The Changing Landscape of Hebraica Cataloging

Daniel Lovins
New York University, daniel.lovins@nyu.edu
The Changing Landscape of Hebraica Cataloging

Daniel Lovins

ABSTRACT

Hebraica catalogers, like other librarians, are witnessing a major shift in their profession. Catalog records for physical objects in the library are increasingly giving way to metadata for digital objects on the web. The RLG Union Catalog, a mainstay of Hebraica cataloging since 1988, has been absorbed into OCLC’s WorldCat. Rapid advances in information technology are driving the development of a new international cataloging code, the introduction of multiple languages and scripts in online authority records, and the emergence of a “Virtual International Authority File.” While these changes are redefining the kind of work Hebraica catalogers are engaged in, expertise in authority control and subject analysis remain essential elements in the emerging global network of libraries.

INTRODUCTION

The library profession is going through a time of great upheaval. For catalogers, the magnitude of change is such that the very word “catalog” seems to be losing traction. After all, what does it mean to catalog something that is not even owned by one’s library? Today, the object cataloged is often a temporary file on a remote server or a website that changes form and content on a daily basis. What was once thought of as the catalog looks increasingly like a “knowledge portal,” a gateway to resources that the library may never have selected or purchased in a traditional sense. Libraries increasingly lease periodical literature through article-aggregator databases,* for example, or point readers through OpenURL link resolvers to the publisher’s website. Digitization of entire research libraries,

* Editor’s note: Indeed, articles in this and the two previous volumes of Judaica Librarianship are aggregated in EBSCO’s full-text Academic Search Complete and LISTA databases.
full-text indexing on the Web, powerful search algorithms and relevancy ranking, are turning the Web itself into a kind of giant universal catalog, with Google as its virtual librarian. Many catalogers are now called “metadata librarians,” and their departments, “Metadata Services.” Metadata is like cataloging, except that it exists everywhere (not just in libraries), and can be generated by humans and machines alike. The concept of the catalog has gotten slippery, therefore, and the core intellectual work is being re-defined in terms of bringing readers together with resources, irrespective of whether the catalog as such plays a mediating role. The transformation has so far been less dramatic for Hebraica librarians than for others, but the general trend is unmistakable.

I. CONSOLIDATION/CONVERGENCE

A. RLG–OCLC Merger

The merger of the Research Libraries Group (RLG) and OCLC, effective July 6, 2006, was a watershed moment for research libraries in general and Hebraica cataloging in particular. Since its founding in 1974 by Columbia, Harvard, and Yale, RLG had provided critical infrastructure for the management and description of unique holdings and complex research collections. RLG’s Research Libraries Information Network (RLIN) allowed catalogers to exchange records online and build a union catalog of their collective holdings. RLIN’s record-clustering method was thought to be better suited to research libraries than OCLC’s “master record” approach, because it allowed each member to retain local cataloging decisions, while still collocating works and editions when browsing. It was especially helpful to rare book catalogers who needed to preserve copy-specific details about provenance, signatures, bindings, inscriptions, etc. It was important to Hebraica catalogers because different policies regarding multiple scripts (e.g., whether to include them, and if so, for which data elements) and culturally-appropriate descriptions (e.g., using terms like “Old Testament” for “Tanakh” and “A.D.” for “C.E.”) required a flexible and robust cataloging platform (cf. Weinberg, 1992a). The clustering technique, by preserving virtually all library-specific details, also allowed copy catalogers to identify and evaluate records by the initials of their creators. Those recognized as coming from a trusted colleague could be derived for local use with minimal editing or quality checks. Of equal importance was RLIN’s early implementation of JACKPHY scripts (Japanese, Arabic, Chinese, Persian, Hebrew, and Yiddish). Hebrew data-entry, indexing, and retrieval, was supported in RLIN since 1988, long before a similar capability was available in WorldCat (Aliprand 1987, 1991, 1992).

