ARTICLES

Janice Kreider  67  The Correlation of Local Citation Data with Citation Data from Journal Citation Reports

Alenka Šauperl and Jerry D. Saye  78  Pebbles for the Mosaic of Cataloging Expertise: What Do Problems in Expert Systems for Cataloging Reveal about Cataloging Expertise?

Hsi-chu Bolick  95  Problems in the Establishment of Nonunique Chinese Personal Headings with Special Reference to NACO Guidelines and Vendor-Supplied Authority Control

NOTES ON OPERATIONS

Tamara J. Kuhn  106  Classifying Newspapers Using Dewey Decimal Classification


FEATURES

122  Index to Advertisers

Margaret Rohdy, Editor  123  Book Reviews

43(2): 65–128  ISSN 0024-2527
American Library Association
Association for Library Collections & Technical Services

For current news and reports on ALCTS activities, see the new ALCTS Newsletter Online at http://www.al.org/alcts/alcts_news.
The Correlation of Local Citation Data with Citation Data from Journal Citation Reports

Janice Kreider

University librarians continue to face the difficult task of determining which journals remain crucial for their collections during these times of static financial resources and escalating journal costs. One evaluative tool, Journal Citation Reports (JCR), recently has become available on CD-ROM, making it simpler for librarians to use its citation data as input for ranking journals. But many librarians remain unconvinced that the global citation data from the JCR bears enough correspondence to their local situation to be useful. In this project, I explore the correlation between global citation data available from JCR with local citation data generated specifically for the University of British Columbia, for 20 subject fields in the sciences and social sciences. The significant correlations obtained in this study suggest that large research-oriented university libraries could consider substituting global citation data for local citation data when evaluating their journals, with certain cautions.

University librarians continue to search for data that helps them evaluate their collections, particularly their journal collections, as subscription costs rise approximately 10% each year and as additional funds are needed to pay for access to full-text electronic journals. Use studies are time-consuming if one wants to obtain enough data to make the study meaningful, and the compilation of the results can be cumbersome. With the recent appearance of Journal Citation Reports (JCR) on CD-ROM, quantitative citation data are now relatively simple to manipulate. Because the citation data on JCR are global, librarians have questioned the relevance of JCR’s data to their own institutions, preferring citation data generated from their own users’ publications. These local data are more difficult to obtain, particularly if one wishes to restrict the data to one subject area, which is crucial because citation patterns vary by subject field and affect citation figures. In order to determine how global citation data relates to local citation data, I explored the correlation between global citation data from the JCR and local citation data for the University of British Columbia (UBC) for 20 subject fields in the sciences and social sciences.

JANICE KREIDER (jkreider@interchange.ubc.ca) is Coordinator of Collections at the University of British Columbia Library, Vancouver. The author wishes to acknowledge the use of the British Library Information Sciences Service and the Oxford University Libraries, William Ng of the UBC Library Systems Division who provided lists of UBC subscriptions by subject, and Kathy Li and Kelly Kwork who were students in STAT 551. Manuscript received June 11, 1998; accepted for publication August 18, 1998.
LITERATURE REVIEW

Previous studies on the relationship of global and local citation data have been few in number and limited to only a few subject fields. Wiberley (1982) studied the relationship of local and national citation data for social work by using an encyclopedia and four periodicals as the sources for the national citations and local citations from publications from lists of faculty publications. He compared the method of using earlier national citations to predict later local citations with the method of using earlier local citations to predict later ones and concluded that “national databases of citations are almost as good as local data in predicting future citation of journals by local authors.” This conclusion therefore suggests that national citation data are useful for journal selection. Wiberley’s study took place in a university known for its high productivity in publishing in the field of social work, but because the field was one of the applied social sciences, the field exhibits different citation patterns than does a subject in the pure sciences.

Joswick and Stierman (1997) showed that there was little relationship between global citation data from JCR and local citation data gathered from the three citation databases from the Institute for Scientific Information (ISI) and organized using DIALOG’s “rank” command. However, all subjects were considered together. The university’s comparatively small departments in chemistry and physics as well as the lack of a medical program meant that there was little correspondence between the predominance of social science titles cited by the professors and JCR’s top cited titles, which lean heavily toward scientific titles.

Garfield (1972) originally suggested using global citation data to evaluate journals. Although this practice, along with other criteria for evaluating journals, has since become fairly common in academic libraries, there is continuing discussion and debate on the relationship between citation data and other indicators of the use of journals in libraries. Since 1972, there have been numerous studies and reviews of the literature. Broadus (1985, 33) summarized previous studies and stated that “counts based on the JCR can be almost as good as expensive local [use] studies for predicting use of periodicals in a given library.” But Broadus cautions against automatically eliminating a journal that has low global citation counts in the library, as there can be valid local reasons for retaining the title.

Bensman (1985) summarized research on the correspondence between global citation data and use, and despite some conflicting studies, states (24) that “ISI citation frequency is measuring an extremely powerful variable in academic library use, and . . . is undoubtedly one of the most important measures that can be utilized . . . [for] managing the journal collections of research libraries.” Bensman was not concerned that the citation data were not local.

Todorov and Glänzel (1988) reviewed studies on the relationship between journal citation measures and objective and subjective ratings of scientific journals; in some of the studies there was a positive relationship, while in others there was not. Kelland and Young (1994) presented a comprehensive review of the literature on the relationship between citation data and library use. Because the correlation data vary, the authors conclude (86) that “citations represent a form of literature use, and to some extent that can be considered library use. . . . Actually, the relationship between library use and citations is so complex that it should not be expected to produce high correlations.”

Swigger and Wilkes (1991) compared techniques (including using local citation data generated by a DIALOG search of ISI) when evaluating journals at Texas Woman’s University and concluded (52): “There is only a weak correspondence between use of serials as measured by reshelving data and by citation data, and no correspondence between citation data and the subjective judgments of faculty or librarians.” The Texas Woman’s University has graduate programs in allied health, education, and library science, so the results are limited to those fields. On the other hand, Bensman (1996) showed
a strong correlation in the field of chemistry at Louisiana State University between global citation data (in the form of total citations from JCR for the journals) and ratings of those journals by faculty members.

To summarize, although there is not total agreement that the relationship of citation data with other indicators of use in libraries is strong, it generally is acknowledged that such a relationship exists and that it is worthwhile to take citation data into account with other measures when evaluating journal titles.

The question remains whether the citation data must be local, or whether global data are just as useful. Despite the difficulties inherent in generating local citation data, authors of articles published in the last several years reveal that librarians are using local citation data in preference to, or in addition to, global citation data due to a widespread reluctance to rely on global citation data to evaluate journals. Haas and Kisling (1994) reported on a project at the University of Florida to evaluate the relevance of their science journal collection using local citation data produced from ISI. Schmidt, Davis, and Jahr (1994) used local citation data, faculty ranking, circulation statistics, and impact factor to prioritize biology journals at the University of Illinois. Sylvia and Lesher (1995) used local citation data generated from psychology theses and dissertations in a cost ratio to evaluate their psychology journal collection at St. Mary’s University. Hughes (1995) examined local citation and publication data along with global data when ranking journals in molecular and cellular biology at Pennsylvania State University. Dole and Chang (1996) conducted four local citation analyses in the fields of marine sciences, chemistry, and sciences as a whole to produce local citation data to use along with faculty rankings and use studies in preparation for a journal cancellation project at the State University of New York at Stony Brook. Loughner (1996) produced local citation data from ISI to evaluate use of science journals at the University of Georgia. Lambert and Taylor (1996), who evaluated journals at Staffordshire University in the U.K., stated (318): “Citation rankings seemed too general; we would not have felt justified in canceling subscriptions purely on the basis of externally generated lists.”

**STATEMENT OF THE PROBLEM AND PURPOSE OF THE STUDY**

Citation data are commonly used along with other criteria such as reshelving statistics and ratings by faculty and librarians to determine core journal titles in a field or to find candidates for cancellation. Librarians often rely on local citation data rather than global citation data, despite the ease of gathering global citation data and the ability to manipulate them electronically, using the CD-ROM version of JCR. Librarians nevertheless remain unconvinced of the value of global citation data, thanks partly to the lack of studies. The purpose of this study is to explore the correlation between global citation data and local citation data to determine the potential of relying on global citation data instead of local citation data, for evaluating a journal collection. A second purpose is to design and test a consistent method for obtaining correlations across a number of subject fields at one specific institution.

