Report to HOTS from the Shared Cataloging Program (SCP)

November 1, 2004

Accomplishments FY 2003/2004

Electronic Journals

This year, SCP added access to 2,194 serial records. Not quite one-third (682) were from one of six new paid package subscriptions. This being the first year SCP has been directed to add free access links to selected titles, we did so for 326 separate items (69 Making of America serials had previously been distributed). During May and June, we managed our first large-scale deletion as our subscription to Computer Database was dropped by CDL (~300 titles affected). Otherwise, new titles were added steadily across-the-board to various packages.

Integrating Resources

The number of IRs cataloged nearly doubled as 154 new database records were distributed to the campuses. Reiterating our comment from our report last year, the integrating resources themselves represent an insignificant workload relative to journals and monographs. It is the potential for discrete, individual title access within a database that we tend to be concerned about since the SCP workload mushrooms accordingly. During this past year, for example, LION fell into this category as from this one database, we provided access to 1,375 separate monograph titles and have an additional 13,000 to complete.

Electronic Monographs

Eight new packages were processed. 8,646 electronic monograph records were distributed for materials in 14 packages, and two experienced no growth. The bulk (8242) came from one of six packages: SPIE (2014), IEEE Xplore (1814), ACM Digital Library (1389), LION (1375), Black Drama (906), SourceOECD (744). To facilitate local processing by recipient libraries, some records were distributed in distinct package-by-package sets. Note that SourceOECD is not an official CDL package, rather, since all the campuses have local subscriptions, CDL decided the cost benefits to all justified cataloging through the SCP.

California Documents

SCP provided access to 255 serials, 379 monographs, and 6 integrating resources during this report period. We noted that the publishing pattern of print and online to online only continues in force.

PIDs and BibPURLs

Over 40,000 PIDs and nearly 2,000 BibPURLs were being maintained at the end of the fiscal year. BibPURLs are created using the same technology as PIDs and their use is guided by the same principles. We adopted the use of BibPURLs in keeping with national level cataloging practices. BibPURLs are used for *free* access materials and can be added to the OCLC master record for national, and UC, distribution. In contrast, PIDs are created for UC *licensed* resources, and being UC specific, can not be entered into the OCLC master record.

Productivity/Statistics

Net production statistics since inception through:

	June 30, 2003	June 30, 2004	FY 2003/04 Net Increase
Serials	9,932	12,126	**2,194**
Monographs	15,315	23,961	**8,646**
Integrating Resources	160	314	**154**
CalDoc serials	513	768	**255**
CalDoc monographs	1,344	1,723	**379**
Cal Doc IRs	14	20	**6**
Total	**27,278**	**38,912**	**11,634**
PIDs	32,473	40,168	**7695**
BibPURLs	993	1,849	**856**
Total	**33,466**	**42,017**	**8,551**

FY 2003/2004 Production Details

	New Access 1	Modified Records 2	Withdrawn Access 3	Total Transactions
Serials	2,832	2,368	551	**5,751**
Monographs	10,851	1,074	29	**11,954**
Integrating Resources	157	86	1	**244**
CalDoc Serials	266	107	16	**389**
CalDoc monographs	403	33	19	**455**
CalDoc IRs	6	---	---	**6**
Total Transactions	**14,515**	**3,668**	**616**	**18,799**

1 New Access: all instances of adding an 856 link, either because a new title was cataloged, or a new link was added to previously cataloged title.

2 Modified Records: all instances of bibliographic record maintenance such as updating holdings data, processing title changes, correcting cataloging errors, etc

3 Withdrawn Access: all instances of the removal of an 856 link

Note that at this point in time, we are unable to provide detailed statistics on PID and BibPURL maintenance activity. We can state that we run the PID validation program on a weekly basis which generates a report on problematic URLs. Each reported URL is reviewed and corrections to the PID resolution tables or other related maintenance is performed as necessary. The number of problematic URLs reported varies from a couple hundred to a couple thousand per report, with most being towards the latter end of this spread. Staff spends approximately sixteen hours per week on this activity.

It has been just over a year since UCSD took over the **SCP record distribution process** from CDL. Generally speaking, this has been a successful changeover. Files of records are distributed weekly instead of monthly. Local control of the process has allowed us to plan and perform more quality control and to time the distribution of select records. This was done for some monographic packages, easing the file processing for receiving campuses. SCP has been, and will continue to be, flexible with special requests from campuses for record sets.

