# Column Title: Library Networking and Consortia

By George Machovec, Executive Director, Colorado Alliance of Research Libraries , 3801 E. Florida, Suite 515, Denver, CO 80210 ([george@coalliance.org](mailto:george@coalliance.org))

This column focuses on formal collaboration and networking among libraries through consortia. It offers in-depth examinations of issues facing modern library consortia including (but not limited to) e-resource licensing, ebooks, next generation integrated library systems, shared print archiving, shared digital repositories, governance and other relevant topics. Contributions are accepted for this column and must be submitted to George Machovec ([george@coalliance.org](mailto:george@coalliance.org)). Contact the column editor for suggested topics, deadlines and formatting.

Keywords: Shared Print Archiving, Assessment, Collection Evaluation, Collection Overlap, Academic Libraries

# Article Title: Shared Print Archiving – Analysis Tools

## Introduction

Academic libraries have shifted their collection development focus from print to digital resources for both serials and monographs. As publishers, vendors, libraries, Google and others have made substantial progress in digitizing backfile collections, libraries have new opportunities to reinvent their mission and physical space. Libraries have been using two primary methods to downsize their physical collections in order to make additional space on campus available for collaboration, research and new initiatives. Moving materials into less crucial off-site storage (or high density storage on campus) allows lesser used materials to still be available but not take up prime real estate in library facilities. Weeding of collections for items that are digitized or are low-use (but widely held) is also a widely used strategy. Many academic libraries work in regional or national initiatives to coordinate their shared print management activities in order to preserve the scholarly record but also to make sure that valuable materials are still available to scholars and students when needed in print. In order to make better decisions, different services and tools have been developed which can help libraries analyze use and compare collections.

The coordinated management of physical collections has different emphases for various initiatives in order to provide ongoing access to the scholarly record. Some initiatives have focused on serial collections where great space savings is easily available. The digitization of journal backfiles by publishers or programs such as JSTOR have provided superb access to historical collections thereby greatly reducing demand for print versions. Coupling this with archiving programs as Portico and LOCKSS (or CLOCKSS) have given librarians greater confidence to store or weed large numbers of serial volumes knowing that easy access and secure archiving are available for the scholarly record. Coordinated programs for monographic materials also hold great promise for space saving. Once again digitizing by publishers, Google Books, HathiTrust and others have provided superb access to many monographs. Although decision making for weeding and storage on monographs is sometimes more difficult, libraries are using different tools and techniques to aid in the process. Through a process of comparing collections with peers, analyzing circulation data and using comparative holdings; libraries can begin to make large scale decisions about their collections.

Other material types such as microforms, government documents and domain specific materials are also candidates for storage and weeding through collaboration. A great deal of material is available online and in published articles about specific initiatives such as at the Center for Research Libraries (CRL) Print Archive Network (PAN) and through regional archiving programs such as WEST, ASERL and others. Some of the tools and services being developed by these initiatives and other commercial vendors are the focus of this article. The order of the services is random and does not indicate quality, cost or any other preference. In addition, although the following services may have overlapping functionality and some of the same target audience, some of these services have variant purposes and libraries could choose to use multiple services for different needs. Tools will be sorted by those that are optimized for monographic and serial analysis, although some may be used for both types of analysis.

The manual comparison of monographic or journal title lists extracted from local integrated library systems or electronic resource management systems (ERMS) could certainly accomplish overlap and gap analysis between partner libraries and other initiatives. However, substantial effort would need to be expended to perform these comparisons particularly since a group of libraries will need to export the data, load the data into Excel, Access or other database management systems and then run comparisons. Except for projects that are small in scope, it is worthwhile investigating commercial and open source tools for monographic and serial overlap and gap analysis

# Monographic Analysis Services and Tools

## Sustainable Collection Services (SCS)

Sustainable Collection Services (SCS) offers “**deselection decision-support tools** to academic libraries. SCS tools enable carefully managed drawdown of low-use print monograph collections while supporting shared print archiving efforts. Uniquely tailored reports combine local circulation and item data with WorldCat holdings, Hathi Trust Digital Library holdings, and authoritative title lists. The result: **cost-effective, evidence-based decisions.”** <http://sustainablecollections.com/>