**SOURCES OF DATA**

The local institution involved in this study was the University of British Columbia (UBC), a provincially funded university with the third largest academic library in Canada, a student body of 27,000 undergraduates and 6,000 graduates, and faculties not only in arts and sciences, but also medicine, dentistry, pharmacy, law, engineering, commerce, education, agriculture, forestry, etc. Its overall rank among Association of Research Libraries (ARL) has ranged between 25 and 30 during the past 10 years.

The local citation data were provided by the Local Journal Utilization Report (LJUR), produced by ISI from its database specifically for UBC. This report, in a database format, contains counts for the number of times authors from UBC cited
specific journals during a 12.5-year period from January 1981 to June 1993. The LJUR was generated using all author addresses, not solely the first author.

The 1994 edition of JCR on CD-ROM produced by ISI was the source of the global citation data for this study. The specific global citation data used were the "1994 Total Cites," which is the number of times a given journal title was noted as being cited in ISI's Science Citation Index and Social Sciences Citation Index during 1994.

The time periods of the two sets of data do not overlap; the sources were chosen because they were available in electronic format, which facilitated analysis using spreadsheets and databases. ISI produced the LJUR for UBC in 1993, but the JCR was not available on CD-ROM until the 1994 edition was produced in 1995.

Because this project was based on ISI data, it has all the limitations of that database:

- errors in the data themselves
- the limited number of journals covered (an example is the field of law, for which only 96 titles are listed in JCR, which led to the exclusion of the subject of law in this study)
- JCR is based only on journals; hence citations to journal articles from other types of publications are not included, and for subject areas that are less journal-centric, these might be significant
- few foreign language titles are included
- total citations are influenced by the length of time the title has been published; titles begun during or near the end of the 12.5-year period covered by the LJUR in this study will not have either global or local citation data comparable to those from well-established titles

**METHOD**

In order to identify journal titles for a specific subject field, two sources were used. The first was the JCR, which assigns each journal title to one or more subject fields. The second was UBC's existing local coding of all active subscriptions to one or more of the 20 subjects under consideration, corresponding to departments at the university. The subjects assigned by the JCR are finer categories than those assigned by UBC. In order to have them correspond as much as possible, the JCR was filtered to obtain similar subject groupings. For example, to correspond with the subject of "Chemistry" as defined by UBC, the following subject categories were chosen from the JCR: Chemistry; Chemistry, analytical; Chemistry, inorganic and nuclear; Chemistry, organic; Chemistry, physical; Electrochemistry; and Spectroscopy.

The following databases were loaded into a local database on a personal computer: the LJUR, the list of titles for each of the 20 subjects from the JCR, and the JCR lists of all science titles and social science titles. Using the UBC subject listings of subscriptions, subject codes were keyed into the LJUR. If a UBC subscription was not in the LJUR, the title was added, with zero as the number of citations. The database was used to create a list for each subject, with the following information for each journal: (1) its title as abbreviated by ISI; (2) the global citation count (i.e., the 1994 total citations from JCR); and (3) the local citation count (i.e., the number of times UBC authors cited the title over the 12.5 year period from the LJUR). The beginning of one of these lists is shown in table 1 for the field of zoology.

The Pearson correlation coefficient was chosen to measure the correlation between the global citation data and the local citation data. The Pearson correlation coefficient reflects the extent of a linear relationship between two sets of data; it ranges between -1 (indicating a perfect negative relationship) and +1 (indicating a perfect positive correlation). A correlation close to ±1 indicates a strong correlation. Other correlation studies have used the Spearman correlation coefficient, which measures the correlation of ranked data. However, if the raw data are available (not just the rankings), it is preferable to use the Pearson correlation because the actual data give more information, such as the varying size of gaps between the ranked data. Before applying the Pearson correla-
TABLE 1
ZOOLOGY JOURNAL TITLES WITH GLOBAL (FROM JCR)
AND LOCAL (FROM LJUR) CITATION COUNTS

<table>
<thead>
<tr>
<th>Abbreviated Title</th>
<th>Global</th>
<th>Local</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACAROLOGIA</td>
<td>142</td>
<td>0</td>
</tr>
<tr>
<td>ACTA ANAT</td>
<td>1,572</td>
<td>38</td>
</tr>
<tr>
<td>ACTA BIOL CRACOV ZOO</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>ACTA PHYTOPATHOL HUN</td>
<td>97</td>
<td>0</td>
</tr>
<tr>
<td>ACTA PROTOZOO</td>
<td>195</td>
<td>0</td>
</tr>
<tr>
<td>ACTA THERIOL</td>
<td>360</td>
<td>13</td>
</tr>
<tr>
<td>ACTA ZOOL HUNG</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>ACTA ZOOL STOCKHOLM</td>
<td>397</td>
<td>3</td>
</tr>
<tr>
<td>ADV INSECT PHYSIOL</td>
<td>425</td>
<td>34</td>
</tr>
<tr>
<td>ADV PARASIT</td>
<td>481</td>
<td>12</td>
</tr>
<tr>
<td>ADV STUD BEHAV</td>
<td>336</td>
<td>13</td>
</tr>
<tr>
<td>AFR J ECOL</td>
<td>189</td>
<td>5</td>
</tr>
<tr>
<td>AM BEE J</td>
<td>155</td>
<td>0</td>
</tr>
<tr>
<td>AM ENTOMOLOGIST</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>AM J PHYS ANTHROPOL</td>
<td>2,642</td>
<td>47</td>
</tr>
<tr>
<td>AM J PHYSIOL</td>
<td>78,546</td>
<td>2,634</td>
</tr>
<tr>
<td>AM J PRIMATOL</td>
<td>826</td>
<td>0</td>
</tr>
<tr>
<td>AM MALACOL BULL</td>
<td>79</td>
<td>4</td>
</tr>
<tr>
<td>AM ZOOL</td>
<td>2,652</td>
<td>202</td>
</tr>
<tr>
<td>AMPHIPACIFICA</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Unfortunately this procedure was not quite as uncomplicated as it would appear. As noted by Haas and Kisling (1994), Loughner (1996), and Harter (1998), raw ISI data can be disconcertingly inconsistent. The LJUR required extensive editing in order to combine titles with different abbreviations, to combine titles that had changed, and to correct errors. When the editing was completed, a great many titles listed with only one citation remained, and many of these were difficult to identify by full title, either because they were esoteric, ambiguous, incorrectly abbreviated, or referred to monographs. Examples of such titles include PHYSIS, PLENARY LECTURE, and TXB PSYCHIATRY. In order to eliminate having to spend undue time identifying these titles of little importance and in order to make the database smaller and easier to manipulate, all
of the titles having only one citation were eliminated except: (1) those in which UBC authors had published (using the publishing data from the LJUR); and (2) titles that began with CAN, CANADA, CANADIAN, BC, BRITISH COLUMBIA, VANCOUVER, etc., because we wished to retain as much Canadian data as was feasible. Although not necessary for this correlation study, obvious monographs; titles indicated as INPRESS, UNPUB, and THESIS; and titles beginning with numbers or abbreviations of months were eliminated.

Another modification to the LJUR was its expansion to include 755 titles in the 20 subject fields to which UBC actively subscribes. These titles had a local citation count of zero, but 175 had global citation data from JCR. The primary reason for adding these titles was to analyze all UBC subscriptions in the 20 subject areas for correlation between global and local citations. The effect for this study was an expansion of the data beyond that of ISI as the source. The periodicals added included esoteric titles, popular titles such as Sky and Telescope, new titles that had not existed long enough to have received citations at UBC, foreign titles, and titles perhaps a bit outside a strict definition of journal such as the Yearbook of Astronomy. Modifications to the JCR were also needed, because it was already several years old and numerous titles had changed or split into parts, so data for the various versions of a title had to be gathered under the latest version of the name.

