

# Shared Cataloging Program - Reports to HOTS

Shared Cataloging Program (SCP) Annual Report to HOTS  
FY 2007/2008  
September 22, 2008

## Productivity

### Serials

For FY 2007/2008, the net increase in access points was 8,072 1 for serial titles, nearly doubling our production from FY 2006/2007. Packages for which we added one hundred links or more were EBSCO (4022), Lexis-Nexis (1657), America's Historical Newspapers (1028), Taiwan Electronic Periodicals (699), Taylor & Francis (201), China Academic Journals (192), CINAHL (134), European Intelligence Unit (128), and SAGE (125). The EBSCO count includes titles from several newly acquired packages including Academic Search Complete (3973). Along with the EBSCO packages, other packages licensed this year were the above listed Lexis-Nexis, America's Historical Newspapers, and Taiwan Electronic Periodicals as well as Bentham, CIAO, Karger, Synthesis, and University of Chicago Press. While not a new package, hiring a Chinese language cataloger allowed us to begin cataloging the China Academic Journals titles. We added 854 new open access links, the bulk for Directory of Open Access Journals, DOAJ (540). As a final note, with CDL's dropping of the Expanded Academic ASAP package in favor of the EBSCO suite of titles, staff completed the withdrawal of its 2,222 titles.

### Monographs

Once again the number of electronic monographs distributed was quite large with UC access to e-monographs increasing by 212,390 links. Two packages account for the vast majority of this increase: ECCO (135,876) and Making of the Modern World (58,719). Other significant additions were Naxos Music (4039), CRC Press (3352), SpringerLink (3566), Knovel (1310), SPIE Digital (690), National Bureau of Economic Research Working Papers (656), Materials Research Society (424), Alexander Street Press (234), ACM Digital (182), Safari (125), IEEE (111), SourceOECD (104), and Synthesis Digital (103). Naxos was new, both as a package and as a new format, digital sound files. Other new packages were Knovel, Materials Research Society, Synthesis, and Thieme. Net new open access monographs were 2426. Note, this year's additions push us over the 400,000 mark for electronic monograph access.

### Integrating Resources

Integrating resources (databases) increased by 44 access points. Due to past cataloging practices, we counted some integrating resources as monographs in past reports. Over the past year, SCP catalogers have been updating these records to current cataloging practices, so an undetermined number of these are due to clean-up.

### California Documents

Links increased by 65 serial and 514 monographic links from last year. There was no change in the net number of integrating resource links during this report period.

### Link Resolvers

Staff now maintains 130,844 PIDs, having created 15,622 over the year. Staff created 1,821 BibPURLs, bringing the total of BibPURLs maintained to 8,006.

## Review of 2007/2008 Goals

- *Keep current with cataloging and with serial and URL maintenance*  
For most of the year, staff kept current, however, significant increases mid-year to a variety of monographic packages (CRC press databases, Springer, Safari Tech Books, IEEE Xplore, ACM, SPIE), the acquisition of Lexis-Nexis access, and the transition from Expanded Academic ASAP to EBSCO's suite of products created backlogs that staff continue to work on. Prioritization by JSC and Ivy has helped staff allocate resources to the most needed of these acquisitions.
- *Begin the cataloging and record distribution of Chinese language resources*  
Our newly hired Chinese language cataloger began work on the Chinese Academic Journal titles in November and was able to distribute records for the series F portion of the package (192 titles). However, with the acquisition of the Taiwan Electronic Periodical Services database, JSC gave that package a higher priority. She completed TEPS at the end of reporting year (699 titles) and resumed work on CAJ.
- *Continue to improve communication with the campuses, with particular attention to status of work on current packages*  
New this year, SCP staff began reporting the actual number of records distributed weekly to each campus (<http://www.cdlib.org/inside/projects/scp/scpstats.html>). Staff continued reporting SCP news via monthly e-mail updates distributed to interested staff and posted at [SCP's website](#) and continued to track relevant package data on its [e-resources tracking page](#), also found at the SCP website.
- *Prepare a workload analysis methodology to measure current workloads, to estimate completion dates for packages that are underway, and to anticipate and plan for future work*  
Staff worked on analyzing and measuring its workload in response to a potential budget cut. As a result of the need for more cataloging prioritization guidance from JSC, SCP began preparing quarterly reports for the JSC.
- *Strategize about SCP workflows, roles, and responsibilities in a WorldCat Local environment*  
Staff participated on some WorldCat Local task groups, and assisted in the analysis and delivery of selected packages of SCP records to OCLC in preparation for the NGM pilot. This analysis helped OCLC perfect its matching algorithm for loading records for the Pilot. Once matched, staff developed processes for redistributing the SCP records to the individual campuses. This was critical for adding OCLC numbers to records for their retrieval via NGM.
- *Develop a process for the speedy resolution of complex link resolution questions*  
Over the past year, SCP and CDL Acquisition staff have fine-tuned processes for the use of SFX OpenURLs whenever possible, or the use of PIDs in cases where SFX openURLs are not available or their use would require an undue delay in the cataloging of resources. Additionally, staff proactively used BibPURLs for open access materials and, given their use by the California State Library, explored the use of Digital Archive Links (DAL). Investigation of the DALs showed that further discussions between the State Library, UC/CDL, and, perhaps, other potential partners (e.g. OCLC), on archiving options for California state publications are needed.
- *Incorporate the Lawrence Berkeley Lab into current record distribution work streams, if so directed by CDL, along with providing them with an appropriate retrospective file of records*  
While investigating licensing and other administrative matters for the Lab Library,