This service was begun in 2007 by Rick Lugg and Ruth Fischer of R2 Consulting after working for many years in the area of workflow analysis in libraries. At that time, they saw the need in academic libraries to reduce the footprint of their collections in order to position libraries for the future. In an article published in 2008 in *Against the Grain* (<http://r2consulting.org/pdfs/2future_tense_lugg.pdf>) they posit eight reasons for why substantial weeding programs in academic libraries should be considered now.

1. Digital accessibility of content
2. Tools and infrastructure for resource sharing between libraries has improved
3. In most cases withdrawn copies of content is easier to replace if a mistake is made
4. Circulation of print materials is down and easy to obtain so that better decision making based on use is possible
5. Volume counts for academic libraries is no longer the leading metric for value among academic libraries
6. There are improved tools for data driven decision making
7. Space for physical collections is at a premium and universities are less likely to continue expanding in traditional growth models
8. Libraries cannot afford to continue the status quo

SCS has developed a sophisticated set of tools for analyzing a library’s collection to compare it against partners and peers. In addition to harvesting holdings data from OCLC WorldCat a library will also submit circulation data to fold into the analysis. SCS can work with an individual library doing a peer analysis through OCLC or can work with consortia to do an analysis tailored for specific needs. In addition to analyzing printed collections, it is also possible to compare collections against digital collections such as HathiTrust or major ebook aggregators in order to determine what should be weeded or stored. Libraries and consortia have different goals and the custom tailored work of SCS is considered one of the best services available. The challenge with the SCS service is the high cost as it includes both a one-time set-up fee as well as a per MARC record charge in the pricing algorithm. Libraries also need to consider what is important for the upfront analysis and how they will proceed in future years as decisions are made and the ecosystem changes.

Some of the consortia which have contracted with SCS include ConnectNY, Washington Research Library Council (WRLC), Academic Libraries of Indiana, California State University System and many others. A full list of customers can be found on their Website.

## OCLC Collection Evaluation

The OCLC Collection Evaluation Service (<http://oclc.org/collection-evaluation.en.html>) replaces a previous service called OCLC Collection Analysis and allows libraries to evaluate their holdings against other libraries, groups or benchmark groups. The OCLC service can import circulation data from either a local integrated library system (OCLC will provide the specified format of the data) or can be automatically supplied if the library is using OCLC’s WMS service as their integrated library system. Most of the data analysis can be done by subscribing libraries without programmer intervention and most reports are run overnight. Data visualization tools are available and data may be exported for local analysis. Analysis can be done with up to 50 other libraries at one time. The costs for the service are custom quoted by OCLC.

OCLC suggests several uses for its services which include:

* Deselection
  + Which materials are in high demand
  + Which materials may be easily borrowed
  + Are electronic versions available
  + What is unique or rare based on OCLC’s massive holdings
* Accreditation
  + How does a library’s collections compare with others with similar academic programs
  + Collection weaknesses
  + Collection currency
* Acquisition

## Intota Assessment

ProQuest is in the process of developing a new integrated library system called Intota. The first module to be released for this new offering is Intota Assessment (<http://www.serialssolutions.com/en/services/intota/assessment>) which is designed to do monographic and serials analysis for holdings, use data and overlap for material in both print and electronic format. Intota Assessment is a standalone service and does not require a library use the Intota integrated library system (which was not even released when this first tool was released). One of the many strengths of Intota Assessment is that it incorporates into this analysis tool many of the other ProQuest services (Serials Solutions has been retired as a name and all of its services are now under the ProQuest umbrella name). ProQuest has folded into Intota Assessment tools such as Books in Print, Resources for College Libraries, Ulrich’s, and the Serials Solutions knowledgebase. In addition, the system can ingest COUNTER use reports so that serials use can be incorporated into a complete analysis.