The LJUR ended up with 10,601 titles, with local citations ranging from a high of 5,350 for the Proceedings of the National Academy of Sciences to the 755 titles with a citation of zero. The JCR for the science fields contained 4,438 titles, with global citations ranging from a high of 265,329 for the Journal of Biological Chemistry to a low of one citation for 13 titles. The JCR for the social science fields contained 1,402 titles, with global citations ranging from a high of 20,038 for Archives of General Psychiatry to a low of one citation for four titles. The two JCR sets had an overlap of 148 titles. The LJUR and the JCR for the sciences overlapped by 3,212 titles (72% of JCR), and the LJUR and the JCR for social sciences overlapped by 946 titles (67% of the JCR titles).

**Discussion of Results**

Table 2 lists the 20 subjects that were analyzed for this study, arranged in descending order by correlation of the data after they had undergone a logarithmic transformation, where N is the number
of journal titles in each subject cohort. The correlations between the global and local citation counts using the raw data were extremely high in most fields, a finding that was not corroborated by visual inspection of the scatter plots. Only five subjects had a correlation below .7, but as mentioned previously, the skewed distribution of the data (large clusters of both global and local citation data dominating the low end) indicates that more realistic correlations can be obtained from the transformed data. This led to the rationale for arranging the table in descending order by correlation of the data after a logarithmic transformation was performed. The actual ranking of the subjects is less important than the fact that the correlation figures are moderate to moderately high for all the subjects. The scatter plots of the transformed data show a linear relationship of varying strength for each subject, a relationship that gets decidedly weaker for the smaller citation values. Figure 2 shows the relationship for zoology. Note the cone shape of the data points that gets wider for the smaller values, indicating the increasing weakness of the relationship as the citation values decrease. The fact that the linear relationship between the global and local citation transformed data gets weaker for the smaller values has implications for libraries. It means that a low global citation count does not always imply a correspondingly low local citation count, despite the moderate to relatively high correlation figures for all the subjects.

The placement of the various subjects in the ranked list in table 2 does not permit any general conclusions about the strength of a subject's correlation based on whether the subject is a pure or an applied science, a science or a social sci-

### TABLE 2

**CORRELATION OF GLOBAL AND LOCAL CITATION DATA**

<table>
<thead>
<tr>
<th>Subject Field</th>
<th>N</th>
<th>Pearson Correlation of Raw Data</th>
<th>Pearson Correlation of Logarithmically Transformed Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microbiology</td>
<td>89</td>
<td>0.953</td>
<td>0.814</td>
</tr>
<tr>
<td>Forestry</td>
<td>140</td>
<td>0.817</td>
<td>0.810</td>
</tr>
<tr>
<td>Astronomy</td>
<td>42</td>
<td>0.963</td>
<td>0.755</td>
</tr>
<tr>
<td>Economics</td>
<td>260</td>
<td>0.916</td>
<td>0.731</td>
</tr>
<tr>
<td>Business and Management</td>
<td>279</td>
<td>0.849</td>
<td>0.695</td>
</tr>
<tr>
<td>Biochemistry</td>
<td>272</td>
<td>0.959</td>
<td>0.687</td>
</tr>
<tr>
<td>Pharmacy</td>
<td>176</td>
<td>0.832</td>
<td>0.682</td>
</tr>
<tr>
<td>Computer Science</td>
<td>254</td>
<td>0.696</td>
<td>0.681</td>
</tr>
<tr>
<td>Mining</td>
<td>39</td>
<td>0.617</td>
<td>0.677</td>
</tr>
<tr>
<td>Biology</td>
<td>346</td>
<td>0.968</td>
<td>0.675</td>
</tr>
<tr>
<td>Botany</td>
<td>154</td>
<td>0.783</td>
<td>0.661</td>
</tr>
<tr>
<td>Mathematics</td>
<td>307</td>
<td>0.800</td>
<td>0.634</td>
</tr>
<tr>
<td>Zoology</td>
<td>220</td>
<td>0.927</td>
<td>0.634</td>
</tr>
<tr>
<td>Physics</td>
<td>343</td>
<td>0.903</td>
<td>0.625</td>
</tr>
<tr>
<td>Mechanical Engineering</td>
<td>177</td>
<td>0.816</td>
<td>0.623</td>
</tr>
<tr>
<td>Librarianship</td>
<td>135</td>
<td>0.613</td>
<td>0.622</td>
</tr>
<tr>
<td>Geology</td>
<td>208</td>
<td>0.818</td>
<td>0.607</td>
</tr>
<tr>
<td>Chemistry</td>
<td>401</td>
<td>0.871</td>
<td>0.567</td>
</tr>
<tr>
<td>Metals and Materials</td>
<td>212</td>
<td>0.579</td>
<td>0.547</td>
</tr>
<tr>
<td>Chemical Engineering</td>
<td>148</td>
<td>0.477</td>
<td>0.530</td>
</tr>
</tbody>
</table>

N = number of journal titles in each subject
An interesting phenomenon is that local citation data are zero for numerous journals in all 20 subject fields. There seem to be three reasons for this. First of all, 755 titles were imported into the subject groupings because they were identified with the subject by UBC's local coding of all active subscriptions but were not included in the original LJUR due to the title's obscurity (e.g., *Amphiperasica*) or because the title was a popular rather than a research title (e.g., *Audubon*). A second reason is that most of the titles from the LJUR that had only a single citation were culled. Finally, some of these titles are simply of little relevance to research at UBC (e.g., *American Bee Journal*).

It might be assumed that researchers at UBC would cite Canadian journals more heavily than would be the case globally, which would produce relatively lower correlations in this study than would be obtained for an institution in the U.S. However, each subject field includes only a few Canadian titles, and often only one is truly important, so the effect is minimal. The subject list for business and management includes 25 Canadian titles, more than the other disciplines. Excluding the 25 Canadian titles and then recalculating resulted in a correlation of .702, compared with the correlation of .695 when the Canadian titles were included. However, the opposite effect occurred in forestry, which included 14 Canadian titles; recalculating the correlation without the Canadian titles gave a correlation of .794 instead of the .810 when the Canadian titles were included.

The *JCR* data (1994) and the LJUR data (1981–93) obviously do not come from the same time periods. At the time of this study, there was only one year of *JCR* data available on CD-ROM, and it might have seemed more worthwhile to have used only one year of local data (and from the same year) to correspond precisely with the one year of global data. However, for a given year, local citation data are considerably sparser than global citation data. Just as use studies based on reshelving counts require a long time frame to be representative, a citation study using merely one year of local citations would not be as indicative of local citation practices as more years would be, both in terms of the variety of titles covered and citation counts themselves, making a reliable correlation figure difficult to obtain. Whether a full 12.5 years of local citation data was necessary is open to debate. Certainly with only one year of local data, the correlations would have been much lower. Although *JCR* data change from year to year (probably more so for titles with few citations), they are still relatively stable—likely more stable due to sheer size—than are local citation data. Line (1985) found a 92% overlap for the top 500 journals cited the most frequently in ISI's *Science Citation Index (SCI)* in 1979 and in 1982. For the *Social Science Citation Index (SSCI)*, there was an 87% overlap for the same time period. Hence, using *JCR* data (which are drawn...
from SCI and SSCI) from a different year than that covered by LJUR should have a relatively minor impact on the results. The local citation data on LJUR would vary more from year to year than do global data. ISI has since enhanced the product to indicate the years of the citations, making it possible to study the growth or decline in citations for a particular title.

The nature of this project required that it be specific to our institution, and factors unique to UBC might have affected the data in unknown ways. Therefore, the results should not be overgeneralized for use by other institutions but rather should be viewed as providing an exploratory study of correlations for various subject fields using a specific method. The relative correlation for each subject is of less importance than the fact that a moderate to moderately strong correlation was shown consistently for all 20 subjects. This allows others to consider the similarity of their situation to UBC and to decide whether the size of the resulting correlations merits their substituting global citation data for local citation data when determining the importance of a journal to their library collection.

