

SERIAL AGGREGATIONS—MULTIPLE VERSIONS AND THE VIRTUAL UNION
CATALOG:
THE EX LIBRIS EXPERIENCE

Michael Kaplan
Director of Product Management
Ex Libris (USA) Inc.

In two centuries we have come a long way in the construction of our bibliographic catalogs. We have evolved from the book style catalogs of a former century that were also made famous in this century by the enormous catalogs such as those of the British Library and the US National Union Catalog—surely the book catalog/union list to end all catalogs—to card catalogs that comprised 1000s of drawers at large institutions to microfilm or microfiche and thence to microfiche and then to online OPACs. While I can say that all in a single list of nouns with a grammatically syndetic structure (that is to say, an ‘and-ed’ string), the fact is that this small list—book, card, film, fiche, computer output microform, and online—represents a credible chronological period. Call it 150 to 200 years, give or take. Another way to measure it is to say that it encompasses about 30% of the time that has elapsed since the invention of the printing press with moveable type.

And this only addresses the issue of the form or format of the individual catalog records and the catalog itself. I’d like you to view the catalog records as analyzed pieces of data, discrete but interconnected, and the catalog proper as the container for the actual catalog records. The individual records linked together in what we used regularly to call a ‘syndetic’ structure are what make a catalog out of discrete bits of data.

On this level we have created the actual standards and protocols of the catalog records that really are the infrastructure of the catalog itself. These are the standards upon which individual records interconnect and hold together as a unified, coherent whole. These codes, both the historical codes and the present-day codes, are the standards and protocols of the catalog just as DNA is the protocol of the genes that make up all living beings. Our catalog code mutations are deliberate and are subject to national and international standards bodies that give them their seal of approval only after due consideration.

It is worth observing that the catalog formats and the catalog codes with which we are *most* familiar were developed in an overlapping fashion. Most of us would probably consider the late 19th and the 20th centuries as the period that comprises the age of substantial catalog form development and the time when bibliographic control or, at any rate, the age when meaningful bibliographic cataloging codes, developed.

What I want to talk to you about today is a new concept of the catalog, one particularly well suited for consortia, but also with implications for the actual further development and refinement of IFLA’s Functional Requirements for the Bibliographic Record, fondly known to its aficionados as FRBR, or ‘furbur’. It also has significant implications on how individual libraries and library catalogs can relate to loads of vendor-supplied data, particularly for aggregations, and to that data’s regular ongoing maintenance.

How well do you remember the old-style book catalog? It was very one-dimensional and extraordinarily static. The National Union Catalog and its regular cumulations were like that. There tended to be but a single access point, and that was it. With card catalogs we developed the ability to photocopy multiple cards, create additional access points, and maintain a multiplicity of complementary files. Those cards offered the first meaningful ability to cross-link and trace from one card to another, and from one alphabetical arrangement in the catalog to another.

In November 2000 I delivered an earlier, largely theoretical manifestation of this paper at the Library of Congress Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web. At that time I was concerned primarily with metadata from different schemata and their resolution, but also with their presentation. Today I want to talk primarily about contextual, on-the-fly presentation of catalog data in a record structure that is (present tense) both virtual and dynamic. In fact, it is the virtuality that makes for the dynamic record that I have under consideration today.

Libraries have long made use of union catalogs. For the most part, those union catalogs were physical union catalogs. I have already mentioned the National Union Catalog as one example, that of a printed, cumulated catalog. In the electronic era it works differently—namely, a pool of candidate records is assembled electronically, the records themselves are examined programmatically and are determined to be identical or not, as the case may be. The general theory behind making a determination of like-identity or non-like-identity for batch processes of loading data is well known and has been in use for well over a decade. OCLC, the University of California Melvyl catalog, and Harvard University's HOLLIS catalog all use flavors of a similar algorithm that has built into it a threshold of weights with positive and negative valences where positive weight is added to matching fields and negative weight is subtracted for non-matching fields. More recently, the German KOBV consortium—an ALEPH500 consortium—has published an updated variation based on more recent studies of matching algorithms.