To use Intota Assessment a library will typically provide a set of MARC records, circulation data and cost data (monographic costs can either be supplied by the library or a library can choose to use Books in Print retail costs). If a library uses Summon or ERMS services from Serials Solutions, the MARC records for monographs or the serial title lists can be directly harvested so that it does not need to be supplied multiple times.

Intota Assessment reports are run in real-time with a growing number of data visualization tools. Data are presented in easy-to-read tabular format and the system can provide side-by-side analysis of peers (with their approval) as well as electronic availability of ebooks if a library would like to incorporate e-availability into the decision making process for deselection or storage. Data may be sorted, manipulated and downloaded in a variety of ways to allow real-time analysis for decision making. Cost per use analysis is provided for both monographs and serials based on supplied or retail pricing.

The cost for Intota Assessment is via an annual subscription fee based on size of collection and FTE and discounts are available for early adopters or those bundling the service with offerings from the vendor.

## GIST Gift and Deselection Manager

The GIST (Getting It System Toolkit) Gift and Deselection Manager is an open source software tool developed by the SUNY Geneseo (the developers of the IDS Project) to assist libraries in expedited decision making in doing batch analysis of gifts and also to improve decision making for the weeding of library collections. The focus of this software is on monographs.

“It is standalone open-source software that automates the gathering of data for evaluating donations; including holdings, edition comparisons, full-text, and other data.  It also recommends “keep or do not keep” according to your collection building conspectus file, and if you keep an item, GDM can open and search the work in OCLC Connexion so you may attach holdings and download a record.  The Gift Manager also automates donor letters with lists of donated items – no more keying in title information for acknowledgments, with all the data now stored in GDM’s database for future reference.

The GDM also enables collection managers to perform item-by-item deselection or use the batch analysis tool to create custom deselection reports for large weeding projects.” (<http://www.gistlibrary.org/gdm/>)

As an open source initiative the software may be downloaded at no charge under a Creative Common license and may be locally run by a library or consortium. The batch deselection tool allows a library to load and compare its holdings with other large online collections including Google Books, HathiTrust and Better World Books Library. Through comparison with these major online collections libraries will have more confidence in weeding monographic materials knowing that have been digitized and are being preserved. Of course, a library or may not have immediate online access to all titles in these national initiatives but the library can be confident that the monographs being discarded have been digitized. The deselection report includes output including title, author, publication date, call number, OCLC number and holdings information. The comparisons will inform a library if deselection candidates are widely held or easily replaced if weeded.

As with any open source software, a certain amount of local technical and programming expertise is required to download, configure/load local data sets and perform analyses. But with some pre-loaded data sets, libraries will be able to compare what they have with the larger environment.

## INN-Reach Union Catalog

The INN-Reach union catalog software from Innovative Interfaces Inc. (<http://www.iii.com/products/innreach>) was created in the early 1990s for the OhioLink consortium to aggregate the collected holdings of all of the academic libraries in Ohio, create a searchable interface and provide patron-initiated requested of returnable materials. Since that time, approximately 20 additional union catalogs have been installed around the world representing over 1,200 libraries in regional or national clusters. Although the focus of the software has been on discovery and patron-initiated borrowing for both academic and public libraries, Innovative has begun to focus some development on the opportunities for shared print archiving. The strength of the INN-Reach systems is that they have sophisticated software for ingesting MARC records, holdings information and updated circulation status for all participating libraries whether they use the Innovative integrated library system or not (non-Innovative libraries use special bridging software called DCB which can import and map data as well as provide patron authentication for requesting). A master MARC record is created for each bibliographic entity and the holdings of each library are attached.