Because both the JCR data and the LJUR data come from ISI, it could be argued that the result has a high inherent correlation because ISI's database is based only on journals and the set of journals is limited. A better method might be that of obtaining the local citation data directly from local publications of various types, as has been done by some librarians when evaluating journals, such as Sylvia and Lesher (1995), whose source of local citations was graduate theses in psychology, and Dole and Chang (1996), who obtained local citations not only from a search of ISI but also from a list of publications by their faculty in marine sciences and from doctoral theses in chemistry. The local citations produced by the list of publications of the marine science faculty were dispersed over more titles than the ISI study, but the years of coverage did not coincide by date or length. Obtaining local citation data from sources other than ISI could produce more data, particularly in fields where the nature of publishing includes a substantial proportion of non-journal formats, such as monographs, conference proceedings, preprints, and technical reports. Obtaining local citation data from sources other than ISI, however, has the disadvantage of being extremely time-consuming to compile and organize; consequently past studies have been limited to few subject fields. It is difficult to evaluate studies of one subject area in isolation, and it is next to impossible to compare studies in which wildly different methods were used. The obvious advantage of the method presented here is that it can be done on a personal computer, and allows numerous subjects to be examined and results compared.

**Implications for Libraries**

It should be noted that the usefulness of both global and local citation data is limited to evaluating titles for research purposes; it has little validity for evaluating titles intended primarily for student use, particularly undergraduates. Joswick and Sterling (1997) showed that titles used by students do not bear any relationship to the number of citations in faculty publications. Moorbath (1993) studied the nursing literature and concluded (44) that there is "a significant difference between rank by student use and rank in the Citation Index." On the other hand, Zipp (1996) found a positive association between faculty citations and citations in theses and dissertations. It should also be remembered that results of this study are not applicable to smaller institutions that cannot offer as broad a base of ongoing research and publishing, because the local research might be less likely to correspond with global research and thus probably result in far lower correlations between local and global citation data. Swigger and Wilkes (1991, 44) stated that using local citation data is "likely more valid, particularly for small to medium institutions where research is more limited in scope and number of projects than at large universities with large faculties." Sylvia and Lesher (1995) used local citation data but also advocated considering shelving counts.
The results of this study do not give us information on whether the titles cited in a specific subject field were cited by researchers in that particular subject department. Because the data from the LJUR span the entire university, citations to a journal could easily be made by someone from a department in a different subject area. This has implications for academic department libraries, which are interested in limiting the evaluation of journals to their specific subject area or perhaps primarily are interested in the research scope of their specific department. The growing interdisciplinary use of journals and their escalating costs mean that most institutions have to take an institution-wide view to their library collections, so for them the department of the researcher is increasingly irrelevant.

CONCLUSION

University librarians are always searching for ways to evaluate journals, due to the increasing costs of journals, the creation of new journals, and the limitation of funds. Data often considered include circulation and reshelving studies, ratings and rankings by faculty members, local and global citation counts, impact factors, availability of indexing, reputation of the publisher, cost, language, interlibrary loan requests, availability from document suppliers, etc. All this information has value when considered along with subjective evaluations by faculty and librarians. Recently, librarians such as Schmidt, Davis, and Jahr (1994), Hughes (1995), Sittig and Kaalaas-Sittig (1995), and Loughner (1996) have proposed various tools or instruments for evaluating journals that combine two or more of the following factors: a ranking or rating by faculty members, circulation statistics, local or global citation counts, local publication counts, costs, and impact factors.

In this present study, it was shown that a relatively high correlation exists between global citation data of one year and longitudinal local citation data for a preceding 12.5-year period for a large research-oriented university for 20 subject fields, suggesting that large research-oriented university libraries could consider substituting the more easily-obtainable global citation data (from JCR) for local citation data (whether from ISI's LJUR or a search of their citation indexes or from an analysis of local publications) when evaluating their journals. High global citation counts have been shown to correlate with local citation counts. However, because the relationship between the global and local citation data gets weaker for the lower values of the data, librarians should exercise caution when evaluating titles with a low number of global citations and seriously take other factors into account. Faculty members are likely to be more interested in local data, so if local citation data are readily available such as ISI's LJUR, they can serve a double purpose as a tool of interest to the faculty along with being of use to the librarian, who, because of the correlation of global and local data, has little need to consider global citation data in addition.

WORKS CITED


Pebbles for the Mosaic of Cataloging Expertise: What Do Problems in Expert Systems for Cataloging Reveal about Cataloging Expertise?

Alenka Šauperl and Jerry D. Saye

Twenty-five years of research in expert systems for descriptive cataloging and related areas are reviewed. Researchers who developed prototype expert cataloging systems in the 1980s found that cataloging rules are extremely comprehensive and complicated, but still insufficient to permit these systems accurate cataloging results. Those researchers also identified smaller areas that need to be pursued for successful implementation of expert cataloging systems. Subsequent research has focused on narrower areas. Cataloging rules were studied and more was learned about problems with their logic structure and organization. Another focus was optical reading of bibliographic elements in documents. Categories of responsibility and graphic design continue to pose problems. The visual characteristics of documents were studied to understand more about the automatic recognition of bibliographic elements necessary for the bibliographic description of documents. The systematic study of the cataloging process, necessary for development of expert systems, may result in improvement of manual working procedures and enrich the education of new catalogers.

**CATALOGING AND EXPERT SYSTEMS**

Cataloging is one of the most time-consuming tasks performed in libraries. It is not surprising then that, since the introduction of computers into library operations, there have been numerous attempts to automate the cataloging process. The early use of computers to create catalog records that were used most often to produce cards for card catalogs has been largely supplanted by the creation of bibliographic records that are used to populate online catalogs. The most challenging aspect of cataloging is to determine the content of the bibliographic record rather than to create the record itself. This is where research in expert systems and related areas plays an important role.
In expert systems computers are used to organize and preserve the expert knowledge of catalogers and to employ that knowledge in cataloging with minimal or no human intervention. Unfortunately, few of these prototype cataloging expert systems have succeeded in the real world. They have contributed, however, to an understanding of cataloging processes and standards, as well as to an understanding of the documents cataloged. When researchers in this area determined that the entire cataloging process was too complex to allow for full expert system implementation, they then concentrated on some elements of the cataloging process, such as cataloging rules, document characteristics, and the characteristics of names used as access points. These efforts resulted in a better understanding of both the advantages and limitations that computers offer to catalogers. Read from another perspective, however, the problems found in having computers create cataloging might be indicative of the problems catalogers themselves face. This might be particularly true for novice catalogers and students of cataloging.

Cataloging standards are an essential part of the knowledge base for the cataloging process. They state the purpose, procedure, and results of the process. The most important standards used in cataloging in the United States are the Anglo-American Cataloguing Rules, 2d ed. (AACR2) and the documentation that supports the Machine-Readable Cataloging format (MARC). These are supplemented by national and local written policies and guidelines as well as by the unwritten practices and policies of libraries, library systems, and bibliographic utilities. Catalogers must learn and master all of these in order to perform their jobs successfully.

The purpose of a catalog is to create a surrogate of a document to support access to that document. A variety of document formats are represented in library catalogs. Document characteristics, in combination with expected retrieval needs, should determine what is contained in the catalog record. Catalogers learn to identify those characteristics, interpret them in accord with the existing rules, and represent them in meaningful, standardized descriptions.

The strength of developments in expert systems relative to cataloging is the process of knowledge acquisition. In this process, specific cataloging knowledge, decisions, and work procedures are systematically studied. The questions motivating the research described in this paper were: What have we learned in the 25 years of research into expert cataloging systems about the characteristics of the cataloging process, the materials cataloged in that process, and records created in the process? If expert systems do not provide the answer to cataloging problems now, can a better understanding of what cataloging is, characteristics of its processes, objects (materials), and results (records) help to improve the quality of work and the training process?

DEFINING EXPERT SYSTEMS

"Expert systems are computer-based systems that use knowledge and reasoning techniques to solve problems that would normally require human expertise" (Morris 1992, 1). The basic elements of expert systems include a knowledge base, an inference engine, an interface, and a general database. Each is described below.

The knowledge base is where human expertise is organized and stored. This base contains facts and heuristics of the domain of expertise. In the case of automated cataloging, the knowledge base would contain the current cataloging rules and modifications for local practice as well as the experience of catalogers.

The inference engine enables the formalized decision-making process. It makes use of the knowledge base to enable the system to perform like an expert. The engine infers results based on the stored knowledge.

