

**Michigan Shared Print Initiative (MI-SPI)**

**Preliminary Proposal for 2015 Shared Print Project**

**July 24, 2014**

**Overview**

The Michigan Shared Print Initiative (MI-SPI) developed in the course of SCS’s first group collection analysis projects in 2012-2013. SCS work on that project was completed in late 2013. Given the success of that pioneering project, and the continuing need to manage print book collections collaboratively, MI-SPI plans to repeat and expand its efforts in 2015. This time, the group will include 12 participants, representing all of the public academic libraries in Michigan except University of Michigan, Michigan State, and University of Michigan-Flint.

As with the initial project, there are a several dimensions to the propose analysis , including a desire to quantify usage and overlap of holdings; to share costs related to retention; to identify and protect scarcely-held titles; and to withdraw surplus copies of widely-held, low-use titles. MI-SPI is also interested in incorporating regular collection analysis into its ongoing operations—i.e., to develop a model that supports analysis on demand, scheduled data refreshes, and predictable budgeting.

Project participants want to use data and evidence-based methods to analyze their respective collections, and to create retention and withdrawal scenarios that support the primary goals of freeing library space for other purposes, while also safeguarding the group’s collective holdings. SCS proposes to provide data analysis and related consulting services in support of MI-SPI objectives. SCS shares with MI-SPI the goal of creating an ongoing program of data work and use of our GreenGlass decision-support software to help Michigan libraries manage their print collections effectively.

**Project Scope**

The libraries listed on the chart below will participate in the 2015 project. Together, their data and collections represent 12 libraries, 12 separate data sets, and approximately 4.6 million bibliographic records. (Counts may change somewhat as SCS and MI-SPI work through scoping parameters.)

In Scope [to be confirmed]:

* Circulating print monographs
* Reference books

Out of Scope [to be confirmed]

* Serials (anything with record type ‘s’)
* Special Collections
* Government Documents (other than those classed in LC or DDC)
* E-Books
* Micro-formats/Fiche
* Audio-Visual/Media
* Lost/Withdrawn items
* Theses/Dissertations
* Music scores
* Maps

**Participating Libraries and Estimated Record Counts**

|  |  |  |
| --- | --- | --- |
| **Institution Name** | **# of monographs** | **Integrated Library System** |
| Central Michigan University | 558,759 | Innovative |
| Eastern Michigan University | 550,268 | Voyager |
| Ferris State University | 153,422 | Innovative |
| Grand Valley State University | 292,904 | Innovative |
| Lake Superior State University | 125,179 | ? |
| Michigan-Dearborn, Univ of | 238,792 | ? |
| Michigan Technological University | 174,538 | Voyager |
| Northern Michigan University | 280,994 | ? |
| Oakland University | 384,638 | Innovative |
| Saginaw Valley State University | 169,899 | Innovative |
| Wayne State University | 700,473 | Innovative |
| Western Michigan University | 966,194 | Voyager |
| **TOTAL** | **4,596,060** |  |

**The SCS Approach to Shared Print Projects**

**Experience**

Since 2012, SCS has completed collection analysis work for approximately 150 libraries. This has included projects focused on independent action in a collective context as well as formal shared print projects. SCS shared print projects include the following (references upon request), though MI-SPI can also draw from its own experience.

Please note that SCS methods and tools have evolved significantly since the original MI-SPI project. We have more capabilities at our disposal, and more importantly, some new strategies and tools that enable a balance of group-level decisions and independent action based on those decisions. We would be happy to describe this in more detail at MI-SPI’s convenience.

***California State University System:***Initially a pilot project involving six Los Angeles area libraries, 3.5 million records. The project is still underway, with 17 libraries participating, with a mixture of independent analysis and regional shared print groupings.

***Connect New York:*** Twelve private academic libraries of various sizes, 3.5 million bib records. Project completed in early 2013. A mixture of retention commitments and withdrawal opportunities, with custom allocation across the group.

***Maine Shared Collection Strategy (MSCS):*** Nine libraries, 3.9 million records. This IMLS-supported project included two public libraries, the State Library, five academic libraries, and a theology library. Primarily focused on retention commitments. Work completed in 2013.

***Tri-University Group (TUG):***Three academic libraries in Ontario, 2.2 million records. De-duplication and weeding of storage facility, plus retention commitments. Work continues.

***Central Iowa Collaborative Collections Initiative (CI-CCI):***Five libraries in the Des Moines area, 1 million records. A balanced approach to retention and withdrawal, with equitable allocation (based on collection size) across the group. Collection sizes ranged from 81,000 to 500,000. Work completed in 2013.

***Virtual Library of Virginia (VIVA):*** Pilot project involving 12 libraries, 5.8 million bib records. Libraries range from community colleges to ARL’s. Goals include some weeding, but also gauging scalability; i.e., can the analysis be expanded to include all 77 VIVA libraries. Strong interest in using the data to inform prospective collection development. VIVA is the first group to use the SCS GreenGlass tool, possible because they have defined ‘scarcely-held’ thresholds within the group. Interesting mix of group action and controlled independent action. Project continues.

***Washington Research Library Consortium (WRLC):*** Nine libraries, 5.3 million bib records, and a 2 million-volume shared collection facility. Goal is de-duplication, and creating room in campus libraries by moving more to storage. Data work complete, scenario-building in process.