SCP staff volunteered to and developed a **Shared Cataloging Program flyer** that briefly describes the program and outlines the some of the benefits to the campuses: http://www.cdlib.org/inside/instruct/scp_spring2004.pdf

Staffing on the cataloging side remains stable, at least in terms of the individuals associated with the program. Organizationally though, data processing functions have increased necessitating significant increases in the amount of time required from our database management staff. The SCP Management Group and SCP Planning/Operations Group continue to meet and to provide administrative and operational direction. The Planning/Operations group specifically worked creatively to devise local monograph processing techniques that make the batch processing of large record sets practical. This group has also developed a list of quality control checks to improve existing records and bring older SCP records in line with current processing and cataloging conventions.

The **PID server** database was successfully transferred by CDL staff in Oakland to a failsafe environment. Over the past year, CDL and SCP staff met to discuss transitioning from the OCLC PURL resolution software to a combination of the SFX and ARK resolution services. While it was agreed the ARK service could potentially provide the same level of functionality as the OCLC software, CDL needs to write the programming before the migration can be attempted. Discussions on using the SFX service instead of PIDs led to agreement that a pilot test would need to be conducted so as to resolve outstanding technical questions and to create a prototype interface for review by public services and other stakeholders.

In response to a request from Becky Culbertson, the SCP recommended to HOTS that they write a letter to Dr. Kevin Starr, State Librarian of California, in support of the **California State Library switching from RLIN to OCLC** . The California State Library is the leading cataloging agency for California documents and, in light of support from groups such as HOTS, decided to switch from RLIN to OCLC as their cataloging utility in July, 2004. The impact of this change on SCP operations has yet to manifest itself; however, we anticipate faster turn-around times for CalDocs cataloging as we will no longer have to search for copy in RLIN and cut and paste records into OCLC.

While SCP has long taken advantage of UCSD's CONSER and OCLC membership to leverage its cataloging efforts for the greater library community through the creation of OCLC WorldCat collection sets, because of the quality of SCP cataloging, we have been put in the interesting position of potentially **selling records** . It is highly probable that ProQuest and CDL will negotiate terms for acquiring the enhanced versions of the LION monograph records we create. Additionally, SCP has received separate requests from two institutions asking us for record sets, one for our IEEE conference proceedings records, the other for our California documents records. While we have no infrastructure to support distribution of records to these two institutions, the possibilities are tantalizing.

Current Activities

Continuing activities include cataloging new titles, updating journal coverage data, deleting withdrawn titles, and PID and BibPURL maintenance. As predicted last year, there was a definite shift in emphasis for SCP from serials to monographs as SCP staff spent significant amounts of time in processing and planning for the distribution of monographic record sets. Following up with the previous distributed monograph record sets, we continue to send additions to those packages. SCP staff, with input from the SCP Advisory Committee, is undertaking various clean-up projects, such as replacing all 710 and 730 access hooks with standardized 793 title hooks. SCP staff is working through our second large scale deletion project, making the switch from ABI/Inform to EBSCO Business Source Premier. Access to over 3000 titles will eventually be added, deleted, or changed by September 2005.

Goals for 2004-05

- Process the LION, EEBO, and Early American Imprints packages

- Transition from ABI/Inform to EBSCO Business Source Premier

- Work with CDL on URL resolution service

- Conduct pilot test with CDL on SFX linking options

- Continue to improve communication from the SCP to HOTS and campuses

- Revise the "Proposals for new cataloging projects" document and re-examine the process

- Initiate overview analysis of the SCP's role and benefits within UC

Horizon Issues

UCSD Database Management staff continues to amaze with their various strategies for **the manipulation of large records sets** , especially for monographs, a continuing trend. Our experience with vendor records indicates that the quality and integrity of data remains varied. Much of our ability to manipulate record sets remains dependent on functionality specific to UCSD's Innovative Interfaces system. We will continue to rely on that functionality to batch process two large record sets, EEBO and Early American Imprints, which combined, will total nearly 200,000 records. Other large record sets are in the pipeline.