The interface supports communication between humans and the system. Usually, the interface has three components. First, a user interface enables communication with the user (the cataloger who uses the system). Second, a developer's interface assists the knowledge
engineer (the person who develops the expert system) and makes the development of the system possible. Third, an external interface provides for data exchange with external sources (e.g., authority files or classification schedules).

The general database keeps track of the current problem and records all the relevant data and steps in the process.

Expert systems have shown a great deal of potential in the cataloging arena for several reasons:

- Human catalogers can be reassigned from routine tasks to work on more exciting and creative tasks. Thus, catalogers could leave simpler items to the expert system and work primarily with documents that are more difficult to catalog. Similarly, greater resources might be dedicated to other cataloging tasks such as enhancement of subject access.
- Expertise, which is rare and difficult to acquire, could be archived and saved for the future. When experienced catalogers leave a library, expert systems could be used as consultants or even as sources of training for less-experienced colleagues.
- Expertise could be distributed more widely and used more readily. In contrast to human experts, expert systems could be used 24 hours a day by multiple users at different sites.
- Critical examination of the decision process could be enhanced. The analysis of cataloging tasks necessary for the development of the system would force developers to understand and describe the work process systematically. This might identify errors and illogical steps in the workflow that were not obvious before.
- Full understanding of all parts of the process could help in the development of standardized approaches to problem solving for particular tasks. These standardized procedures might result in more consistent catalogs as well as support the training of novices.

With all their positive qualities, expert systems would seem to be the obvious cataloger's tool of the future. However, the realization of this ambition has been impeded by a number of challenging obstacles:

- The building of an expert system requires a substantial amount of time and work from the builders of the system (the knowledge engineers) and from the experts in the particular domain (here, experienced catalogers).
- The expert system's domain needs to be carefully chosen within a narrow and well-defined area if the system is to be reliable. Despite original hopes, researchers have found that descriptive cataloging is too broad a domain. Although cataloging rules exist, they have not proven sufficient for accurate cataloging. Researchers have also found that cataloging rules are difficult to transfer into a knowledge base. In studies, it has been indicated that only when cataloging is divided into small subdomains will the resulting expert system be successful.
- Knowledge engineers, who are responsible for preparing the computerized version of the human expert, need to know the computer system well. They also need to be familiar with the expert domain in order for the required knowledge to be represented in the system in a meaningful and useful way.
- Knowledge engineers, who are also responsible for acquiring knowledge from human experts, frequently have been unable to get those experts to express completely what they think and do.

To explore more fully the potential of expert systems for use in cataloging operations, we reviewed 25 years of research. The studies examined yield important information on the fundamental requirements for designing the essential knowledge base for creating a standard cataloging record.

**Prototype Expert Systems**

Davies and James (1984) were the first to investigate the feasibility of creating an expert system for cataloging. They tried to develop a system that would provide a complete catalog entry (computer record or catalog card). Their system required a
cataloger to choose options from a menu about the type of publication under consideration and access points. After a series of questions about the publication, the system offered a template, called a frame, for the cataloger to record the remaining data for bibliographic description. Davies and James found their system very complex and difficult to manage. Two main reasons were identified: (1) AACR2 is very complex; and (2) the general database, which kept track of the current problem, needed more space than their computer systems was able to provide at that time. In essence, the cataloging process was more complex than the existing technology available was able to manage.

Davies (1987) explained further how frames could be used in cataloging to infer and complete particular fields in a MARC record. He believed such a system should be able to infer who the publisher was from the International Standard Book Number (ISBN) and provide the complete publisher element. This could reduce some routine typing and typographical errors. Although not specifically stated by Davies, it is clear that accuracy—particularly in spelling—is vital. Errors in these areas are not trivial for they can have substantial impact on later retrieval. A good expert system would support the cataloging process with editing input.

Cataloging expertise, however, often lies in the choice of appropriate rules for bibliographic description and the choice of access points rather than in the simpler inference of publisher or the consistent spelling of words. The system created by Davies and James was intended to guide catalogers through the construction of the bibliographic record by providing rules relevant to each particular element of the description. Ultimately, this was found to be too complex for computer systems of the early 1980s.

While Davies and James worked on their system, Hjerppe and Olander (1989) built the Expert System for Simple Choice of Access Points for Entries (ESSCAPE). Hjerppe and Olander's system asked questions relevant to access points and the cataloger answered them. This interaction resulted in the system suggesting the AACR2 rule number appropriate for given situations. Hjerppe and Olander eventually abandoned further development of this system because they concluded that AACR2 alone was not sufficient to allow for the automatic decision making in the selecting of access points. They found that human interpretation was essential because the cataloging rules were neither self-contained nor clearly formulated.

Evidently, the experience of catalogers enables them to develop more specific rules to given situations than are provided by the cataloging code for decision making. Varied experience can lead to differences in the interpretation of the information not provided by the cataloging rules. Different or inconsistent inferences for missing information can result in inconsistent catalog records.

Gibb and Sharif (1988) also created an expert system to assist catalogers in creating full catalog records. By following menu options in the system and using "yes/no" and numbered decision alternatives, a cataloger was led to the appropriate AACR2 rule. Based on their experience, Gibb and Sharif commented that using only AACR2 as a knowledge source was appropriate. This would allow the lengthy process of knowledge acquisition from experts to be bypassed and avoid personal bias. Gibb and Sharif suggested that the system was particularly useful for training new catalogers.

Gibb and Sharif's optimistic conclusions contradict those of Hjerppe and Olander. This difference may be due to different expectations. Hjerppe and Olander were looking for a system that used AACR2 to generate correct access points for the variety of items encountered in a real environment, and concluded that AACR2 was insufficient for this purpose. Gibb and Sharif, on the other hand, developed a system for the accurate use of AACR2 and were in general less concerned with cases in which those rules did not give sufficient guidance.

If cataloging is so complex, it seems reasonable to expect that a successful system might better be limited to addressing
specific document types. This would likely result in fewer rules being required to make the system operational. Ercegovac (1990) and Ercegovac and Borko (1992) followed this approach and created a semi-automatic cataloging advisor to assist catalogers in the cataloging of maps. Their system, “Mapper,” helped catalogers establish main entry, title statement, statement of responsibility, publisher, place, and year of publication. Their expert system differed from previous efforts because the developers incorporated principles from:

- the system’s performance testing;
- user interface design;
- different sources of knowledge for the development of the knowledge base; and
- knowledge elicitation methods to gain the necessary knowledge from map cataloging experts (Ercegovac 1990).

Ercegovac’s method for system performance evaluation was to examine the results (the catalog records) and the user interface. Catalog records produced through Mapper were compared to the catalog records for the same maps prepared by the Library of Congress (LC). This standard for comparison was chosen because LC records are often treated as the U.S. national standard for quality cataloging. Although tested on only a small sample of maps, Mapper performed well and its user interface was well accepted by its users. Overall, it was perceived as helpful, easy to learn and fast.

Davies and James (1984) used cataloging rules and local guidelines as sources of expertise while Gibb and Sharif (1988) and Hjerppe and Olander (1989) used different parts of cataloging rules and, very likely, their own personal knowledge, although this is not mentioned in their reports. All commented on the weaknesses of cataloging rules and the adjustments that would be necessary to implement them in the knowledge base of an expert system. Ercegovac (1990) showed how to overcome these weaknesses with additional sources of expertise. First, one must observe common features of the documents. The printed logos, number codes, and the fold of a map can give clues to its producer. Second, one must interview experienced catalogers about their understanding of documents and the process of map cataloging. Because the rules don’t clearly identify the bodies responsible for a map, that information was gathered from catalogers and developed into a definition. Additionally, Ercegovac inquired about the experts’ approach to cataloging and the cataloging process.

One area of cataloging that novice catalogers must master is knowledge of the sequence of appropriate cataloging rules to use in particular cases. Each of the expert systems guided the cataloger through a sequence of questions, suggesting the appropriate rule but requiring the cataloger to provide the necessary bibliographic data. This cooperation resulted in the creation of a partial or complete bibliographic record. Davies (1987) later suggested that cataloging systems interfaces could provide assistance with templates by interpreting some data across different fields of the record. Interfaces could also assist by prompting catalogers for required data following an established pattern of rule sequences.