By 2006, however, as market conditions seemed to threaten RLIN’s long-term viability, the RLG Board of Directors concluded that libraries would be better served through a merger with OCLC (Hane, 2006; OCLC, 2006). Before voting itself out of independent existence and into a new OCLC division called “RLG Programs,” the institutional members of RLG needed assurance that OCLC would be able to satisfy their needs. In fact, OCLC had already implemented the JACKPHY scripts in WorldCat and was preparing to roll out Cyrillic, Tamil, and
Thai as well. Furthermore, OCLC agreed to restructure its database so that institutional records (analogous to ordinary RLIN cluster members) could be maintained alongside master records (which were similar to primary cluster members, though not necessarily the work of a single agency). Most of the membership was therefore convinced that merger should go forward.

By April 13, 2007, the entire RLG Union Catalog had been matched against WorldCat (OCLC 2007b), and 7.8 million new “master” records had been created (that is, titles or editions not yet represented in WorldCat). These included 300,000 new original script records. Moreover, 600,000 pre-existing WorldCat records were “enriched,” which is to say, they were matched against incoming RLG records, and original-script data were added as paired fields.

Since most Hebraica catalogers in the U.S. had been creating records through RLIN, the database migration led many to seek training in OCLC’s Connexion21 editing client. Alternatively, catalogers tried to use the editing module in their local ILS (integrated library system). Most libraries chose Connexion21, but some, notably the Library of Congress, preferred using their local systems so as to harmonize documentation, workflows, and training across all cataloging units.

B. WorldCat and the JNUL

Another important change involving OCLC was its agreement with the Jewish National and University Library (JNUL) to absorb 1.5 million JNUL bibliographic records. According to the press release (Adler, 2007; Hebrew University, 2007), 600,000 of these are in Hebrew script, 50,000 in Arabic, 50,000 in Russian, and the rest in the Latin alphabet. They include records for a collection of microfilmed Hebrew manuscripts that represent over ninety percent of such manuscripts known to exist.

While non-Israeli library catalogers (and users) will benefit greatly from access to JNUL WorldCat records, JNUL has the opportunity to benefit as well: Approximately 450,000 of JNUL’s pre-1984 Latin and Cyrillic alphabet records are still exclusively in card format. Interoperability with WorldCat, which likely contains matching records for most of these, will make it easier to replace these cards with electronic surrogates.

II. INTERNATIONALIZATION

A. Non-English Access Task Force

A “Task Force on Non-English Access” was charged by the Association of Library Collections and Technical Services (ALCTS) in August 2005 to “examine ALA’s [the American Library Association’s] past, present, and potential future roles in enabling access to library resources in all languages and scripts and in addressing the needs of users of materials in all languages and scripts through the development of library standards and practices.” A final report, published in 2007, included several recommendations that could impact Hebraica cataloging.
In particular, recommendation 2 calls for designated task groups to advise systems vendors and implementers on best practices for non-Roman scripts; recommendation 3 is to charge a group to work with the Program for Cooperative Cataloging (PCC) on updating their Guidelines for Multiple Character Sets (LC, 2008c); recommendation 4 is to study the effect the new cataloging code (see below) will have on non-English, non-Roman cataloging; recommendation 5 is to consider implementation of the Unicode Collation Algorithm (Davis & Whistler, 2008), which supports culturally-appropriate sorting of search results; recommendations 6 and 9 emphasize the need for non-English access not just for research libraries, but also for public and school libraries; and recommendation 8 is to recruit more language specialists with an eye toward “ALA’s commitment to a diverse library workplace.”

B. Cataloging Code

The world’s diverse cataloging traditions are converging like great tectonic plates, buckling at the seams, but promising (eventually) a unified bibliographic land mass that will offer collective solutions to common challenges. Differences need to be negotiated among languages, cataloging rules, and encoding systems, but the work is well under way.

The movement to internationalize cataloging principles—or identify the common principles already there—has taken on renewed urgency. The international agreement of 1961 known as the “Paris Principles” helped to unify cataloging practice over the past forty years (IFLA, 1963). More recently, the International Federation of Library Associations (IFLA) produced Functional Requirements of Bibliographic Records (FRBR), Functional Requirements of Authority Data (FRAD), and the IME ICC Statement of International Cataloging Principles (IFLA, 2005, 2007a, 2007b), all of which extend the Paris Principles beyond bibliographic description into the realms of subject analysis and authority control, also taking into account the rise of digital media.