***Academic Libraries of Indiana (ALI):*** Thirty-four libraries, 5 million bib records. Project includes private academics and the smaller branches of the Indiana, Purdue, and Notre Dame systems. Also includes Indiana State, Valparaiso, and others. Data preparation work is still underway, but the group has also defined its threshold for scarcely-held titles (which will be retained), enabling participating libraries to use GreenGlass for local action on more widely-held titles.

**Products**

The primary SCS focus is on project success for the library or group. Because of that, we have configured our approach and services to be flexible, and adaptable to the objectives of the group we are serving. At the most basic level, we offer three kinds of services:

* *Data Preparation, Normalization, and Augmentation:* SCS spends a great deal of time and effort shaping each library or group’s data set. We believe that this up-front work greatly improves the consistency and reliability of results. This is vital in cases where withdrawals are contemplated, but it also provides workflow efficiencies for the libraries, as lists can be sorted accurately, OPAC links work consistently, etc. Starting with the bibliographic, item, and circulation data supplied by the library, we parse and normalize call numbers, dates, locations and other elements. We then augment the library’s data with holdings information from WorldCat, Hathi Trust, Internet Archive, CHOICE, and other sources. Curating this initial data set is the core of our service.
* *Custom Queries, Scenarios, and Reports:* The rich data sets created in collaboration with libraries offer many opportunities for analysis. SCS data analysts have great facility with SQL, and can draw on experience with other clients to help the group develop retention, storage, or withdrawal scenarios of many kinds. We have designed our services to allow sufficient time and bandwidth to run and compare many iterations of scenarios—to use the full power of the data to make informed collections decision. Examples of some reports are included in the Processes section below. More are available upon request.
* *GreenGlass:* GreenGlass is our web-based decision support tool, which gives an individual library the ability to interact with and visualize its own data. While GreenGlass is not currently set up for group-level scenario building, some groups have begun to use it in a controlled way, by flagging their scarcely-held titles for protection, and allowing individual libraries to act independently on widely-held titles. See the SCS homepage at [www.sustainablecollections.com](http://www.sustainablecollections.com) for a 2-minute video introduction to GreenGlass.
* *Seminars and Stakeholder Education:* Many libraries and groups have invited SCS to speak with library staff or teaching faculty about the changing role of local print book collections. SCS offers a 90-minute program called “Rethinking Library Resources” that we have presented dozens of times in the past five years. It is designed to explain why rethinking local print collections is a reasonable idea, and how data can be used to make decisions as safe and cost-effective as possible. This can be included as an option in any project.
* *Onsite Presentations and Facilitation:* Communication and decision-making in a group context can be challenging, especially when working with large, complex data sets. In a typical group project, SCS would participate in an onsite kick-off meeting and a return visit to present the first round of Collection Summary results. This is supplemented by conference calls and WebEx sessions, GreenGlass training, and in some cases, additional site visits to discuss alternatives.

**Process**

Our process typically begins with a planning call with each participating library, in which SCS explains the data extract we will need from each library. Alternatively, this can be done in a group-wide kick-off meeting. This discussion is based on a Data & Cataloging Questionnaire, through which we seek to understand the extent of circulation data available, call number schemes in use, codes for location, item type, item status, etc. Because local practice varies, and because multiple ILS systems are involved, additional calls may be needed to decide how to reconcile divergent data and practices.

Depending on the level of facility with data extracts, it typically requires a month or more before all data sets are received by SCS. There is usually some back and forth with some libraries as we parse and validate the data.

Once we have a usable data set from every library, our processing work largely depends on how large the record set and comparator groups are. For a project with 4.6 million bib records and 12 libraries, it requires six to eight weeks to create a clean, accurate Group Collection Summary. All of this is described in the chart that follows the narrative section of this response.

**Holdings Comparisons: General**

SCS routinely provides comparisons of group holdings in the following categories:

* Overlap among the group participants (based on the records submitted to SCS)
* WorldCat holdings at several levels:
  + Among US Libraries
  + Among Libraries in State/Province
  + Among specified comparator groups (see below)
* Hathi Trust Digital Library
  + Public Domain
  + In Copyright
* Internet Archive
* CHOICE
  + All titles reviewed in CHOICE
  + CHOICE Outstanding Academic Titles (OAT)

**Holdings Comparisons: Groups**

SCS can support up to five groups of comparator libraries, and a total of 100 holdings symbols. When comparing the group’s collection to an individual named library, that library constitutes a group. It is also necessary to obtain written permission from those named libraries in order to disclose their holdings as part of the analysis. For MI-SPI, comparator groups will be worked out during the planning process

SCS anticipates that we will remain part of the group’s work until the project has been completed, or the data needs to be refreshed. In some cases, this simply involves use of GreenGlass, but in other cases we continue to run reports and queries as needed over time. As MI-SPI knows from experience, SCS will do whatever is necessary to support the project to a successful conclusion—or to build an effective ongoing solution.

**Support from MI-SPI**

As noted above, SCS will need extracts of bibliographic, item, and circulation data from each library. Other libraries have reported that these extracts typically require no more than 2-3 days of time from a systems librarian. In some cases, the system vendor may need to make changes to export profiles, so some liaising would be needed there. At this stage of the project, it is necessary for SCS to work with each library individually, since systems and data practices vary widely across the group.