CDL determined that records for the Lab should be handled within the Berkeley campus.

- *Perform a retrospective cleanup for outstanding titles lacking targets in SFX*  
No specific project was undertaken, rather SCP and CDL Acquisitions staff performed clean-up as discovered. Work on this clean-up is sometimes dependent on SFX following through on reported gaps in the SFX KnowledgeBase.
- *Install a shared SCP "power station" dedicated to process-intensive operations such as global updates and file processing, and to release staff workstations for other use*  
Power work station installed.
- *Explore possibilities for creating additional WorldCat Collection sets*  
Various candidates for collection sets were identified over the past year, of which three were set up: Material Science Society Online Proceedings, Lecture Notes in Physics, and Lecture Notes in Mathematics.
- *Continue development and improvement of batch processing techniques*  
Staff created new OCLC Connexion and Millennium macros to speed cataloging of monographs. Staff developed an ISSN batch loading technique to aid in the cataloging of several serial packages.
- *Continue to monitor ERMS implementation plans and, if procured, leverage relevant system functionality to reduce redundancy and achieve greater efficiency*  
SCJP continues to monitor the ERMS situation and looks forward to finding ways to use such a system to improve workflows and access.

### Other Accomplishments

**CONSER/NACO Related Activities.** The UC CONSER Funnel entered its third year of operations with SCP staff continuing to contribute in significant ways to its ongoing success. Renee Chin carried on as the Funnel's Communications Coordinator, managing the [Funnel's Web presence](#) and e-mail list. This year, she completed a survey to assess the Funnel's communication needs. Based on her analysis of the results, she put forth various recommendations to encourage greater use of the Funnel's communication tools. Catalogers provided valuable feedback towards the ongoing development of the CONSER Standard Record guidelines. Adolfo R. Tarango, through his role as UCSD's CONSER representative, was able to present the SCP perspective directly at the annual CONSER Operations Committee meeting in May. Also, through his role as a CONSER serials cataloging trainer, at ALA Annual, he participated in a full day session review of CONSER training workshops, the CONSER standard record guidelines, and the basis serials cataloging modules. With regard to CONSER work, SCP catalogers created 19 original CONSER records and authenticated and converted 19 non-CONSER records into CONSER records. Additionally, catalogers re-authenticated 2 CONSER records, made enhancements to 777 CONSER records, and performed CONSER related work on 56 non-CONSER records. Adolfo, with Manuel Urrizola (UCR) and Melissa Beck (UCLA), taught the CONSER SCCTP Basic Serials Cataloging Workshop in San Diego. Adolfo assisted with the CONSER serials cataloging review for catalogers at UC Santa Barbara and UC Berkeley, both ongoing. For NACO, catalogers added 44 new and revised 14 headings in the national authority file.

**Personnel.** Bie-Hwa Ma joined the SCP on November 13th as our new Chinese language materials cataloger. She has summarized some concerns and impressions about cataloging Chinese electronic resources for SCP, which follow this report.

An analysis of **CDL's Resource Sharing Fund** revealed that the SCP budget had exceeded their allocation for several years, with the overage being paid by CDL funds. SCP staff did an extensive review of cataloging priorities and workloads to aid in determining staffing and funding options. Ultimately \$48,000 was cut from the SCP budget for fiscal year 2008/2009. Since the SCJP budget is entirely staff costs, this meant staff reductions. Several staff voluntarily reduced their time, and one staff member moved halftime to the UCSD Libraries payroll. Because of the loss of a halftime staff position, SCP and CDL determined that, beginning July 1, 2008, SCP could no longer catalog California documents. Several plans have emerged from the UC campuses and SCP to identify alternative ways to provide this data.

