users identify that these materials are not held on site, but available through ILL? Solution is that the holdings say WEST, so requests are sent to ILL, and the ILL staff decides from whom to request using their usual channels

### **ExLibris Discovery / Kat Cuff and Amy Wood**

#### **CRL experience with creating / MARC files and ExLibris knowledge-base targets**

- New streamlined process for the ExLibris Discovery and knowledge-base systems
- Kat Cuff is leading the team that runs all of these systems; comes from the SerialsSolutions/library background
- Bringing two systems in alignment: SerialsSolutions (360 Suite: Summon/360Link) plus the ExLibris complementary products (Alma/SFX/Primo)
- Discovery (Primo/Summon) now use a central discovery index
- Knowledgebases (SFX, Alma- (these two share an index) and 360 Link) - providers send a file to one place; huge job to align the disparate targets; working with individual providers to do so
- CRL uses III Millennium; they embed holdings in the bib records; they don't have the ability to download LHRs. Don't use LC or Dewey. Haven't sent anything with shared print retention commitments, but their retention commitments are in their MARC 583.
- Quarterly loads; will separate them into separate groups in the future (Open Access / All CRL)
- Can add referring parameters to the URL which says the source was referred from Summon/Primo to get data on how users are making the discovery
- Usage reports will be rolled out over time Discovery Reports for Databases; which collections a library has activated and usage for each.
- Provider portal in the future so providers can see what their content looks like in the various systems; in the meantime demo systems; SF
- With Clarivate coming in Kat really doesn't know what the future timelines are;
- In general; anything that requires development work is a multi-year project
  
- Partnership - if we downloaded our records from OCLC, would we need to include attached LHRs, and how could that information be leveraged? The ExLibris indexes are mostly electronic resources; not built to accommodate all of the nuances of MARC.
  - ExLibris would need to make product feature changes to be able to leverage this data
  - We can continue the discussion with the product managers
  - Maybe we could run an overlap analysis with shared print retention commitments and electronic resources; although many would not have OCNs so what is the match point?



- MARC records from OCLC may not be free to redistribute
- Explanatory links courtesy of Kat Cuff

For your reference, here are some of the recent pages we have put together to increase transparency into our system and processes:

- Resources for Content Providers:  
[https://knowledge.exlibrisgroup.com/Cross-Product/Providers\\_Page](https://knowledge.exlibrisgroup.com/Cross-Product/Providers_Page)
- Relevant too for mutual clients, how our products work together:  
[https://knowledge.exlibrisgroup.com/Cross-Product/Providers\\_Page/Content\\_in\\_Ex\\_Libris\\_Products/Using\\_Ex\\_Libris\\_Products\\_to\\_Manage\\_Electronic\\_Resources](https://knowledge.exlibrisgroup.com/Cross-Product/Providers_Page/Content_in_Ex_Libris_Products/Using_Ex_Libris_Products_to_Manage_Electronic_Resources)
- Collection list for SFX/Alma, updated monthly:  
[https://knowledge.exlibrisgroup.com/Alma/Content\\_Corner/Supporting\\_Resources/Alma\\_Community\\_Zone\\_Collection\\_List](https://knowledge.exlibrisgroup.com/Alma/Content_Corner/Supporting_Resources/Alma_Community_Zone_Collection_List)
- Collection list for 360 (will be updated monthly, awaiting a new list from the Operations team):  
[https://knowledge.exlibrisgroup.com/360\\_KB/Product\\_Documentation/General\\_Content/360\\_KB%3A\\_List\\_of\\_All\\_Databases](https://knowledge.exlibrisgroup.com/360_KB/Product_Documentation/General_Content/360_KB%3A_List_of_All_Databases)
- Summon/CDI demo UI: <https://demo.summon.serialssolutions.com/>
- Primo/CDI demo UI:  
[https://eu00.alma.exlibrisgroup.com/discovery/search?vid=EXLDEV1\\_INST:Alma](https://eu00.alma.exlibrisgroup.com/discovery/search?vid=EXLDEV1_INST:Alma)

## Resource Sharing Task Force findings relevant to Discovery.

The creation of a shared index of retention commitments was acknowledged as a potential solution for discovery during several of the discussions held by the Resource Sharing Task Force. Details of vendor conversations are recorded.