Based on this emerging international consensus, the Anglo-American Cataloguing Rules (AACR2R, 2005) is to be replaced in early 2009 by a new cataloging code: Resource Description and Access (RDA). As stated in the October 2006 JSC (Joint Steering Committee for Development of RDA) Outcomes document: “The JSC affirmed the role of the IME ICC draft Statement of International Cataloguing Principles as the basis for the cataloguing principles used throughout RDA…”

---

1 Some difficult issues need to be resolved. According to the April 2007 JSC Outcomes document, the JSC will need to monitor the status of several IME recommendations on uniform titles, e.g.: 5.2.4.1, that the uniform title “should be the commonly known title in the language and script of the catalogue when one exists for the resource.” This raises the concern that agencies will construct uniform titles according to the language of their local catalogs, and that this will reduce the ability to share a common authority file. Both Britain (through CILIP) and the U.S. (via ALA) came out against this proposal. The ALA response pointed out that if the intention is to enable users “to bring the vocabulary at their disposal to the catalog and have a reasonable chance of success in discovering resources,” this goal could better be achieved through authority records, “without compromising the order of the underlying data.”
The Changing Landscape of Hebraica Cataloging

The shift in orientation to internationalism is nicely illustrated by the case of Bible uniform titles. As stated in a proposal submitted by the Library of Congress (LC), “The goal of incorporating a global perspective within RDA argues for taking a more balanced, neutral and culturally-sensitive approach to formulating certain headings.” Specifically, the AACR2 nomenclature and arrangement of Jewish and Christian scriptures follow the Protestant Christian tradition. Accordingly, the Hebrew Bible is expressed in AACR2 as “Bible, O.T.” or “Old Testament” (Anderson, 1992).

In an RDA-based catalog, however, a single book such as “Genesis” would be entered directly after “Bible,” that is, without the intervening “O.T.” One will still able to use terms such as “Old Testament,” “New Testament,” and “Apocrypha,” but they would only be for groupings of books (cf. AACR2 25.18A4), not to represent an a priori canonical order. Ironically, when this proposal was submitted for constituency review, the Association of Jewish Libraries (AJL) was less enthusiastic than other constituencies (notably, the Catholic Library Association). In a set of comments submitted on August 27, 2006, AJL’s Research Libraries, Archives, and Special Collections Division (RAS) Cataloging Committee questioned whether there is “a sufficiently clear understanding of emerging authority record models to foresee how [the uniform title guidelines] would be implemented.” AJL members were particularly concerned that allowing variant names as access points to the same work in the same authority file would scatter related items and make cooperative cataloging more difficult (cf. Weinberg, 1992b, 1994). One alarmed member thought the proposal “would introduce a wild and random variable into the quality of records available for copy cataloging”; another, that it would cause “absolute chaos.” Even though the AACR2 approach betrays an obvious Protestant Christian bias, there are those who find that it facilitates interdisciplinary and interfaith studies by providing consistent, predictable (albeit biased) access to closely related texts, regardless of the point of view of the user. An ostensibly more inclusive system, such as the one proposed for RDA, might actually hinder this kind of research by scattering commonly held books according to the idiosyncratic access points of each cataloging agency. The AJL RAS Cataloging Committee suggested that the Virtual International Authority File (see below) might be a better way to negotiate cultural differences in the bibliographic commons.

Another RDA proposal (simply called “Internationalization”) involves non-Western representations of words, numbers, and characters. It calls for retention of original language and script when transcribing data elements, with substituted or added romanized fields permitted, but only as an “option.” This is a departure from current practice where romanized fields, with or without paired original script counterparts, are mandatory. One benefit of the “Internationalization” proposal, should it appear in the published RDA, is that it will help

---

2 Interestingly, in its response to the proposal, ALA suggested doing away with the term “romanization,” since it reflects a bias toward the Roman alphabet. In the emerging global catalog, transcription/transliteration will occur in all directions, e.g., from Hebrew to Roman or from Roman to Chinese.
facilitate the exchange of records between Anglo-American and Israeli libraries through WorldCat. Since romanized fields will no longer be required, Anglo-American libraries will be able to import Hebrew-only Israeli records into their catalogs without violating the rules.