Historically, Innovative has provided a “Create List” functionality which allows libraries to do ad hoc queries of the MARC and item records embedded within the system. Although this does provide a deep dive into the collective holdings as well as the ability to export data, the Create List functionality has not been optimized for quick and easy comparisons thereby requiring substantial work to perform analysis for collection comparisons. Understanding these shortfalls the company is now developing a series of new tools to assist libraries in analyzing the rich metadata embedded in their systems. The first step in this process has been the ability for the central system to import multiple 583 tags where libraries have recorded their retention decisions, similar to OCLC (earlier only the 583 of the master record would have been centrally stored). This will allow libraries to quickly and easily check the retention commitments from all of the libraries in their shared catalog without having to consult individual local catalogs or OCLC. As a second phase Innovative is planning for a series of tools to help libraries easily compare collections to determine overlap to make better decisions on what to deselect or put in storage.

If a library is part of an INN-Reach union catalog where the participants are cohorts in shared print archiving decisions, the developments with INN-Reach are certainly worth considering.

## Ebrary Title Matching Fast (TMF)

The ebrary Title Matching Fast (TMF) service has been created by ProQuest (parent to ebook aggregators ebrary and EBL) and allows libraries to match their print holdings against ebook availability for both ebrary, EBL and their various product lines (including ebrary Academic Complete). In this service a library exports all of its ISBNs from its catalog and submits them to ebrary for analysis. The ISBNs are cross checked with *Books in Print* (which is owned by ProQuest) for normalization and correction. These are then matched against the ebook holdings in the ebrary and EBL ebook systems (these two aggregators are in the process of being merged since ProQuest acquired EBL) and a report is produced which tells the library which printed books are available in electronic format. In addition the report indicates which type of license an ebook may be acquired (e.g. is it available for perpetual access, is it in a Demand Driven Acquisition model). The value of the report is that if a library needs a quick way to determine what is available in electronic format in comparison to their current and historical print monographs a report can aid the library in storage and weeding projects. Of course, the vendor hopes that this will generate subscriptions to its offerings, but even if a library cannot subscribe to everything immediately, it is quick way to determine digital availability so that a library can make better decision with the legacy print materials. The best part of the service is that it is free and takes relatively little effort from the library to get started. Intota Assessment would also offer this same overlap analysis in its broader suite of tools but at a cost.

# Serial Analysis Services and Tools

## CRL Print Archives Preservation Registry (PAPR)

The Center for Research Libraries (CRL) has played a central role monitoring and coordinating serial print archiving programs around North America. One of their key programs is the Print Archives Preservation Registry (PAPR). “The Print Archives Preservation Registry (PAPR) system supports archiving and management of serial collections by providing comprehensive information about titles, holdings, and archiving terms and conditions of major print archiving programs.” (<http://www.crl.edu/archiving-preservation/print-archives/papr>)

The PAPR database is available online (<http://papr.crl.edu/>) and allows libraries to immediately query the status of specific journals to determine which regional initiatives have agreed to retain which titles along with specific retention policies. As of January 2014 over 30 program data sets have been ingested and users may query the holdings commitments by regional program, storage facility, archiving institution and journal title. In addition data may be viewed by subject area, the number of holdings institutions, years of coverage and level of validation. Some of the title lists may be downloaded for local analysis and use. Links are provided to regional programs and any associated agreements or memoranda of understanding (MOU).

The PAPR system is provided as a service by the Center for Research Libraries (CRL) and is available at no cost. It will help individual libraries and groups to determine what other programs have committed to preserving journals and may aid in retention or deselection decisions.

CRL has played a key role in working with OCLC on a Print Archives Disclosure Project the establishment and use of the 583 field in the MARC record as a universally recognized place to record retention decisions. This key field is now being incorporated into local catalogs and retained at the institutional level in OCLC.