The **large scale withdrawal of CDL titles** caused by CDL's cancellation of the Computer Database package was successful for two reasons: first, at ~300 titles, it was a moderately sized project, and secondly, only one package of titles was involved. The switch from ABI/Inform to EBSCO Business Source Premier will be more challenging as there will be ten times as many titles involved and because, since subscription access to both packages will overlap through the end of the year, the switch needs to be done in stages. One stage, which is first priority and currently underway, consists of adding the EBSCO titles not found in ABI. The second stage consists of adding overlapping titles. The final stage will be deletion of ABI access. Because of this staged process of first adding and then deleting, several hundreds of these records will be distributed at least twice. For those campuses that do automatic overlay of records, this will likely prove inconsequential, however, those campuses that don't will find this process a significant workload.

Expanding use of the PID server to other campuses for use for local resources is being deferred until the URL resolution service issue has been settled. The option for each campus to independently use SCP's URL resolution service will be considered a critical component option of any service tendered by CDL for migration from the OCLC PID service. UCLA has successfully used this option for over two years now. It has proved invaluable to them for all the same reasons it has been invaluable for SCP—it is an effective, efficient, and economical means for performing URL maintenance.

The potential for continuing workload increase remains enormous and must be carefully managed. As noted above, CDL's Collection Development Committee's (CDC) call for UC bibliographers to identify freely accessed materials for "collecting" resulted in over 300 titles being added to our workload. When completed, addition of the EEBO and Early American Imprints will add nearly 200,000 titles to our catalogs. Cataloging of individual titles in Asian language packages remains suspended. Originally only a single package, (China Academic Journals (CAJ), ~1,800 titles), CDL has now acquired Sibu Congkan (~500 monographs). Additionally, we are aware that some bibliographers are interested in SCP cataloging for Siku Quanshu (~3,700 titles), SuperStar (~50,000 titles), National Diet Library (~30,000 titles), and JapanKnowledge (~600 titles). On another front, records for cartographic materials from the Rumsey Collection (~45,000 titles) and the Library of Congress American Memory Project (7,000+ titles) are being considered for SCP processing. And these are all dwarfed by CDL's acquisition of the Readex U.S. serials set (~325,000 titles). Fortunately, hopefully, some of these packages SCP will be able to handle via one time batch processing projects, but even for these, some will remain ongoing workloads as additional titles are added to the package. As just these enumerated materials total nearly half-a-million titles, obviously, the SCP will continue to need appropriate financial support and staff with the appropriate cataloging and language expertise to successfully complete its mission.

The question of **classification of electronic monographs** is approaching resolution with the recommendation from HOTS to SOPAG supporting classification. Assuming SOPAG also supports the idea, CDL will need to be approached for financial support. SCP will continue its efforts to find creative, cost-effective, and automated ways to implement consistent classification, but as a last resort may need to increase staff to handle the workload in a timely fashion. How systemwide priorities are set for funding various workloads—classification, cataloging licensed

resources used by smaller subsets of users, cataloging free resources—will need further input and discussion.

As of this report date, no decision on **a shared database/electronic resources management system** has been made. The impact of such a development on SCP is uncertain, but it is something that we are actively tracking.

We remain concerned about the potential impact of the **UC Shared Print Collection** . We plan to meet with Nancy Kushigian, Director, Shared Print Collection, to raise her awareness of the SCP, learn of her outlook on how the Shared Print Collection might be integrated into Melvyl and, potentially, other OPACS, and hear her ideas on SCP's role, if any, in this process.

Finally, in light of a changing environment, the fact that the SCP is now four years old, and the wisdom of continuous evaluation, an **analysis of the role and benefits of the SCP** seems in order. To this end, SCP staff plan to investigate at least three aspects of our current environment. First is a more detailed look at services offered by outside vendors (e.g., SFX, Serial Solutions). Might their services replace or augment current SCP services? Second, we will look at means to solicit greater feedback on needs and improvements to services rendered by the SCP to the campuses. One method we are exploring for gathering this feedback is a series of meetings by SCP staff at individual campuses. We would be there to explain in greater detail aspects of the SCP, learn first hand of the processes used by the campuses for processing SCP records, address issues raised, and bring back ideas and concerns for further action. The final aspect of our current environment we hope to explore is identification of additional value-added services that SCP could provide the campuses. Ideally, ideas would emerge from our campus visits and from the SCP Advisory Committee. But we seek to be proactive and are already looking at mechanisms for providing greater authority control over our records, and possibly, distributing authority records to the campuses.