It appears obvious from these research efforts that traditional cataloging is successful because it incorporates printed standards, national and local policies (written and unwritten), and the experience and knowledge of the cataloger. It is very likely that in order for an expert system to be successful it must incorporate all these elements. These also are areas that novice catalogers need to master over time through acquisition of experience. Library and information science programs can usually provide some instruction on rules and document characteristics, but rarely are able to provide the development of an experience base at a level sufficient for quality cataloging.

It may come as a surprise that the development of expert systems essentially ends with these described prototype systems. The basic unit of an expert system for cataloging still has not been resolved—that is, the structure and organization of the cataloging rules appropriate for computer use. Without that fundamental unit in place, the tedious and
expensive process of cataloging knowledge acquisition for developing the knowledge base seems to lead to no usable product. Although Ercegovac (1990) demonstrated how an expert system could overcome the difficulties of the cataloging rules and incorporate appropriate expert knowledge, no one has reported that they have resumed the development of expert cataloging systems. It appears that catalogers and library managers both perceive other avenues as easier and more reasonable ways to reduce the human effort and thus the cost of cataloging. Efforts such as outsourcing of cataloging and the development of cataloging workstations seem to have taken the place of a fully matured expert system.

Recently, researchers and developers have focused on the improvement of the cataloger's tools, e.g. cataloger's workstation (Brisson 1995a; 1995b). In this approach, advances in computer technology—from dedicated terminals to personal computers—are predominantly used to benefit human work patterns. At first these workstations were only able to handle one task at a time, but eventually were improved to allow the use of multiple applications simultaneously. The major advantage of these workstations is that catalogers can access electronic versions of cataloging resources. Although such workstations are very useful and bring immediate benefits to the catalogers, they do not reduce the intellectual effort in the cataloging process, which continues to be the catalogers' responsibility.

**CATALOGING RULES**

While some researchers tried to build prototype expert systems and found it difficult to implement cataloging rules in these systems, others were interested in the function and structure of the rules. This is a narrower yet vital part of a comprehensive cataloging expert system development because it is an essential part of a knowledge base. Different chapters of the cataloging rules have been studied from a variety of perspectives, with the common purpose of enhancing their understanding and suggesting how the rules could be improved in further editions.

Svenonius and Molto (1990), Davies (1992), and Weibel (1992) prepared extensive reviews of automated descriptive cataloging and related research. They point to Sandberg-Fox (1972) as the first to work on this topic. Sandberg-Fox (1972) compared the Anglo-American Cataloguing Rules (1967) (Agricultural) to the older A.L.A. Cataloging Rules for Author and Title Entries (1949) for the computerized assignment of access points. She found that both rules posed problems for implementation in a computer environment. Given the problems that the prototype expert systems encountered and the comments of researchers on their difficulties with implementing AACR2, Sandberg-Fox's result suggests that these problems were already present in previous editions of the cataloging rules.

Jeng (1991a) analyzed chapter 1 (General Rules) of the Anglo-American Cataloguing Rules, 2d ed., 1988 revision (AACR2R), to identify the definition and functions of a rule. For the use of rules in cataloging monographs she created condition/action pairs for rules from chapters 1 and 2 (rule 1.0A1, Sources of Information, to rule 1.1G4, Items Lacking a Collective Title, and rule 2.0A1, Scope, to rule 2.1G2, Relationship Between Statement of Responsibility and Works in an Item that Lacks a Collective Title) into 77 condition/action pairs. This process would be necessary to implement cataloging rules in the knowledge base of an expert system.

Further, in a sample of 50 title pages, Jeng identified which rules were applicable and how frequently they would have been used. She determined only 13 of the 77 condition/action pairs were applied to every title page in the sample, while more than half of the condition/action pairs were never applied. For the Rule 1.0C1, which was divided into 25 condition/action pairs, only 3 pairs were applicable to all the title pages in the study. Although the sample of title pages was small, the results suggest that the rules may not be equally applicable, with some being general and some very specific. We feel that for a much larger sample of documents in a limited domain it could be anticipated
that not all the rules would need to be implemented, thus making the knowledge base of a system somewhat more manageable. While the findings of some researchers agree with Jeng’s, others have found the contrary—in some cases more rules would be necessary because of the specialized role of the rarely used rules.

Meador and Wittig (1991) studied how frequently the rules used for choosing access points in AACR2 (chapter 21), were applied to books in chemistry and economics. In their experiment, only 12 of 143 rules in that chapter were used on a sample of 30 items from each discipline. The authors concluded that, although the core rules were different for each discipline, the same expert system could work for both because: (1) a small number of rules was used in total, and (2) rules used for economics represented a subset of rules used for chemistry. This study, although on a small sample, indicates that some of the rules are general and could be used across different disciplines.

Weiss (1994) reported on a project conducted at the National Library of Medicine (NLM) intended to create an expert system to assist catalogers with personal name authority work. Implementing only AACR2 chapter 22 (Headings for Persons) proved insufficient. Weiss found that to construct personal name headings correctly, the expert knowledge of experienced catalogers was required for the correct translation of personal names appearing in documents into personal name headings. He concluded that acquiring that expert knowledge is a time-consuming and expensive task. One might imply from Weiss that the foundation for correct decision making in cataloging requires not only a sound knowledge of cataloging rules but also an extensive experience in the application of these rules.

The results of Jeng’s (1991a) and Meador and Wittig’s (1991) studies point in the same direction. Because the entire set of cataloging rules is difficult to implement in a knowledge base, using only a subset of the rules may be a better approach. If the rules used more frequently are also the most important rules and conclusion of other rules does not lead to problems, perhaps only that smaller set of important rules needs to be implemented in an expert system. Weiss’s (1994) findings do not appear to support this conclusion, however. These contrasting results show that using a subset of rules in a cataloging expert system needs to be studied more fully. One way to address this problem might be to develop a knowledge base using the subset of important rules to identify problems that arise. Judging from the results of prototype expert systems, it is clear that there is some subjective interpretation of cataloging rules because of missing information or insufficient guidance in the rules. Moreover, in the studies cited, the issue of researcher bias is not addressed, and its effect is undetermined.

Clearly, there are some difficulties with the logical structure of cataloging rules. Hjerppe and Olander (1989), Ercegovac (1990), Jeng (1991a), and Weiss (1994) all commented on problems with inconsistent, contradictory, insufficient, and redundant information contained in the rules. Fidel and Crandall (1988) examined AACR2 from a generalized database approach, using the entity–relationship model. They categorized rules into six types: (1) content; (2) establishing entities, relationships, or attributes; (3) authorized sources; (4) domain; (5) format; and (6) access points. They demonstrated that the current arrangement of AACR2 scatters rules of the same category in different parts of the text and also mixed different categories into the same rule. They believed that this adds to the confusion of what are and what are not conceptual rules. They suggested that systematic investigation of the structure and function of the rules would be possible with the entity–relationship model.

Taniguchi (1996) built a prototype expert system for analyzing AACR2R. This system was designed to analyze the internal structure of the rules and the relationships among them. The first phase of the analysis in this system, a manual one, was a transformation of cataloging rules from natural language into a more formal,
structured language that a computer system could process. The next phase, a computerized approach, involved the normalization of rules, i.e., converting a rule from its still relatively natural, but already structured form, into an entirely mathematical, logical form (condition/action pairs). The third phase of Taniguchi’s system involved testing of a rule with rule templates. These rule templates were skeletal schemes that provided the very basic structures of rules. They consisted of the characteristics, i.e., the variable information of condition/action pairs, necessary for logical functioning and verification of information for all the necessary components. Three types of templates were developed: (1) for descriptive rules, (2) for definition rules, and (3) for organization rules.

It seems reasonable to expect that Taniguchi’s first phase was rather subjective, given that rules can be interpreted differently by different people. As a consequence, the second and subsequent phases could then have different results as well. Yet, these differences might illustrate the missing information in the rules, which might lead to improvements in rule structure.