C. Multi-Script Authority Records

Another boost to internationalization is the introduction of non-Roman script cross-references in national authority records. According to an October 2007 news release, the major record exchange partners (often called “NACO nodes,” a reference to the PCC’s Name Authority Cooperative Program), namely the British Library, Library of Congress, National Library of Medicine, and OCLC, in consultation with the Library and Archives Canada, agreed to a framework for adding non-Latin script references to the LC/NACO authority file. At present, only NACO nodes are allowed to add the new scripts, but guidelines are being developed to enable NACO members to participate as well. During the initial implementation period, added scripts are limited to the MARC-8 repertoire of UTF-8 (Unicode), namely, Japanese, Arabic, Chinese, Korean, Persian, Hebrew, Yiddish, Cyrillic, and Greek. Other scripts will be added in the future.

OCLC has devised a way to “pre-populate” the authority records by extracting the paired script fields from newly enriched WorldCat bibliographic records. These names will be added as “variant forms” according to MARC21 “Model B” (LC, 1999), that is, without 880 tagging and without paired systematic romanization. Once the cross-references are in place, Hebrew-script access points to bibliographic records will become redundant. The question will still need to be addressed, however, whether it continues to makes sense to include transcribed elements (e.g., quoted notes) in both Hebrew and romanized script in bibliographic records.

Also, some of the issues that confound Hebraica catalogers when working with multi-script bibliographic records will now be an issue in the authorities format. For example, seemingly simple questions about whether a span of years that qualifies a Hebrew personal name should be displayed from right to left or left to right, and whether the invisible Unicode formatting characters that preserve the correct display of bi-directional text will be parsed correctly by a given application. For example, should Maimonides’ Hebrew name be displayed as:

\[ \text{רמיהו,ל ה,1135-1204} \]

Publishers and librarians have been inconsistent about this. And even when the underlying source code is consistent, different computer applications will sometimes display such character strings in unpredictable ways. In general, getting bi-directional script to display and sort correctly is a long-term challenge (Aliprand, 1987, 1991; Lazinger & Adler, 1998, p. 165 f).
Another question, for newly-established headings, is whether one should transcribe as cross-references simply what one finds in Hebrew script on the cataloged item, or rather construct them according to AACR2/RDA cataloging rules. And if the form appears as *ketiv male*, should it also be entered in *ketiv haser*? Or if in *ketiv haser*, then also in *ketiv male*? While Israeli libraries have reached a unified policy of sorts (Lazinger & Adler 1998, p. 67 ff.), the conversation among AJL libraries has just begun.

D. Virtual International Authority File

One possible solution to the conundrums of multi-lingual, multi-script cataloging environments lies in the Virtual International Authority File (VIAF), a resource that links together the authorized headings of various national authority files (OCLC 2007c). If successful, this project could allow each national library to take responsibility for headings and variants in its own native languages and scripts. A prototype is available at http://orlabs.oclc.org/viaf/, which, while so far including only U.S. and German headings, demonstrates the potential usefulness of the tool.

Search the name “Maimonides,” for example, and one obtains the preferred forms from the American and German libraries respectively:

| 2 headings found for *maimonides, moses, 1135* |
|-----------------|-----------------|-----------------|
| **Name**        | **Authority File** | **Sample Title** |
| Maimonides, Moses, f.1135-1204 | LC      | dalait al hairin |
| Maimonides, Moses, f.1135-1204 | DNB     | talmud bavli     |

In this case, with the exception of a comma after “Moses,” the German and U.S. headings happen to be the same. Having a coordinated registry of vocabularies and crosswalks, such as the VIAF promises to do, should eventually allow catalog users to choose the most “culturally appropriate” display of headings. It would then be a trivial additional step, if desired, to automate the preferred authority file at the library, user, or session level, so that only one preferred term from one authority file is invoked for any given search.