As the data extracts are being defined and prepared, at each institution, it will be important to have a cataloger, or someone who understands the institution’s data practices over time; i.e., what classification schemes are in use, what is the call number hierarchy, has a reclamation project been done recently? Similarly, someone from Access Services can inform us about in-house usage counts, treatment of historical circulation data during systems migrations, etc. And of course, someone from Collections to help with scoping, and decisions about comparator groups.

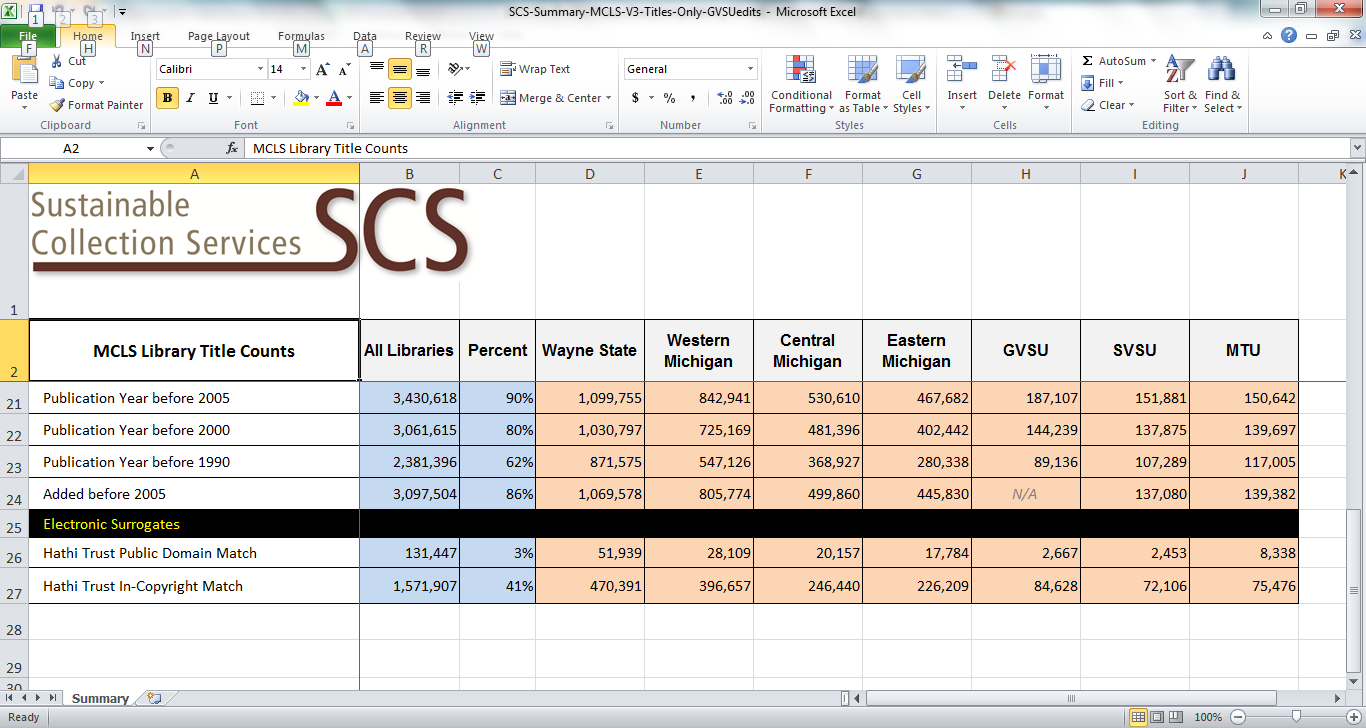
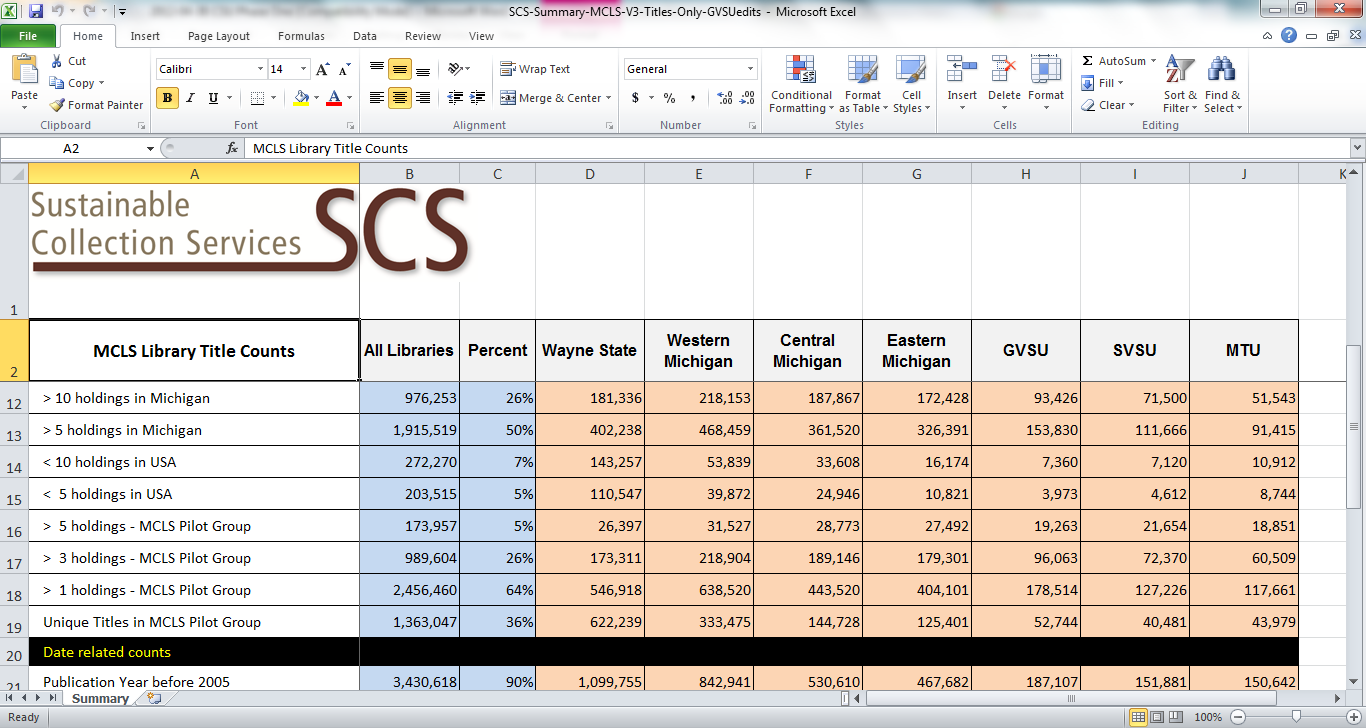
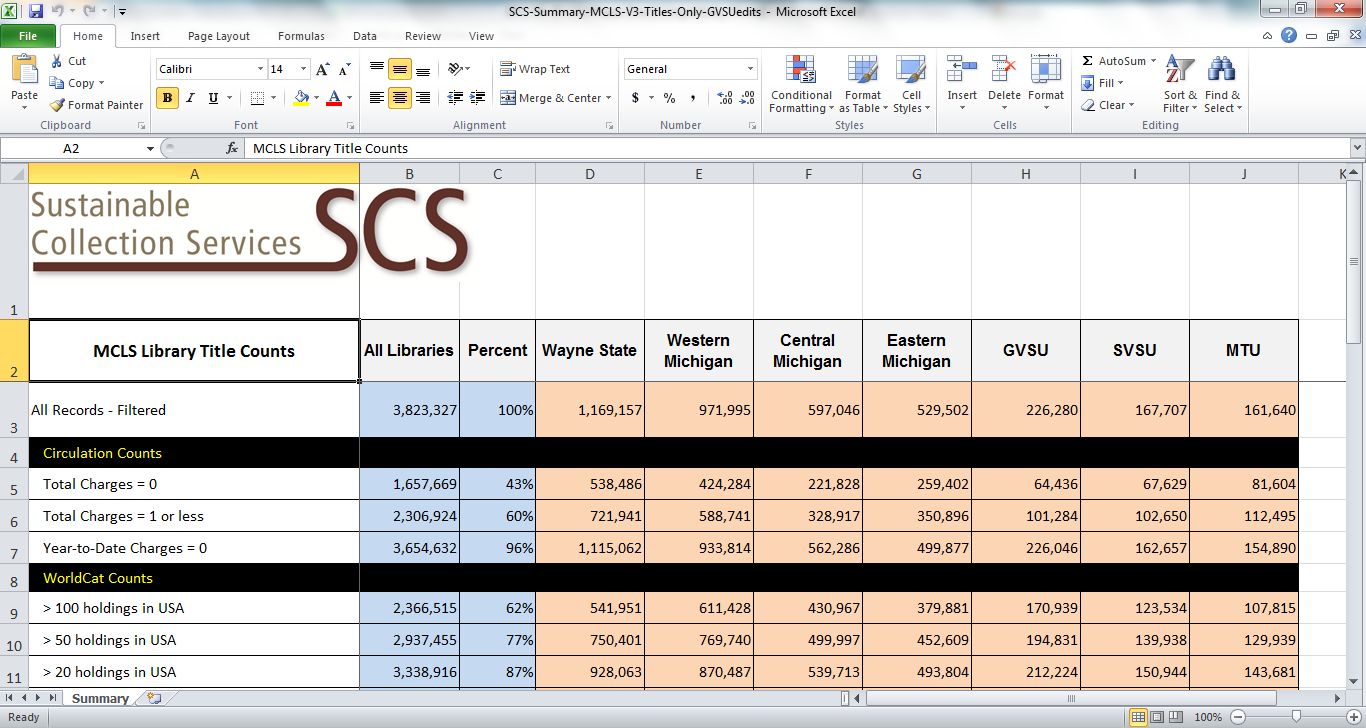
SCS will make every effort to use each library’s time carefully, but it is critical to get these details at the outset of a project. At the group level, we may need conversations about what to do if, for instance, if circulation data varies too widely across the group.