During the 2000 LC Conference I spoke of the dynamic record as embodying a record concept that can be constructed on the fly from different data elements, both traditional bibliographic and non-traditional, including enriched table of contents information, abstracts, reviews, cover jacket illustrations, etc. Today I have something slightly different in mind, something that is more tightly tied to traditional bibliographic content and is particularly well suited to a consortial environment. I am particularly pleased to be able to show you examples of such a catalog construct because, while it was my personal fantasy at the time of the 2000 LC Conference, today it is rapidly approaching reality at the California Digital Library where CDL's Ex Libris incarnation of Melvyl and the State University of New York SUNYConnect Union Catalog are making use of our new developments. Its applicability to aggregation sets is also ready to demonstrate based on discussions Ex Libris is having with Ebsco Publishing regarding the MARC data they make available and update on a monthly basis for their *Ebscohost*® aggregation.

Let me discuss *Ebscohost*® first, then circle back to the CDL implementation.

Librarians have long agonized over the challenge of the so-called ‘multiple version’ problem. Succinctly put, the problem is how to catalog, associate, and present related titles or related manifestations of a given work. For example, in the days of card catalogs, microform reproductions were frequently cataloged on the same card set as a paper original by use of the ‘dashed-on’ methodology. For the cataloger this means of cataloging a subsequent reproduction (or a later edition, or a supplement or index) was a quick and dirty means of providing access. For the patrons this methodology offered an intuitive collocation of ‘equivalent’ works in different formats or works related to one another in a logical sequence.

The famous Airlie House meeting on multiple versions (or Mulver) was held in 1989. While it led to a period of intense discussion on the advantages and disadvantages of such an arrangement in an era of electronic access, and while various bibliographic tiers were proposed, nothing official resulted from it. The topic has come to life once again in the late 1990s with intense debate on the treatment of electronic serials and serial aggregations. Should libraries standardize on separate, discrete records for the printed version of serials and their electronic counterparts, or encourage a single record for both?

Libraries have a pressing need to develop a cradle-to-grave approach to handling electronic sets. That means obtaining the corresponding electronic records from the aggregator and receiving regular maintenance updates to the sets. Maintenance would ideally include additions, deletions, changes in coverage, etc. But what should libraries do with the records they receive from the creators of these aggregations? Should they standardize on separate, discrete records for the printed versions of serials and their electronic counterparts, or encourage a single record for both? Expediency (also known as Efficiency) has been the name of the game to date, and that has largely meant that libraries tack a holding location for the electronic manifestation of the serial to the record that otherwise serves to describe the paper original. Indeed, the CONSER program has specifically given its approval to this approach. However, the appearance of large and increasingly costly aggregations of electronic serials that defy easy control has raised the level of interest in new approaches. The Program for Cooperative cataloging has released guidelines for aggregators to create distinct subsets of records for their electronic journals, and those subsets may or may not be based on the record for the paper journal.

HOLLIS# AAP3195 /ser	
TITLE	The American historical review.
PUB. INFO	v. 1- Oct. 1895- [Washington, etc.] American Historical Association [etc.]
DESCRIPTION	v. maps. 27 cm.
LINKING NOTES	Superseded in part by: Recently published articles, ISSN 0145-5311, formerly issued as part of the American historical review.
SUBJECTS	*S1 History--Periodicals. *S2 United States--History--Periodicals.
MED. SUBJECTS	*M1 History--periodicals.
LOCATION	<p>Andover-Harv. Theol: Mfilm. Period. 65 Microfilm. Ann Arbor, MI: University Microfilms International, [19---]. - microfilm reels; 4 in., 35 mm. Microfilm Also: Period. 65 Library has: Current issues of original printed edition: R.R. Library currently subscribes to this title. Enter DISPLAY H1 for more information on this holding</p> <p>Andover-Harv. Theol: Period. 65 Current Issues: R.R. Also microfilm: Mfilm. Period. 65 Library currently subscribes to this title. Enter DISPLAY C2 for circulation information Enter DISPLAY H2 for more information on this holding</p> <p>Countway Medicine: Serial Enter DISPLAY H16 for more information on this holding</p> <p>Hilles: Periodicals Library currently subscribes to this title. Enter DISPLAY C9 for circulation information Enter DISPLAY H9 for more information on this holding</p> <p>Master Microforms: Film Mas 15187 Microfilm. Cambridge, Mass.: Harvard University Library Microreproduction Service. microfilm reels; 35 mm. Enter DISPLAY H19 for more information on this holding</p> <p>Networked Resource: E171.A57 Access restricted to users with a valid Harvard ID. To access: URL is http://nrs.harvard.edu/nrs/hul/eresource/anhistre The latest 5 years are not available in this electronic version. Consult individual Harvard libraries for more recent issues available in paper copy only Enter DISPLAY H17 for more information on this holding</p>