As another example, if one searches the name “Rashi” in the VIAF, one sees that the German and Anglo-American cataloging and transliteration rules have yielded different authorized forms:

| 2 headings found for *rashi* |
|-----------------|-----------------|-----------------|
| **Name**        | **Authority File** | **Sample Title** |
| Rashi, f.1040-1105 | LC      | perush rashal ha torah |
| SClclogo. Ben-YehaCop, f.1040-1105 | DNB | talmud bavli |

3 The fill characters (empty boxes) indicate a problem rendering certain characters. It should be fairly simple to fix for Hebrew script before the VIAF moves out of beta testing.
In theory, an Israeli authority file with preferred forms of names in Hebrew script could be included in the VIAF. The problem at the moment is that there is no central authority file in Israel. According to a November 29, 2007, press release, however, the Knesset voted to make JNUL a true national library, which may help other Israeli libraries rally around a unified Hebrew script authority file. Alternatively (or provisionally) AJL could try to create its own Hebrew script authority file that would link to the VIAF. Pre-existing headings from member libraries could serve as a base, and a NACO-like protocol for peer-reviewed contributions similar to NACO could be established.

In addition to dividing the labor of maintaining language and script-specific authority files, the VIAF might also help support the culturally appropriate sorting of concepts. For example—returning to the example of uniform titles above—invoking a heading for “Tanakh” in Hebrew script might pull in an entire syndetic structure for biblical works according to the Jewish canon out of the VIAF. Within any single national authority file, there typically needs to be one “official” hierarchy of concepts, but the VIAF might make it possible to offer users the syndetic structure of their choice.

III. FUTURE OF CATALOGING

Changes in the cataloging world happen in fits and starts, largely because there is little consensus on the best way to proceed. While there is widespread agreement on the continuing need for professional subject analysis and authority control (e.g., IU, 2006; Mann, 2005, 2005b; Marcum, 2005; PCC, 2005; UC, 2005), there are moments when even these pillars of bibliographic control are thrown into doubt.

A. LC and Series Authority Records (SARS)

This is what happened when LC decided to discontinue authority control for series titles, effective June 1, 2006. AJL and other library organizations feared that this was the opening salvo in what would become a deadly assault on traditional cataloging. Indeed, it seemed the battle plans had been released in advance in the form of Karen Calhoun’s LC-commissioned report: “The Changing Nature of the Catalog and Its Integration with Other Discovery Tools” (2006),

4 One could get such a file up and running quickly via the NSDL Metadata Registry. Based on the open-source Dublin Core Metadata Initiative (DCMI) Registry application, the registry exploits emerging Semantic Web tools such as Research Description Framework (RDF) and Simple Knowledge Organization System (SKOS). The mission of the Registry is to facilitate controlled vocabularies, crosswalks, and interoperability among all participating projects and data providers. Items from beyond the NSDL and NSF, including orphan schemes and schemas (i.e., those still in use but lacking current institutional support), are to be included as well.
which recommended “dismantling” LCSH and included other controversial proposals.

Like many library organizations, AJL (2006) submitted an open letter to LC’s Associate Librarian Deanna Marcum, criticizing the SARs decision and explaining the deleterious effect it would have on the management of non-Roman script materials:

Controlled series headings are especially important when providing access to documents written in non-Roman languages and scripts (such as Hebrew and Arabic). Non-roman script titles are a particular challenge because, within a single monograph series, they may appear in original script, in romanized form, and/or as translations. And each of these possibilities contains a further level of complexity: original script titles may have been printed in more than one orthography; publisher-supplied transliterations may have been derived from more than one romanization scheme; publisher-supplied translations may, at different points in the series, vary as to specific word choices. In addition, the cataloger must also provide a transliteration, and the standardized ALA/LC scheme he or she uses may differ from the ones supplied (if at all) by the publisher. The potential confusion caused by so many possible representations of a series title is staggering.