Finally, it can be immensely helpful in group projects to have a designated project manager, who can help with communication, facilitation, dissemination of information, and in general assuring that the group perspective be represented. This will be an important role at the outset of the project, when structural decisions are made, but also over time, as the group considers scenarios, allocation, data remediation, addition of other libraries, and works to resolve the inevitable mistakes or misunderstandings. An email listserv can be a valuable tool in this context if that can be provided.

**SCS Processes and Methods: Group Projects**

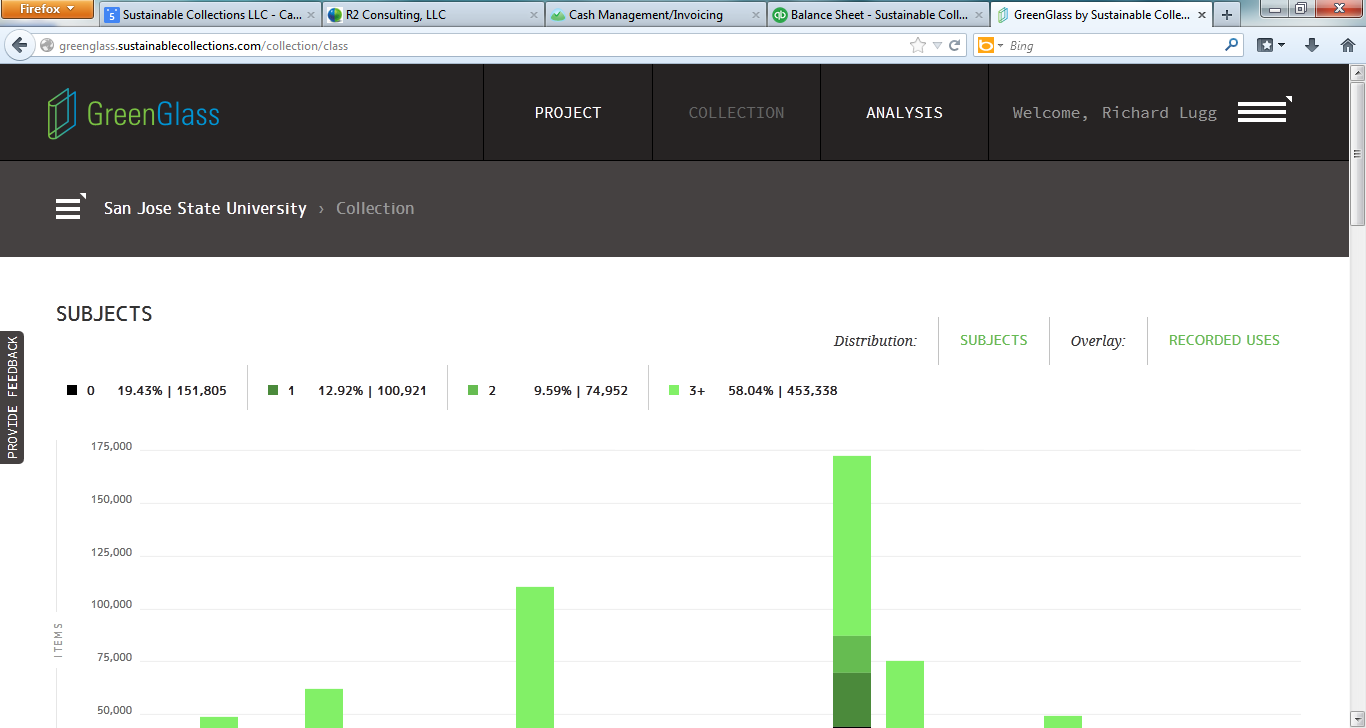
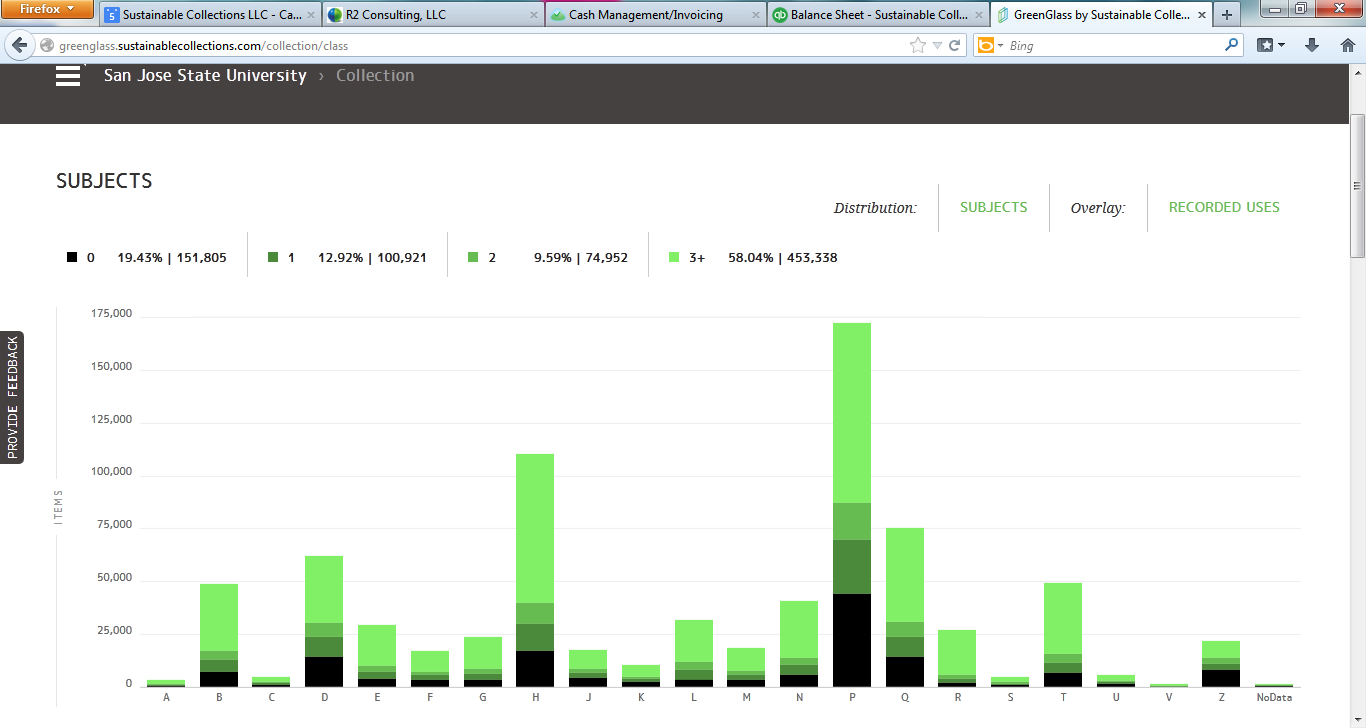
Shared print projects involve working with data from multiple libraries. The complexity of a project can vary, depending on the participants’ previous history of collaboration, the consistency of their respective data management practices, and other factors. This following table provides a high-level view of the main project components and SCS methods.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Task | Description | Data/Decisions  Needed | Timeframe | Output |
| Planning Meeting | Key players discuss data extracts, mappings, anomalies; define peers, title protection rules. | Mappings, peers,  Scope of extract | 1-2 days | Project outline Preliminary schedule |
| Data Preparation | SCS receives data extract, validates, transforms, normalizes, performs holdings lookups | Bib, item, circulation data | Depends on size. 6-8 weeks for 4.6 million records | Remediation lists; library & group-wide databases for querying |
| Group Collection  Summary | Categorical overview of the group data set. Used to gauge opportunities, guide scenario development. Can be further subdivided by subject, location, etc | Review of patterns, benchmarking | 1-2 weeks after group database is available | See example following this table. |
| Scenario Development | Project leaders suggest preliminary withdrawal, retention, and preservation criteria, drawing from data elements in the Collection Summary. SCS performs trial iterations, revisions as necessary. GreenGlass supports individual library queries and scenario modeling. | Criteria development, discussion of results, adjustments, repeat process. | Depends on number of iterations & availability of decision makers. 2-3 days per iteration for data work. | Finalized scenario descriptions; criteria for withdrawal & preservation. GreenGlass access. |