Figure 1: Illustration from the Harvard University HOLLIS catalog of a Multiple-version style record

Now, as fond as I am in a theoretical sense of the Mulver or one-record approach also to electronic journals, I have to admit that for libraries and vendors generally I am troubled about enabling the aggregator's data file to reach deep down into our integrated library systems and to touch data on this level if the data is buried within a Mulver-style record. Ex Libris may well be able to do it, but for most libraries and vendors it is simply fraught with too many difficulties. In fact, in preliminary discussions between Ebsco Publishing and Ex Libris we have concluded that Ex Libris has advanced utilities in place that will allow us easily to rip out and replace the existing *Ebscohost*® aggregation, whether maintained as discrete records or combined with the paper records, and replace them with the updated file on a regular basis. So I reluctantly conclude that, if we want to look to a hands-off, computer-to-computer data interchange, we need to keep the basic building blocks of that exchange as simple as possible. From the perspective of a former Director of Technical Services in one of the so-called Big Heads libraries, I have to admit that managing large serial aggregations is an impossibility at the local level without firm and decisive actions from the aggregators and help from the ILS vendors. I am desperate for an EDI-type solution for all aspects of aggregations that drills right down into the local catalog to solve this problem that I see as growing increasingly intractable.

So let me add a specific recommendation to this complaint. I have concluded that a partnered solution, for most ILS systems will mean keeping the electronic journal records separate from their paper counterparts. That way an incoming record, particularly an incoming maintenance-level record, can be programmed to behave in predictable ways vis-à-vis the existing database record. I believe the solution that best serves technical services and that provides the best, easiest, and most-up-to-date access will ultimately prove the best public services solution. Besides, it might prove possible to then merge records for the public view that are kept separate in the technical services components of our catalogs.

That aside, there is one further aspect to all this, too, and that is user preference and user understanding. In general, I have maintained that a Mulver record best serves users provided that the display can be made clear and sensible. Users do not want to see a multiplicity of records for what to them is essentially the same title. Never mind the fact that the paper and the electronic may not be 100% the same for technical reasons, or that the electronic record might be suffer from a moving wall such as JSTOR maintains. The simple fact is that users want to see one record and one record only. This also means that all the information that is pertinent only to the reproduction or the electronic journal needs to be clearly associated with the reproduction or the electronic version. As a user in a long ago graduate student lifetime, I will say that finding all the relevant records, regardless of format, in a single, intelligible bibliographic construct would have been the arrangement I would have found most appealing.

Now what of the consortial arrangements? Consortia want to present not just a single institution's holdings but also those of a group of affiliated libraries. They need means to match, deduplicate, and merge holdings from a variety of libraries. As I said earlier, the algorithmic means to detect duplication are well known and have been refined in recent years. The match runs in the first instance on key numbers such as OCLC or RLIN numbers, or Library of Congress control numbers, or ISBNs or ISSNs. Where these numbers cannot be guaranteed to be unique, secondary checks on authors, titles, imprints, etc., are employed. And the matching algorithms are capable of being manipulated and fine-tuned to account for differences in

different bibliographic formats—such as monographs vs. serials—and the threshold associated with them can vary and be adjusted based on experience. CDL, the California Digital Library, has just such an algorithm used by Ex Libris in the construction of the new Melville catalog.

Once that comes into play, there is the question of selecting a preferred record, or creating a composite record from the pool of duplicate records. At this point, rather than talk further, I want to show you a sample CDL record, that for the SIAM journal on matrix.

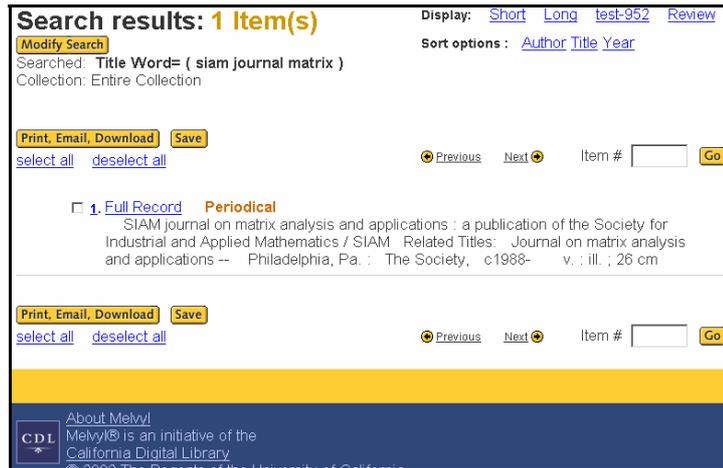


Figure 2: CDL search result for “Siam Journal Matrix”

In this first screen capture we can see the result of a search for “Siam Journal Matrix”. It brings up a single hit. The following screen capture shows a view of the so-called global record that is an amalgam of the CDL records, with underlined holdings representing the holdings and individual records for each of the owning campuses.