**ReCAP.** Melissa Gasparotto. With development, the SCSB (Shared Collections Service Bus) solution could be used by other programs to facilitate resource sharing and/or to create a shared index of retention commitments. Discovery options in use include Blacklight, Summon, and locally developed options. Current ILS' supported are Sierra, Alma, Voyager.

**ExLibris / Rapido.** Mike Richins, Katy Aronoff, Tate Nunley. A shared index of shared print retention commitments for the Partnership could be created and used by participating libraries as a discovery and resource sharing tool. Currently integrated with Alma; other ILS integrations are in development.

**Project ReShare.** Sebastian Hammer, Index Data; Boaz Manes, Lehigh University. Open Source shared index is built using a custom Harvester which can utilize OAI-PMH or harvest from an FTP site populated by the library on a schedule. De-duplication merge/match logic based on GoldRush. Can be used as a Discovery target through export to EDS or via VuFind. Current ILS' supported: WMS, Innovative, Alma, Sirsi Dynex and many more.

**Auto-Graphics Shareit.** Paul Cope, Albert Flores, Debbie Hensler, Cheryl Slinkard. SHAREit could be used as a discovery system for Shared Print retention commitments across a group using the Union Catalog feature of SHAREit. They would create links back to each individual library's ILS, providing real-time availability. The records could be downloaded for inclusion in a discovery layer, with the inclusion of links back to the SHAREit system to enable

statistics reporting. Individual libraries or programs can FTP records on the schedule they choose. Registered retention commitments could be downloaded from OCLC and uploaded to Auto-Graphics' FTP site. Company Policy: the records always belong to the library. ILS' supported: Innovative, Polaris, Alma, Ex Libris, TLC, Koha, Evergreen, Syrsi Dynix, WMS partly (doesn't fully support NCIP). SCELCL is in the process of testing this proposal and hopes to begin implementing in 2022.

## Conclusions and Recommendations

The landscape for shared print discovery options continues to develop. As can be seen by our investigations to date, many technologies exist or could be developed as a shared index of shared print retention commitments for use by individual programs or the partnership. All of the options investigated report the ability to export data for use in Discovery layers, while integration with a variety of ILS varies.

The Task Force and Infrastructure Working Group will continue to investigate additional discovery options and may propose one or more shared print retention commitment discovery pilot projects for consideration by the Partnership.

# Partnership for Shared Book Collections

## Unique Materials Task Force Report September 2021

---

### Purpose

The Unique Materials Working Group of the Partnership was tasked with determining the current and potential capabilities of existing systems to identify unique or scarce materials in member libraries, both to identify items which should be under retention but are not, and to make libraries aware of what they have that is unique or not widely held. This report focuses on ways to use existing metadata in existing systems, as a proxy for a physical resource, in determining unique materials.

### Membership

Sara Amato (EAST), Chair  
David Almodovar (Pace Univ.)  
Judy Dobry (CDL)  
Jennifer Martin (Salisbury Univ.)  
Tom Teper (Univ. of Illinois)

### Scope of Project

While the group was tasked with a general goal, it quickly became clear that there is a need to define what is meant by 'unique' as well as to determine the population of materials we are looking at. A separate task force has been set up to explore the question of how to define unique and scarce, most likely reporting out by the end of 2021. For the matter of which materials to look at, questions to consider include: Are we focused on retained titles only? Or all holdings of all member libraries of Partnership programs? Or the entire population of known published print monographs? Are we interested in limiting to English language only and/or by place of publication, e.g. North America, or are we looking more broadly?