| Candidate Lists | Based on finalized criteria for withdrawal and preservation. Detailed Excel spreadsheets (sample in Appendix). Can also be produced via GreenGlass. | Requires final decisions on criteria. | 1 week for Withdrawal Candidate lists. 1 day per library for Preservation lists. | Group withdrawal candidate list.  Individual preservation candidate lists (for scarcely-held titles). |
| Task | **Description** | **Data/Decisions**  **Needed** | **Timeframe** | **Output** |
| Allocation | Responsibility for withdrawal and retention candidates must be divided equitably among participants. This allocation process incorporates many factors: space, distribution of copies, presence in the collection. Allocation can occur based on collection size, proportion of withdrawal candidates, or other factors. | Individual and group goals sometimes come into conflict here. Needed: a group-wide agreement on who benefits and who bears costs and in what proportions. | Unknown. These are difficult decisions, and can require numerous discussions and repeated trials of an allocation algorithm. | Allocation formula that is acceptable to all. |
| Discussion/Facilitation | Needed at many points, but especially around scenario development, allocation/retention, and related policy decision (last copies, retention responsibilities, etc) | SCS can help organize & facilitate some conversations by presenting alternatives drawn from the data | Varies, depending on process. | Documented decisions. |
| Picklist/Keeplist  Production | Once allocation decision has been made, SCS will derive title/item lists for use by individual libraries. | Criteria and allocation scenario finalized. | 1-2 weeks. | Two outputs: an annotated group withdrawal list & library-specific retention lists. |
| Ongoing Data  Management | SCS data sets can be used for 2 years from creation. SCS will maintain (but not update) library data during that period. Data can be used for additional projects. | None | 2 years from initial creation. | Data remains available for additional analyses. Pricing for additional work to be negotiated. |