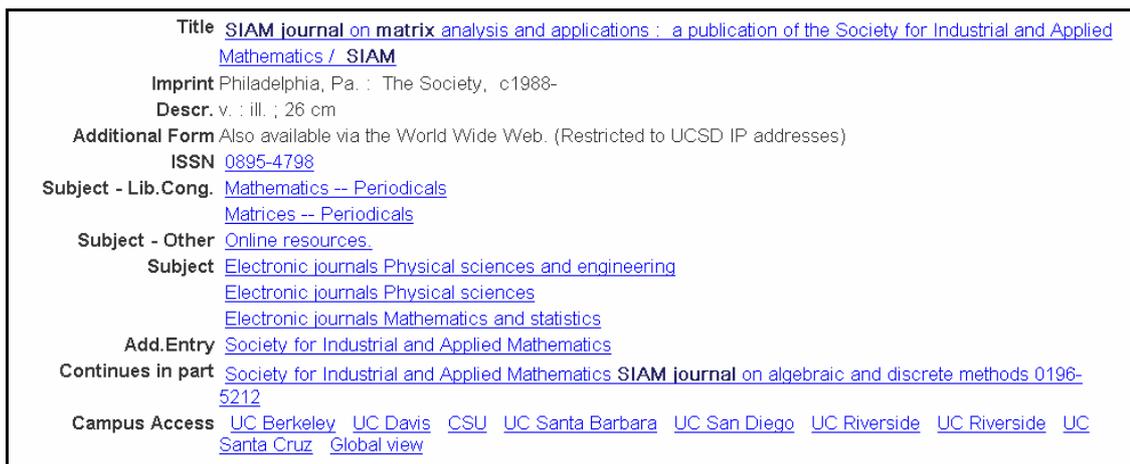


Figure 3: Global record for CDL/Melvyl and the owning campuses

No. Regional Library Facility			
	QA1 S732	Circ status	Volumes relocated to NRLF; inquire at Circulation Desk v 9-10,1988-1989
UC Berkeley			
		Available online	Vol. 22 (1998)- available online at:
		Available online	Vol. 15 (1994)- available online at:
Astr/Math	QA1 .S732a	Circ status	Unbound issues on Display Shelves - Alphabetically 9 (1988)-
UC Davis			
		Available online	Full text: v.18 (1997)-; Abstracts: v.15 (1994)- Access restricted to UCSD IP addresses
Internet		Circ status	v.18, 1997- Electronic access
Shields	QA1 S732	Circ status	Recent issues in SHIELDS Bio/Ag Current Per; older issues bound in book stacks. v.11,1990- Active commitment
UC Riverside			
		Available online	Connect to Abstracts only: v.15 (1994)-v.17 (1996); Full text: v.18 (1997)- UCR access only
		Available online	Connect to Abstracts only: v.15 (1994)-v.17 (1996); Full text: v.18 (1997)- UCR access only
Internet	Electronic journal	Circ status	Abstracts only: v.15 (1994)-17(1996); Full text: v.18(1997)-
Science	QA1 S6252	Circ status	v.9(1988)-21(1999/00)- Latest in Math Dept.; older in stacks
UC San Diego			
		Available online	v.18(1997)- Restricted to UCSD IP addresses
		Circ status	18:1- (1997-) INTERNET
S & E	QA 1 S53	Circ status	Electronic journal 9- (1988-) Current Journals Bound in Stacks
UC Santa Barbara			
		Available online	
Main Lib	QA1 .S524	Circ status	v.9 (1988)-v.21 (1999/2000) Sci-Eng
UC Santa Cruz			
		Available online	1994- Access restrictions may apply
Science	QA1 S732a	Circ status	9(1988)-
Internet	QA1 S732a	Circ status	Back issues: Stacks.
Previous Next			
			

Figure 4: Partial view of the holdings for the UC libraries

In these views of the record we can see the composite record for the paper and the electronic journal and we can also see that it is held at multiple campuses throughout the University of California system. The record shows, as you can see, holdings for both paper and electronic versions, with partial holdings information. In CDL's incarnation of its ALEPH500-based catalog, it is further possible to go to the holding for any individual campus or any individual manifestation since the original records are preserved intact and the merging is done on the fly on the basis of pre-built tables of equivalencies that were part of the Ex Libris-CDL development for the new Melvyl.