### **Goals for 2008-2009**

- Continue to reinvent SCP through the ongoing reengineering of workflows and development of batch processing techniques
- Monitor and study the new California documents processing strategy and explore similar methods for creating a broader, virtual SCP
- Working with appropriate groups, contribute to the development of a long-term, stable economic foundation for SCP
- Keep current with cataloging and with serial and URL maintenance
- If NGM is implemented, resolve whether single records for serials will continue to be SCP policy and accordingly, set holdings in WorldCat for all SCP materials possible. Initiate discussion among the campuses as to whether distribution of records is still necessary. If the need remains, explore alternative record distribution methods
- Explore possibilities for creating additional WorldCat Collection sets
- Continue to monitor ERMS implementation plans and, if procured, leverage relevant system functionality to reduce redundancy and achieve greater efficiency
- Explore options for providing greater authority control and doing more authority control on SCP records

### **Horizon Issues**

Addressing the impact of the **SCP budget cut** will figure prominently in the coming year. We will continue to exploit new techniques to gain greater efficiencies in the processing and cataloging of resources. In addition, staff will continue to seek out the availability of vendor records or vendor data to generate records. Use of the latter, however, is compounded by the quality and nature of vendor records and data (see Chinese materials discussion below) and by the unknown outcomes of the NGM pilot. Given that a key component for NGM functionality is presence of an OCLC number in local records, securing permissions for the loading of vendor records into OCLC becomes critical. OCLC is working with various vendors to negotiate the loading of their records into OCLC, but lacking such agreements, the UCs will either have to do without NGM access to these sets of resources, or SCP staff will need to do the work themselves. This does not necessarily mean the manual cataloging of each resource. Policy decisions made might allow SCP the option of setting holdings in OCLC only, without distribution of records to the campuses.

Staff continue to look favorably upon and are excited at the potential changes that might be brought about through execution of the **NGM pilot**. If successful, this might lead us to move the record cataloging upstream so that all (or most) of our cataloging is done in OCLC, rather than locally. Depending on the Pilot results, staff may need to significantly alter their workflows, take on new roles and responsibilities, drop others, or the SCP may morph into something else entirely.

**Substantial change is coming.** The unknown impact of the NGM pilot, the unknown linking practices for “local URLs” and local data within that pilot, the purchase of a UC ERMS, development of a “provider-neutral” record for e-monographs, and other fast-moving efforts combine to form a rapidly changing environment. SCP needs to position itself to be flexible and responsive, and must remain mindful of its mission to efficiently provide bibliographic access to electronic resources for UC users.

We have gotten a start on cataloging our **Chinese language materials**, however a variety of challenges have arisen, all related to our less than optimal attempts to use vendor records and data. Appended below is a full bulleted report by Bie-Hwa Ma, but problems encountered range from inaccurate and incomplete title lists, to the use of different Romanization and encoding standards (Unicode vs. MARC8). As we strive to provide access to the large numbers of materials we have licensed, such ongoing difficulties will slow the distribution of records. SCP staff was able to take advantage of an April UCSD site visit by East View and TTKN representatives to discuss these issues.

Submitted on September 22, 2008 by Adolfo R. Tarango

**SCP Web site:** <http://www.cdlib.org/inside/projects/scp/>

## **Issues of Cataloging CAJ and TEPS**

### **Metadata or Image Issues**

- Bad quality of title list/MARC records that necessitates lots of manual verification, revision, and manipulation
- Publication surrogates are far from being clear and comprehensive: There are no title page or cover surrogates for each issue. Often times the cover surrogate, the only image besides the content, is missing
- Inaccurate or missing ISSNs, titles, first issue and publishing dates, holdings information, etc., e.g., 10% of the ISSNs provided in TEPS are not in sync with those in the ISSN Portal. Around 40% of the titles in TEPS have incorrect information for first issue and publishing date
- Incorrect citation due to hidden title changes practice, e.g. all the earlier titles in CAJ/CJP are mounted on the same Web page and all the articles from earlier titles are cited under the most recent title
- Inconsistency in metadata, inconsistent information within the title list, between the title list and the web site that hosts the content, etc.

## **Different Standards**

- MARC Format: Good mapping tools between CMARC, CNMARC, and MARC21 are not handy, and with existing tools, the mapping from other MARC formats results in a certain degree of data loss. Actually, the MARC records are not readable at times
- Cataloging languages: Chinese instead of English
- Cataloging rules
  - Different ways of descriptive cataloging
  - Different rules for choosing access points
    - Title proper, ways of treating title changes: to comply with AACR2, need to manually combine or split vendor records then determine appropriate URLs
    - Imprint information and issuing body are mostly based on latest issue, etc.
  - Different holding descriptions
  - Not following ISBD
- Romanization standard: different pinyin guidelines among LC, ISSN centers, OpenURL providers, libraries, and publishers in China and Taiwan
- Subject cataloging:
  - Provide no subject headings or
  - Provide headings according to different subject heading lists and classification schemes
- Unicode vs. MARC8
- No standards are followed among vendors/publishers/aggregators

## **Little connection or weak relationship among service providers, librarians, publishers, and aggregators across countries**