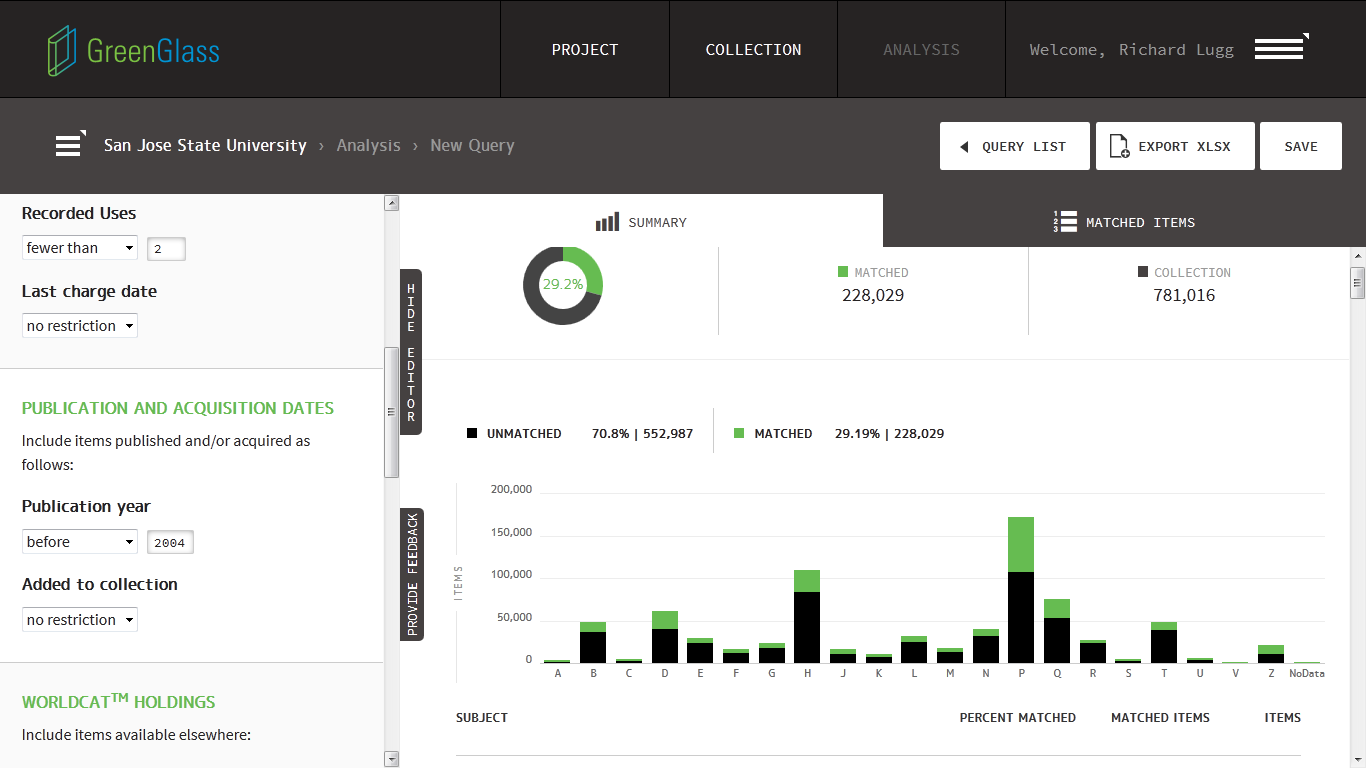


Example of a Group Collection Summary

Example: GreenGlass Collection Page with Recorded Uses Overlay

****

Example: Partial View of GreenGlass Query Builder



**Scope/Deliverables** [to be discussed & confirmed in planning stage]

This project will focus on circulating monographs and reference books in print form. Print journals, government documents, audio-visual, microforms, and other non-book materials are out of scope. Special Collections are not included. The Group estimates that SCS will load approximately 4.6 million bibliographic records, plus corresponding item and circulation data.

SCS is responsible for analyzing and presenting the data needed to make decisions and formulate policies. SCS will also facilitate discussions related to data analysis, interpretation, and policy options. Decisions and final policies are the responsibility of the libraries and the group. Specific deliverables include:

1. SCS will load and normalize bibliographic, circulation, and item data for 12 data sets and 12 libraries. During planning sessions, further details will be explored (e.g., inclusion of data from circulation transaction file, extent of in-house use data, and mapping of data elements). Harmonization of circulation data drawn from multiple automated library systems will also be part of this process.
2. SCS will offer via GreenGlass data remediation lists for corrections made in the course of normalization. This may include corrected control numbers, identification of records without holdings set, and possible duplicates (within an individual library collection).
3. SCS will compare group holdings to OCLC WorldCat, HathiTrust, Internet Archive and to one another.
4. SCS will provide aggregate comparison according to the comparator groups defined by the group.
5. SCS will compile all streams of collection, item, circulation, and external holdings data into a consolidated group database, which will be maintained on SCS servers.
6. In consultation with the group, SCS will design and execute queries against this group data set, to produce a Group Collection Summary, similar to the example shown above.
7. SCS will draw from the Summary data to analyze collection overlap by subject, date, circulation level, location and other factors supported by the data. Depending on the strategy adopted by the group, the libraries may be enabled to produce their own scenarios and lists in GreenGlass.
8. SCS will present the Summary-level analysis to group participants, either in person or via webinar. SCS will facilitate discussion of deselection and retention scenarios.
9. SCS will iterate queries and deselection scenarios as needed, in consultation with the libraries and the Project Coordinator. There may be many variations of group-level queries run, adjusted, and repeated. Each group-level scenario will include an estimate of its potential yield for retention commitments and preservation candidates. These iterations are the heart of SCS’s service, enabling libraries to interact with the data until they have evolved criteria that support their collection goals.
10. SCS will provide all participating libraries with access to GreenGlass, our web-based decision support tool. This will enable each library to model its own criteria and scenarios for withdrawal and retention for titles that are widely held (i.e., exceed the group-defined threshold for scarcely-held).
11. SCS will recommend specific areas of focus and will estimate the size of various opportunities, based on our experience with other academic libraries and with the data itself. SCS will identify issues and strategies for the group to consider, and will help to frame key decisions.
12. SCS will produce a single group-wide list of retention candidates based on criteria for scarcely-held items as defined by the group.
13. After negotiations with the group regarding allocation priorities, SCS will allocate retention commitments among the participating libraries, and produce appropriate lists for each library.
14. SCS will also produce library-specific lists of preservation candidates (titles held scarcely elsewhere in the US).
15. SCS will maintain the group data set (as created to prepare the Group Collection Summary) for up to two years. If additional analyses are wanted subsequent to the initial project, the same data set can be used.
16. SCS will work with MI-SPI to develop a data management and pricing model that will support ongoing collection analysis and print book management across the group.