There is also a question of what we want the output from a collection analysis to be:

- A one time .csv file?
- A dashboard?
- An ongoing tool to query?

These questions will need to be answered before beginning any work with vendors on identifying unique materials.

We also have an opportunity to broaden the scope of the question to include Risk questions. For example, can retentions be mapped geographically (how many copies in more than one risk zone)? What storage facility types are items held in?

## Vendors

The group held 1 hour meetings with:

- OCLC (GreenGlass and WorldCat) - Andy Breeding and Matt Barnes
- Colorado Alliance (Gold Rush) - George Machovec, Rose Nelson and Chet Rebman
- ExLibris (Alma and Alma Analytics) - Jane Burke
- IndexData (ReShare and Folio, Central Index) - Sebastian Hammer

The **questions guiding these conversations are in Appendix A**. Links to the **raw notes** from these meetings are in **Appendix B** (access limited to Partnership Infrastructure Working Group, OC & EC).

## Vendor Responses in Key Areas:

Scope of materials available currently in their systems	
<b>IndexData</b>	PALCI and CNY are using and could possibly work with Project ReShare to use data
<b>ExLibris</b>	Have over 2,000 libraries using their Alma LSP
<b>OCLC</b>	Current local holdings data of WMS libraries, holdings of all cataloging partners ( <a href="#">15,637 in 107 countries</a> ). Most likely, the majority of the Partnership libraries
<b>Gold Rush</b>	EAST and CARL retentions. Keep@Downsview. Possibly Ivy Plus data from a few years ago

Ability to work with libraries currently not using their service	
<b>IndexData</b>	They can use their harvester to get data in and transform it. Data could then be pushed anywhere. For libraries without OAI/PMH capabilities, they can submit MARC files.
<b>ExLibris</b>	Can import MARC into a separate Alma space. A one-time project is possible. Less expensive than subscribing to LSP.

<b>OCLC</b>	Can import MARC from libraries into GreenGlass. They would do some verification of data and also look up OCNs where none exists.
<b>Gold Rush</b>	Any library that can deposit MARC records can participate.

### Matching algorithms - ability to detect unique materials

<b>IndexData</b>	Currently using the Gold Rush MatchKey. They understand that it can be problematic for our use case. Would consider developing additional matchkeys and cascading matching algorithms. MatchKey is a 'living document'.
<b>ExLibris</b>	Queries can be built using whatever criteria the user desires, with cascading match points. Also, building AI algorithms to judge the quality of records - uniqueness will be part of that algorithm.
<b>OCLC</b>	Uses OCLC number, distinct manifestations. Can also report out at the work level. "Unique" is identified by holdings level on OCN.
<b>Gold Rush</b>	Uses MatchKey, which has evolved over time. Willing to consider secondary match points. Have recently added an 'overlap/uniqueness' facet that could be used to determine unique and scarcely retained across collections.

### Identifying digital copies (IA, HT)

<b>IndexData</b>	Working with IA and could possibly have integration. Having discussions with HT.
<b>ExLibris</b>	Have HT data and could most likely work IA data into reports. Do have a relationship with IA as a fulfillment partner for RapidILL.
<b>OCLC</b>	Have HT data; do not currently have IA matching capability.
<b>Gold Rush</b>	Do not currently have. If we could provide the records, they could load them and do comparisons.

## Can they run a one-time report of Unique using the dataset and tools they have today?

Mostly all said yes to this type of query. ExLibris and OCLC have done this already. If data is in Gold Rush they can also do it. Most likely IndexData can do this too, though they have a far more limited data set as of today.

### Culture

All were very responsive to our requests for meetings and generous with their time. Below are quick notes/impressions.