There are different views on the categorization of cataloging rules. While Taniguchi’s three categories of rules were description, definition, and organization, Jeng (1991a) listed five categories: definition, description, organization, identification of the source of information, and transciption. Fidel and Crandall (1988) specified six categories: content; establishing entities, relationships or attributes; authorized sources; domain; format; and access points. Taniguchi (1996) expressed a need for consensus on this issue as a condition for success of any analysis and suggested that the proposed system offered an appropriate method that could help with developing less ambiguous and more consistent rules.

Although one might argue that human reasoning is different from computer processing, the computerization of a process offers a way for people to organize their thoughts and produce a more systematic assessment of their own work. The system may not be perfect or be able to perform analysis independent of human experts, but it offers a way to gain a deeper and more systematic understanding of human experts’ own reasoning, including their strengths and limitations.

In these studies of AACR2, researchers have indicated how, and why, it is difficult to “teach” a computer how to catalog. This should come as no surprise to instructors of cataloging, who encounter the need to convey document patterns and the intricacy of what is stated in the rules, and (equally as important) what is not stated in the rules to their students. Complicating all these instructional challenges are the different life experiences that the students bring to cataloging process. Much of the “artistry” of teaching cataloging seems to lie in teaching how to search for clues that provide a basis for the creation of an appropriate bibliographic description. Such clues often are found in the combination and context of the bibliographic elements in a document and the cataloging rules and their interpretations appropriate to the situation. The question arises of how this artistry can be formalized in an expert system. For the majority of “typical” documents, expert systems using knowledge base likely could work independently, while for more complicated documents, these systems would need to interact with human experts.

CHOICE AND FORM OF ACCESS POINTS

Access points are necessary because they offer a way to access catalog records systematically and to arrange them for display. Thus, both their selection and their formatting are essential components of existing catalog codes. Access points have been a special area of inquiry from two perspectives. First, they have formed a subset of studies of cataloging rules. Second, access points have been studied with a view toward better control of what would be one part of a comprehensive expert cataloging system. Such a system would incorporate automatic reading of parts of documents to produce data for the creation of access points.
Svenonius, Baughman, and Molto (1986) reported studies of automatic recognition of names and titles on the title page. Their purpose was to identify the possibilities of simplifying the determination of name access points. Their work made the assumption that optical character recognition (OCR) was advanced enough to identify text on title pages accurately. They investigated responsibility statements in a sample of 400 monographs from an academic library. All the documents were in English. Author names most frequently appeared on title pages, preliminaries, and tables of contents (38% of the 1,310 author names in the sample). Tables of contents contained the largest share of all names (66% of the total 2,536 names). However, the researchers further found that many of the personal and corporate names that appeared on title pages, preliminaries and tables of contents were neither authors nor people or bodies whose function was related to authorship.

Svenonius and Molto (1990) investigated the possibility of automatic recognition of type of responsibility from machine readable title pages of monographs. They also examined whether title page data were adequate for identification of name access points. The success rate was measured by the number of correctly identified access points (all names), and in terms of precision (the number of correctly identified name access points automatically identified by their algorithm compared to the records created by LC or NLM). Their system correctly identified 90% of the names that appeared on the title pages. Of the correctly identified names, 95% were also chosen as access points in LC and NLM cataloging records. Svenonius and Molto estimated that approximately 88% of the access points selected by LC and NLM could be automatically derived from title page data.

Molto and Svenonius (1991) also investigated automatic recognition of names appearing on samples of title pages from monographs in an academic library and a public library. Their method correctly identified 86% of corporate names and 85% of personal names. Only 65% of the names identified as personal names in the sample were also chosen by LC as personal name access points. Approximately 14% of corporate names and 15% of personal names were either not recognized or were incorrectly recognized. Molto and Svenonius suggested that their rules for establishing name access points and the graphic presentation in the documents contributed to the incorrect or failed recognition.

The state of OCR technology at the time of these tests could not reliably interpret text for cataloging purposes. Problems arose due to the function of names and the artistic design employed in the documents. Systems have difficulties identifying the function of some authors. While illustrators and translators are usually clearly credited, the roles of editors, compilers, and other contributors can be harder to determine. In some cases these functions had to be inferred. Electronic files used to produce printed texts might be more amenable for computer interpretation. Perhaps cataloging rules need to be more flexible to allow the use of this new technology.

Excluding these obstacles, two issues emerged from the research on the automatic recognition of access points. First, it was not possible to anticipate all the possible personal and corporate names to be included in name authority files. These files are a necessary component of expert systems. Words scanned in a document need to be compared to the names in the authority file by an expert system, in order to be accepted as a name, verified against other names, and accepted as an access point. Second, even if it were possible to identify all the appropriate names correctly and their current form automatically, the ability to distinguish different functions of authorship remains a challenge. A mechanism is needed to distinguish among writers, translators, editors, and numerous other people who appear on the document and are responsible in different ways and degrees for its creation.

Both of these are major impediments to building expert cataloging systems as well as a challenge for the everyday work of catalogers. The identification
and interpretation of bibliographic data in a document is the underlying skill. The "fuzzy" use of language, which poses such difficulties for computer systems, is less problematic for people. Catalogers must judge the importance of names and titles appearing on the document and decide which should be chosen as access points and which should be disregarded. This is where experience plays a crucial role. Here the cataloger finds the organization and layout of the document helpful, whereas an expert cataloging system finds them problematic.

**DOCUMENT CHARACTERISTICS**

There are linguistic characteristics most people use without much thought, but an expert system must be programmed to recognize them. Many catalogers know that terms like "Associate Professor of Clinical Medicine" appearing under a name is not a title of the book but rather a person's affiliation. People familiar with American culture know that "Washington" can refer to a city, a U.S. state, or a personal or corporate name. These people can easily distinguish through context which "Washington" is meant. The challenge to the expert systems is to interpret that context properly.

Studies of visual and linguistic characteristics of documents are extremely important in building expert systems. Thirty years ago, Kilgour (1969) suggested the automatic extraction of bibliographic data from documents. Researchers of this operation still encounter difficulties of technological limitations similar to those researching prototype expert systems. Although the optical reading of catalog cards for their conversion into database records has had good results, OCR equipment is not yet sufficiently developed to create catalog records from original documents. Two characteristics of documents add to the task's difficulty. One is the indeterminate nature of language, which poses relatively little problem to speakers of the language, but is difficult to formalize for a computer system. The other is the graphic design of documents. Although design can be diverse, catalogers expect that certain types of data will usually appear in certain places. Understanding the visual and linguistic structure of documents may, therefore, be necessary in the development of cataloging rules, if they are to be used effectively in an expert system. Rules derived by generalizing characteristics observed in documents can ensure more homogeneous interpretation of data in documents.

Jeng (1986) suggested that title pages tend to have structure, which could be used for automated data recognition. She defined a "block" (one or more words separated from other blocks by blank vertical space) as a basic unit of analysis. She found that blocks have two characteristics: physical appearance and content. Physical appearance is described with type face and size. Content is the meaning of the text. Jeng (1987, 1988) later presented a study of linguistic characteristics on a sample of 203 title pages in which she identified the occurrence of words, phrases, and punctuation marks. In her sample, an average of 23 words appeared on a title page. The most frequent types of words used were proper names (36%) and common nouns (26%). Other syntactic elements that appeared were prepositions (11%), initials (initials in names or abbreviations) (7%), articles (6%), adjectives (5%), conjunctions (4%), and verbs (4%). She examined the occurrence of these elements in different blocks on title pages (e.g., titles, other title information, statement of responsibility, publisher).

Jeng (1991b) continued her research on title pages, concentrating on their visual characteristics. On the basis of frequency of bibliographic elements, she proposed a three-level prototype title page, with an appropriate location and sequence of bibliographic elements on the page. The first level contained the title proper, author and publisher information. The second level added other title information and place of publication, with the third level adding author affiliation and year of publication. Based on her findings, Jeng concluded that because cataloging rules were not derived from a systematic knowledge of title pages, the interpretation of title page information is
more dependent upon a cataloger's interpretation and experience. She suggested that her research into the composition of title pages has implications for both the future development of cataloging rules and expert cataloging systems.