**Cost Estimate**

SCS services and pricing are focused on a single goal: project success. It is not always possible to know at the outset what problems will be encountered or what new opportunities may be identified in the course of the work. SCS services and pricing are structured to assure that we can adapt as needed to changing circumstances in pursuit of the group’s objectives. Group projects in particular demand that flexibility. Our commitment is that SCS will do everything possible to help the Group attain its objectives for print book management.

Pricing for group data analysis is based on the following model:

$5,000 set-up fee per data set, plus:

$.03 per bibliographic record loaded to SCS FTP server

8-20% Group project charge

(includes data consolidation/harmonization, communication & facilitation, onsite meetings, record de-duplication where necessary, individual library and group-level tallies, multi-party scenario development, allocation of retention responsibilities & withdrawal opportunities across multiple libraries, and a range of other tasks not required for individual library projects.) The rate varies depending on project scale and complexity, and the likely number of site visits required.

For MI-SPI, these factors cost out as estimated here:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Institution Name** | **# of records** | **Set-up** | **Bibs @ $.03** | **8%**  **group charge** | **Total** |
| Central Michigan University | 558,759 | $5,000 | $16,763 | $1,741 | $23,504 |
| Eastern Michigan University | 550,268 | $5,000 | $16,508 | $1,721 | $23,229 |
| Ferris State University | 153,422 | $5,000 | $4,603 | $768 | $10,371 |
| Grand Valley State University | 292,904 | $5,000 | $8,787 | $1,103 | $14,890 |
| Lake Superior State University | 125,179 | $5,000 | $3,755 | $700 | $9,455 |
| Michigan-Dearborn, Univ of | 238,792 | $5,000 | $7,164 | $973 | $13,137 |
| Michigan Technological University | 174,538 | $5,000 | $5,236 | $819 | $11,055 |
| Northern Michigan University | 280,994 | $5,000 | $8,430 | $1,074 | $14,504 |
| Oakland University | 384,638 | $5,000 | $11,539 | $1,323 | $17,862 |
| Saginaw Valley State University | 169,899 | $5,000 | $5,097 | $808 | $10,905 |
| Wayne State University | 700,473 | $5,000 | $21,014 | $2,081 | $28,095 |
| Western Michigan University | 966,194 | $5,000 | $28,986 | $2,719 | $36,705 |
| **TOTAL** | **4,596,060** | **$60,000** | **$137,882** | **$15,830** | **$213,712** |

A few comments:

* Please consider this an estimate only. In our experience, record counts can change somewhat during the planning process (typically downward). We are happy to discuss further as requirements become clearer.
* We have estimated a group project charge of 8%--at the low end of our range. This is because the number of different ILS systems involved is small, because MCLS supports consolidated invoicing, and because of previous experience—SCS has a reasonable idea of the level of complexity involved. Nonetheless, additional work will be required to make the data usable at the group level, and additional communication and facilitation (including onsite meetings) will be necessary. The group project charge reflects those additional commitments.
* We need to verify which libraries have active subscriptions to CHOICE Reviews Online (CR)). Our data license with CHOICE requires a royalty payment of $550 for each library that does not have a current subscription.

**Terms:**

SCS projects are typically invoiced on the following schedule.

25% Due at signing

25% Due upon receipt of all data extracts

25% Due upon production of Group Collection Summary [adjusted for actual record counts]

25% Due when first candidate lists are produced and libraries are active in GreenGlass

**Schedule:**

SCS will accommodate the group’s preferences for scheduling to the highest degree possible.

Timing will be influenced by the SCS production schedule and group readiness, but a sample schedule might include these phases:

* August 31, 2015: Data extracts, documentation, and final comparator lists to SCS
* October 31, 2015: Collection Summaries to all participating libraries
* November/December 2015: Scenario development, List production, ongoing analysis
* November/December 2015: Library data available in GreenGlass (if wanted)

The Group data set will remain available for querying and analyzing for up to 24 months, after which we recommend it be refreshed and rebuilt to capture additional circulations, newly added titles, etc. The full refresh of the database would again incur set-up and per-record fees.