<b>IndexData</b>	Open Source software, open culture. Prefer NOT to be a monolithic entity but to help create an open/sharing environment of bibliographic data.
<b>ExLibris</b>	Corporate. Are creating a sizable bibliographic database and thinking about their bibliographic ecosystem.
<b>OCLC</b>	Well known already in community, corporate culture. OCLC Research might be interested?
<b>Gold Rush</b>	Not Open Source, but built on Open Source tools. Responsive and collaborative.

## Strengths/Weaknesses

### IndexData

#### Strengths

Open Source; community building up around them with ReShare and Folio. Both a pro and a con is that their work is incipient; PSBC could influence the direction or creation of services.

#### Weaknesses

Not a lot of data in their systems as of today in comparison to OCLC or ExLibris.

### ExLibris

#### Strengths

Robust reporting system already in place, could be extremely flexible, and for libraries currently using the system could provide ongoing analytics at no extra cost.

**Weaknesses**

Unclear how quickly they could move to reindex to maximize 583 searches or how it might fit into their goals. Not all partner libraries use ExLibris, unclear what the cost would be to load and/or maintain data in their cloud. Searches not currently optimized for 583 data nor uniqueness.

**OCLC****Strengths**

Majority of our libraries are already members and data is available. Have an understanding of Shared Print and now have a large set of normalized retention data. Also have large [data sets in GreenGlass](#).

**Weaknesses**

Holdings may or may not be up to date; some libraries moving away from OCLC. Identification of 'unique' would most likely need work, though this is true of all vendors. Based on GreenGlass prices, costs may be high.

**Gold Rush****Strengths**

Flexible and willing to work with users to meet needs. Have an understanding of Shared Print.

**Weaknesses**

Not a lot of data in their systems as of today in comparison to OCLC or ExLibris.

**Summary Table:**

Vendor	Data Currently in System	Match point for Unique	Pros/Cons
ExLibris	2,000+ libraries use as LSP.	Can be defined on bib data points	Existing system could be leveraged at minimal ongoing cost, but many not in that system.
Gold Rush	EAST and CARL retentions. Keep@Downsview data. Other small projects.	Matchkey based on bib data - flexible	Flexible and open, knowledge of shared print. Small data set currently.
Index Data	PALCI and CNY use ReShare. Several	Matchkey based on bib data - flexible	Flexible and open, willing to work on

	Folio libraries. (growing)		developing shared print infrastructure. Small data set currently.
OCLC	All OCLC cataloging subscribers.	OCN	Large data set, knowledge of shared print. Limited to matching on OCN.

## Considerations/Thoughts

Determining uniqueness is hard. Some of the issues that come into play are unique cataloging, variations in cataloging practices, merged print and electronic records and works with many publications/editions that might appear rare or unique when they are not.

We may need to explore this on a smaller/pilot scale first if we want usable data output. Some points to consider when deciding when and how to engage with these vendor are:

**-OCLC and ExLibris** have strong **existing products and user bases**. Little effort on part of member libraries to engage if they already use one or both of these vendors.

**-OCLC** has the **most development in built out systems**, with Gold Rush also having strong Shared Print capabilities, though not as large of a built-in Data Set.

**-Strong possibilities for engagement with ReShare (IndexData)** - but might take more time. IndexData in conversations with Ivy Plus.

**-IndexData** could contract for consulting to scope the project, perhaps a trial. Could also just be **used as a harvester** and data could **then go to Gold Rush** or other analytics.

**-ExLibris** is working with **ReCAP** (for resource sharing) and **Orbis/Cascade Alliance**.

We might want to consider a separate retentions database built by one of these vendors that could show overlap, and ability to upload data to it to report overlap.

## Recommendations

Thinking of our primary use case as being finding unretained unique materials and also identifying uniquely retained materials for more focused preservation work, these are our recommendations:



- Identify an expert to help evaluate vendor matching / record matching. Finding records for the same entity within and across different databases is something that has received quite a bit of attention in computer science ([see Duplicate Record Detection](#)) and the advent of machine learning has increased interest in this area. It might be worthwhile to consult with some academic experts on this topic to get their recommendation on the best approach to finding duplicates in and across library catalogs. This could give us both a better understanding of how the algorithm needs to be written and how a database needs to be set up.