The 583 field is used to record information about preservation actions. For purposes of current and future shared print initiatives, the 583 is recommended for use in recording print retention commitments and related actions. Such commitments are made by an institution as part of its participation in a collaborative print archiving program. The specific information included in these preservation action notes include:

* Action and type of physical review undertaken to verify completeness or condition, if any (retained, condition reviewed, completeness reviewed)
* Retention commitment (date the retention commitment expires)
* Institution (i.e., making the commitment)
* Authorization (the Print Archive program(s) to which the materials are contributed)
* Validation level (none, volume, issue, or page level)
* Outcomes of validation (includes standard terms that are or will be integrated into the Preservation & Digitization Actions: Terminology for MARC 21 Field 583 (PDA). e.g. explicit enumeration of missing volumes, issues, presence of reprints, tight bindings, damage)
* Materials Specified (if different from the LHR 85x/86x)
* Uniform Resource Identifier (link to program documentation for print archiving program)

## JRNL

The University of Florida Libraries has been developing open source software called JRNL to support the shared print journal archiving efforts of academic libraries in Florida through the FLARE (FLorida Academic REpository) initiative as well as the shared print programs of ASERL Association of South Eastern Research Libraries (<http://www.aserl.org/>).

“The Journal Retention and Needs Listing (JRNL) tool was developed as a way for participating institutions to track journal retention commitments between the Association of Southeastern Research Libraries (ASERL) and Florida State University System (SUS) partners. It allows users to upload current journal holdings and identify particular gaps in those holdings. Further, users can determine if thier institution has holdings that could be donated to other institutions to fill those gaps.” (<http://guides.uflib.ufl.edu/JRNL>)

The software allows libraries to upload and track print journal retention decisions in a group so that libraries can make better decisions on what to store or discard. Once delimited journal holdings data are imported the software provides sophisticated tools for overlap and gap comparison with partners along with the ability to download this data for local manipulation and use. In addition to determining gaps in a local collection one can identify other partner libraries to help fill gaps from libraries that are planning on deselecting materials.

Although the software was developed for Florida and ASERL initiatives, it would be useful in other cooperative journal archiving programs but will take effort to load and manage the software environment and work with local partner libraries to create useful data sets.

## Gold Rush Decision Support

A number ERMS/link resolver systems have developed online comparison tools to compare serial title lists from publishers and aggregators. These comparison tools were primarily developed for decision making on what publisher and aggregator packages should be purchased or deselected. Some of the leading tools have been developed by Serials Solutions (now renamed under the ProQuest brand), CUFTS Open Source Serials Management (developed by Simon Fraser University in Canada) and Gold Rush Decision Support (Colorado Alliance of Research Libraries).

The Gold Rush Decision Support (<https://www.coalliance.org/software/gold-rush>) system offers a rich knowledgebase of over 2,000 publishers, aggregators and indexing/abstracting services and has some unique functionality. The Gold Rush system also allows libraries to load their own serial title lists whether they be local journal collections, print backfiles and even exports from other ERMS vendors so that libraries can compare what they own against commercial, open access or other participating libraries. The software can compare one-to-one, one-to-many or many-to-many title lists in the same simple interface. The comparisons are run in real time with graphic visualization of unique and overlapping content with the ability to email or download data in popular formats such as Excel, .csv, PDF or text.

In addition to full-text content, Gold Rush has also retained where serials are indexed so that one can compare complete title lists or filter them to the indexing or full-text components only. Among the pre-loaded title lists are all major publishers, all major aggregators, all major indexing/abstracting services, open access journals as well as Portico (to determine if your collection or publisher sets have been archived), Get it Now (Copyright Clearance Center), and selected regional archiving projects. If a library has loaded its own serial holdings, quick comparisons with these other initiatives can quickly performed and visualized with detailed downloadable data.

Gold Rush is not open source but can be licensed from the non-profit Colorado Alliance of research Libraries at a very low cost (full disclosure: the author of this column is the executive director of the Colorado Alliance of Research Libraries).

## Conclusion

Academic libraries around the world are collaborating on the management of legacy print collections for both serials and monographs. Reducing the footprint of physical collections in order to re-purpose the focus of traditional library buildings and to emphasize digital library offerings are some of the big challenges in the early part of the new century. However, legacy print collections are important to retain in order to help ensure the preservation of the scholarly record, to provide access to content that may not have been digitized and make sure that digitized materials are of sufficient quality. However, libraries cannot act in isolation in this effort so that regional, national and international efforts are addressing the key issues.