### **Service/Utilities for Chinese language electronic resources in North America**

- A lot lower hit rate in OCLC: TEPS has 39.9% (279 out of 699) and CAJ Series F (the best one among the packages) around 50% hit rate
- Other language and hybrid records in OCLC slow down the speed of identifying and updating records
- The ISSN International Center has no Chinese scripts in the database and applies a different Romanization standard
- It appears that Ex Libris uses a machine-transliterated Romanization method to build its KB title entries, which may be the underlying source of various problems with their KB title data like a lack of quality control over the word division and other script inconsistency, e.g., 145 out of 515 titles in CAJ, Series F had wrongly Romanized titles

## Suggestion

- Request vendors to:
  - Verify metadata provided, if possible, against ISSN centers and national catalogs data
  - Identify title changes, clearly laying out journal history and/or relationships, building hyperlinks between related titles
  - Digitize cover-to-cover: covers, title pages, table of contents, commercial advertisements, etc.
  - Provide precise holdings data, indicating gaps
  - Post regular and frequent updates and alerts regarding new, withdrawn, or ceased titles, as well as coverage changes
- Information science training for publishers on providing quality metadata
- Encourage publishers/vendors from Asia to attend international conferences (ALA, CEAL, NASIG, etc.) to:
  - Obtain the knowledge of user needs and trends in library and information management
  - Create opportunities for closer communication with librarians and other publishers/vendors
- International standards and cataloging tools
  - Desirable for all publishers to use one metadata standard (ONIX, etc.), enabling easier transfer/exchange of records between vendors and libraries
  - Urge ISSN International Center to include Chinese scripts in their database
  - The construction and application of internationally standardized MARC, cataloging rules, and pinyin Romanization
  - The establishment of subject authority files across languages
  - Mapping tools between different standards, classification schedules, and subject headings

### Productivity/Statistics

<b>Link statistics since inception through:</b>			
	<b>June 30, 2007</b>	<b>June 30, 2008</b>	<b>FY 2007/08 Net Increase</b>
Serials	23,012	31,084	<b>8,072</b>
Monographs	194,946	407,336	<b>212,390</b>
Integrating Resources	507	551	<b>44</b>
CalDoc serials	1,082	1,147	<b>65</b>
CalDoc monographs	3,369	3,883	<b>514</b>
Cal Doc IRs	32	32	<b>---</b>
<b>Total</b>	<b>222,948</b>	<b>444,033</b>	<b>221,085</b>

PIDs	115,222	130,844	<b>15,622</b>
BibPURLs	6,185	8,006	<b>1,821</b>
<b>Total</b>	<b>121,407</b>	<b>138,850</b>	<b>17,443</b>

<b>FY 2007/2008 Production Transactions Details</b>				
	<b>New Access<sup>1</sup></b>	<b>Modified Records<sup>2</sup></b>	<b>Withdrawn Access<sup>3</sup></b>	<b>Total Transactions</b>
Serials	6,773	5,676	2,375	<b>14,824</b>
Monographs	264,378	10,021	826	<b>275,225</b>
Integrating Resources	35	909	19	<b>963</b>
CalDoc Serials	62	41	1	<b>104</b>
CalDoc monographs	544	32	10	<b>586</b>
CalDoc IRs	---	2	---	<b>2</b>
<b>Total Transactions</b>	<b>271,792</b>	<b>16,681</b>	<b>3,231</b>	<b>291,704</b>

<sup>1</sup> 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.

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

<sup>3</sup> Withdrawn Access: all instances of the removal of an 856 link



**SCP Record Distribution Statistics (January-June 2008 only)**

Campus	Monos	Serials	Special Distributions						Totals
			AIP	MIT CogSci	Oxford Ref	MOME	LION	EAI	
UCB	5,710	10,331	346	380	3	58,332	13,950	32,316	<b>121,368</b>
UCD	8,002	10,546	346	380	---	---	13,950	32,316	<b>65,540</b>
UCI	9,075	10,571	346	380	3	58,332	13,950	32,316	<b>124,973</b>
UCLA	5,725	10,569	346	380	3	58,332	13,950	32,316	<b>121,621</b>
UCM	5,643	10,219	346	380	3	---	13,950	32,316	<b>62,857</b>
UCR	5,030	10,241	346	380	3	---	13,950	32,316	<b>62,266</b>
UCSB	9,065	10,210	346	380	3	---	13,950	32,316	<b>66,270</b>
UCSC	5,269	9,949	346	380	3	58,332	13,950	32,316	<b>120,545</b>
UCSD	5,726	10,571	346	380	3	58,332	13,950	32,316	<b>121,624</b>
UCSF	4,511	10,094	346	380	---	---	13,950	32,316	<b>61,597</b>

**SFX KB Maintenance**

	Lack Object	Lack Target	Lack Portfolio	Need Activation	Coverage Updates	Other Maintenance	Items Reported
Totals	93	68	6	189	113	37	506