The letter acknowledged that the LC had the right to make its own decisions based on its own perceived best interests, but argued that the net effect of this decision by a de facto national library undermines cooperative cataloging efforts and would cause a tangible decline in service to library patrons. Furthermore, maintained the letter, it is inefficient because series records save time “by disambiguating similar titles, keeping track of cataloger research (so as to avoid duplicated efforts), and recording complicated series treatment decisions.”

B. LC Task Force On the Future of Bibliographic Control

In fact, following an almost universally negative reaction from the U.S. cataloging community, and better to coordinate policy changes in the future with other bibliographic control stakeholders, Deanna Marcum commissioned a new “Working Group on the Future of Bibliographic Control.” The Task Force was assembled in November 2006 and released its final report on January 8, 2008 (LC, 2008b). Interestingly, while the development and distribution of the Report was a model of open communication and inclusiveness (addressing two of the complaints on how the SARs decision was made), many readers were sur-

5 “Although a primary catalyst for formation of the Working Group was reaction in the library community to a Library of Congress decision to discontinue series authority control for the materials it catalogs, the focus of the Group’s work was much broader” (LC, 2008b, p. 9).
prised by its recommendation to discontinue development of RDA (cf. Academic Newswire, 2008). Here is the relevant section of the report:

3.2.5.1 JSC: Suspend further new developmental work on RDA until a) the use and business cases for moving to RDA have been satisfactorily articulated, b) the presumed benefits of RDA have been convincingly demonstrated, and c) more, large-scale, comprehensive testing of FRBR as it relates to proposed provisions of RDA has been carried out against real cataloging data, and the results of those tests have been analyzed.

The RDA publication date is set for February 2009; work has been going on for several years; and much time and money spent on drafting and reviewing chapters. Furthermore, an initiative to develop an RDA Application Profile in conjunction with the Dublin Core and World Wide Web Consortium partners was hailed by participants as a major breakthrough, and demonstrated practical data models and web services based on RDA and FRBR. John Attig, ALA representative to the Joint Steering Committee for RDA, reported afterward, “It is impossible to overstate the significance of this meeting. . . . It was clear to the JSC from the outcomes that the other metadata communities saw very definite benefits to be gained from particular aspects of the RDA project, and were prepared to collaborate actively in projects that would have important benefits” (Attig, 2007).

CONCLUSIONS

The RLG Union Catalog is gone, absorbed into WorldCat. AACR2 is giving way to RDA. MARC21 is slowly making way for MODS, METS, and other XML-based methods of data encoding and transmission. As Robert Wolven (2008) points out in a recent article, the clarion call for change had already gone out in a 1991 Library Journal cover story: “Cataloging Must Change,” and this helped bring about the Program for Cooperative Cataloging in 1995. But by most accounts, the change hasn’t happened quickly enough, and a sense of malaise still haunts the profession. Perhaps most disturbingly, more than sixteen years after the emergence of the World Wide Web, catalog records are still created using a Web-unfriendly encoding scheme from the 1960s, designed primarily for the formatting and printing of catalog cards.

The issue of whether bibliographic control is still needed in the age of Google is particularly poignant for non-Roman catalogers. Advances in optical character recognition (OCR), full text scanning, keyword searching, and social tagging, have indeed opened up vast new avenues for organization and discovery. Much of this is still tied to Western languages, but increasingly includes non-Roman script languages as well. Hebraica catalogers have an important role to play, however, in developing standards for machine processing, building the Virtual International Authority File, shaping development of a truly interna-
tional cataloging code, and providing high-quality structured metadata to Hebraica items whether they be "in the catalog" or out on the open Web.

**SOURCES**


The Changing Landscape of Hebraica Cataloging


Daniel Lovins is Hebraica Catalog Librarian and Team Leader at the Yale University Library. He is also the chair of the AJL Research Libraries, Archives, and Special Collections Division (RAS) Cataloging Committee. He has been the News Editor for Cataloging and Classification Quarterly for three years.