This type of advice could be unpaid advice based on an hour's conversation or a situation where we use part of a grant to pay someone to look more deeply. This would be similar to the work done by Candance Yano to determine how many copies should be kept for shared print. Since shared print is just one of many places within the library ecosystem where finding duplicate records is helpful, we might find funders quite interested in such a project.

- Decide if we care more about precision (finding only unique items, even if that means missing some unique items) or recall (finding all unique items, even if there's also false positives). There may be a way to have the system indicate its confidence in identifying unique materials (essentially a confidence score), which would allow us to determine a threshold where we bring in human evaluation versus relying on the machine evaluation.
- Small trials with iterative processes would be desirable. Possible projects could include:
  - Define/Scope a pilot project, perhaps with assistance from IndexData. Need to determine clearer goals before proceeding, e.g. focus on one-time projects or an ongoing platform?
  - Pursue a small grant for a small project, perhaps with a consortium - e.g. could use PALCI data in ReShare or Alliance, California community colleges, CARLI, SUNY Libraries, and ReCap data in ExLibris. Might consider asking ExLibris for Unique reports for Partnership libraries.
- Could try establishing a one-time project/report with one vendor (to set up future 'on demand' type reporting) and a platform with another. E.g. reports from OCLC while pursuing a platform with IndexData or Gold Rush.
- Should re-engage with the CCH project as to how this might fit in what they are doing.

## Further Engagement

Might be worth talking with Internet Archives on identifying materials unique to them and thus good candidates for consideration for print retention, or conversely determining if any retained titles are NOT in Internet Archives and thus good candidates for digitization.

## Appendix A - [Questions guiding conversation with Vendor](#)

**The Partnership for Shared Book Collections** is a federation of monograph shared print programs in the U.S. and Canada whose mission is to ensure the long-term preservation of, access to, and integrity of monographic print resources. The Partnership supports both established and fledgling monograph shared print programs by coordinating collaboration for the protection of print books. Currently over 400 libraries participate in programs that are part of the Partnership. See <https://SharedPrint.org> for more information, including a [list of participating libraries](#).

The Unique Materials Working Group of the Partnership is interested in determining the current capabilities of existing systems to identify unique or scarce materials in its member libraries.

This is an introductory conversation which may eventually lead to more in depth discussion and request for quotes.

### **Questions:**

What types of metadata do you have available today to determine

- Scarcely held material across the Partnership
- Scarcely retained\* material across the Partnership

What resources (both in terms of data and financial resources) would you need in order to identify unique unretained (at risk) materials?

Could you handle records for libraries who don't already use your services?

How do you handle identifying identical editions across different datasets which may have similar but not identical records for the same edition?

One need that has been brought to our attention is determining what can be done quickly if a library is closing and needs to disperse its commitments. Being able to quickly identify unique or scarcely retained\* titles would be ideal. Do you have any lightweight 'emergency' collection analysis tools? If not, what resources (both in terms of data and financial resources) would you need in order to do so?

Talk a minute about your staff/systems department and security.

Can you give us a very high level view of your business model just to give us an overall view of the products sustainability.

**Here are a few user stories to flesh out scenarios around the above questions:**

*A librarian at xyz college wants to know if a monograph in their collection is unique in the partnership retentions\* in order to make informed withdrawal decisions.*

*A Program Manager at a Shared Print program would like a list of unique materials at their member libraries in order to propose those titles be retained.*

---

\*The term “retained” or “retentions” refers to materials that have been committed to be retained to a shared print program by the holding library. These can be identified by the presence of a 583\$a of “committed to retain” in the library's bib or holdings records.

## Appendix B - Notes from Meetings with Vendors

Access to these notes was provided to the Partnership Executive and Operations Committees