What does this mean for the library and for the end user?

For the library it is really a case of, as the English phrase goes, of having the best of both worlds. Records that are not really intended to represent the same title—at least not until we get to a fully-fledged FRBR environment—are kept apart. They do not compromise one another's bibliographic integrity or existence. It is always possible to get back to first principals or to the original description. Yet, on the other hand, titles that a library or a consortium of libraries really wants to treat as equivalents can be made to behave in that fashion. And, yes, the library has several options available to it to handle incoming files of aggregated data, thus relieving its staff

of much tedious, time-consuming, repetitive work involved with regular upkeep of large sets of aggregated data.

Item 1 of 1 Total
[Return to Search Results List](#)

Title [SIAM journal on matrix analysis and applications](#) : a publication of the Society for Industrial and Applied Mathematics / SIAM.

Imprint [Philadelphia, Pa. : The Society, c1988-
 Descr. v. : ill. ; 26 cm.
 ISSN [0895-4798](#)

Subject - Lib.Cong. [Mathematics -- Periodicals](#)
[Matrices -- Periodicals](#)

Add.Entry [Society for Industrial and Applied Mathematics](#)

Continues in part [SIAM journal on algebraic and discrete methods 0198-5212](#)

Campus Access [UC Berkeley](#) [UC Davis](#) [CSU](#) [UC Santa Barbara](#) [UC San Diego](#) [UC Riverside](#) [UC Riverside](#) [UC Santa Cruz](#) [Global view](#)

Library	Call Number	Availability	Notes
UC Berkeley		Available online	Vol. 22 (1988)- available online at
		Available online	Vol. 15 (1994)- available online at
Astr/Math	QA1 .S732a	Circ status	Unbound issues on Display Shelves - Alphabetically 9 (1988)-

[Previous](#) [Next](#)

Figure 5: UC Berkeley’s individual campus view of a title

Title [SIAM journal on matrix analysis and applications](#) : a publication of the Society for Industrial and Applied Mathematics / SIAM.

Imprint [Philadelphia, Pa. : The Society, c1988-
 Descr. v. : ill. ; 26 cm.

Numbering Note Issues for v. 20, no. 1- have cover dates of "papers published electronically."

Additional Form Also available to subscribers of SIAM journals online via the Internet.

ISSN [0895-4798](#)

Subject - Lib.Cong. [Mathematics -- Periodicals](#)
[Mathematics -- Periodicals](#)

Subject - Other [Online resources](#)
[Electronic journals](#)

Add.Entry [Society for Industrial and Applied Mathematics](#)

Continues in part [Society for Industrial and Applied Mathematics c2as SIAM journal on algebraic and discrete methods 0198-5212](#)

Campus Access [UC Davis](#) [UC Berkeley](#) [CSU](#) [UC Santa Barbara](#) [UC San Diego](#) [UC Riverside](#) [UC Riverside](#) [UC Santa Cruz](#) [Global view](#)

Library	Call Number	Availability	Notes
No. Regional Library Facility	QA1 .S732	Circ status	Volumes relocated to NRLF; inquire at Circulation Desk v.9-10,1988-1989.
UC Davis		Available online	Full text: v.18 (1997)-, Abstracts: v.15 (1994)- Access restricted to UCD IP addresses.
	Internet	Circ status	v.18, 1997- Electronic access.

Figure 6: UC Davis’ individual campus view of the same title

What does it mean for the end user? For the end user it promises a more sensible collocation of data. It promises a single record that represents a display that has more meaning to the user. It allows them to see variations in the presentation format of the data that in many cases are of significance only to technical services librarians. While it is true that some users will want to see the paper original, in many cases users would at least rather start with the electronic version that is readily available and revert to the original only if in need of the material in its ‘ur’ format. Except for ‘die-hard’ bibliophiles or antiquarians, it is the data that is important much more than the object itself. That is why Mulver of FRBR promises to be so important for our clientele. Why not offer them easy, collocated access from the inception of their search? As CDL has envisioned it, they not only have access to a combined, consortial record, but individual records are also only a click away.

Or, to quote another English phrase, this is like “having your cake and eating it too”.