Weibel, Oskins, and Vizine-Goetz (1989) also investigated the possibility of automated recognition of data on the title page and tested Jeng's findings (Jeng 1986, 1988). They simulated OCR techniques using locally produced electronic versions of title pages. This simulated OCR technique was necessary because contemporary OCR devices were not able to perform at the level that the researchers required. Their goal was to identify automatically the following data from the title page: title, other title information, statement of responsibility, edition statement, publisher, place, and date of publication. Their basic unit of data was represented by a "token," which was a space-delimited character string such as name or title (what Jeng called a "block"). Tokens had attributes representing position, size, typeface, and the case class of each character string. Case class reported upper- and lower-case characters, numbers, punctuation, symbols, and blank spaces. After the system identified the data, a confidence level was calculated, and the data were assigned to appropriate fields.

In their sample, 45 records of English-language monographs were used as training data to adjust and improve upon the theoretical assumptions for the system. Forty-six similar records were used to test the system. Their results showed that identification of publisher, place and date of publication was far better than the identification of other bibliographic elements. For these bibliographic data they achieved 80% correct identifications. This success was attributable to being able to compare publication related data to external files. For example, publisher names could be compared to an authority file of publishers. They found that it was much more difficult to identify titles, other title information, and statements of responsibility correctly.

It seems likely that an expert system for optical recognition of bibliographic data will require human verification of the results. Catalogers will continue to need to master the identification and interpretation of bibliographic data from documents. Places where this information is likely to appear are frequently, but not always, prescribed in the rules. However, often document designers or publishers don't follow standard practices in the placement of data.

Another suggestion derived from the development of prototype expert systems is the study of catalogers' perception and understanding of document layout. Ercegovac (1990) addressed this in his attempt to identify the characteristics of documents for identification of access points the producer of the map. The use of document characteristics has been found to be very useful by catalogers. That use, however, has not been incorporated in the cataloging rules. How do students and novice catalogers learn to use these clues? Most likely, experienced instructors and experienced catalogers impart this information on using the clues as well as the use of the cataloger's general knowledge to create a complete catalog record.

**CATALOGING EXPERTISE**

These expert cataloging system research efforts illustrate some of the difficulties of capturing the skill and knowledge required to catalog. Cataloging has been found to be so complex that it needs to be studied as a range of different subtasks. The details of these varied tasks need to be put together as pebbles in the mosaic of cataloging expertise. A similar conclusion was made by Jeng (1992a). She asserted that the understanding of the role of cataloging expertise (the general and personal knowledge of expert catalogers) needs to be explored. In addition, a larger knowledge base (with transformed rules) and better user interfaces must be developed.

Ercegovac (1990) was the first to discuss the incorporation of personal knowledge from expert catalogers and specific document characteristics into an expert
system for descriptive cataloging. For Mapper, she interviewed experts to devise additional rules for the knowledge base to supplement situations in cases where AACR2R provided insufficient detail. She studied the characteristics of printed maps and interviewed experts about common characteristics of these maps in order to create a knowledge base.

In a later paper, Ercegovac (1992) discussed observation techniques she used in the development of Mapper. Interviews and surveys were used in addition to the careful study of the characteristics of maps. The first two approaches were designed to elicit details about the data necessary for map cataloging. For instance, questions were asked about definitions of authorship and publishing data, and a ranking of specific responsibility functions. This enabled her to understand this narrow field in detail and successfully imitate expert catalogers in a computerized system.

Jeng (1992b, 184) defined expertise as "the high degree of skills, dexterity, or knowledge of a specific subject area." In that study, in which she studied cataloging expertise and the transfer of that expertise, she interviewed the head of the cataloging department and three professional catalogers at National Agricultural Library (NAL). Documentation related to the cataloging process and job descriptions was also gathered. Jeng concluded that people doing the lowest ranked tasks have the least expertise, that transferring expertise is well formalized in NAL, and that informal cooperation among catalogers was also common. Jeng emphasized that learning from questions that arose during regular workflow is the most common method for transferring expertise. This method would likely prove to be an obstacle in building expert systems.

Ercegovac (1992) found that a focused interview with a detailed schedule was a better source of information about the cataloging process than verbal reports and associated protocol analysis. The focused interview is less biased by the weaknesses of human memory. In her work, exact questions could be asked about authorship and the characteristics of specific maps. One wonders how difficult it would be to answer in an exact way the questions about the cognitive processing of how and why someone decided to follow a detail on a map for determining authorship when that detail is not mentioned in cataloging rules.

Jeng (1997) found verbal reports and protocol analysis for understanding cataloging expertise to be a successful and appropriate method for acquiring an understanding of cataloging expertise. In that project, she studied the kinds of knowledge and skills catalogers needed, what tasks were involved in cataloging, and what strategies catalogers used to solve specific problems. She used multiple methods: (1) a questionnaire, (2) a self-administered verbal report of cataloging sessions, (3) verbal protocols of sessions in which catalogers were asked to "think aloud" during the process and the researcher observed the process, and (4) verbal protocols of training sessions in which the researcher was taught by a senior cataloger. Jeng did not present the result of this research, but she illustrated her research methods with examples of the work process of two catalogers. Workflow was found to be influenced by the workscreen. One cataloger’s problem-solving strategy was to leave a difficult decision until the end of the task, while the other solved the problem when it occurred. Little explanation was provided in her report about the cognitive process of decision making. The two catalogers used standard cataloging knowledge for their work and problem solving techniques. The report of the catalogers’ work did not mention the use of knowledge beyond cataloging rules.

Ercegovac (1992) and Jeng (1997) suggest that knowledge of cataloging rules and procedures—as well as knowledge of cataloging systems—are necessary. Some specific knowledge, which is not written in cataloging rules or policies but is part of an oral tradition among catalogers in the same environment, is also necessary. This knowledge needs to be supplemented with general knowledge. The unwritten rules and general knowledge that assist in problem solving needs further research.
Different strategies of employing this knowledge in solving cataloging problems are used by different catalogers.

The different aspects of cataloging expertise likely will require different research approaches. While self-administered reports might be useful in studying the mechanics of cataloging procedures, observations and teaching sessions might be more useful for obtaining reasons for particular steps in the procedures. While questionnaires might be useful for learning about certain concepts in the cataloging process, they are useless if catalogers are not aware of those concepts. Different documents are examined and described in bibliographic records containing different data. In the same way, different kinds of processes and categories of knowledge need to be examined and described differently. Catalogers need to employ their long tradition of categorization of data, enriched with related disciplines, on the cataloging process to develop mechanisms for systematic representation of their expert knowledge.

**SUMMARIZED IMPLICATIONS**

In this paper, we have reviewed 25 years of work and thinking in the area of the use of computers to create cataloging records. Particular attention was paid to the types of knowledge necessary for these systems and to the problems with contents of the knowledge base. Have we found any clues about the missing pebbles for the mosaic of the cataloging expertise? We believe we have. Here are some clues found in the literature we reviewed:

- Comprehensive cataloging systems (i.e., those that cover all cataloging operations and all document formats) are too complex to develop either in research projects or for work environment.

- Reporting successes and failures in the knowledge acquisition for these prototype systems has been as important for the development of the discipline as successes and failures in actual development and implementation of the expert systems themselves.

- Some cataloging tasks seem to be more amenable to automation than others. Such tasks seem to have routine procedures. Standardization of such procedures could aid in the development of expert systems for the automatic generation of the content of the cataloging records. Such standardization would also be beneficial for the education of the students and new catalogers.

- AACR2 has been found to be problematic primarily because: (1) of the inconsistent logical structure of the rules; and (2) the information required to make cataloging decisions is missing in the rules. Difficulties arising from those problems are troublesome not only for the expert systems, but also for human experts.

- Characteristics of documents are important in developing expert systems that could automatically derive bibliographic data from the document. These same characteristics are useful to human catalogers. Researchers have found some patterns in the characteristics that would help both.

- Systems and people are fairly successful in recognizing personal names. The difficulty lies in deciding about the role of a particular name in the generation and production of the document.

   We wish to add some more reflection on each of the six clues for the missing pebbles by returning to each individually.

   Initial optimism that theoretical research in building expert systems would help with the complex task of descriptive cataloging dissolved quickly under the disappointing performance of prototype systems. These complex and comprehensive systems were generally not successful. Researchers then focused on smaller, narrower areas, in which more understanding and less variability existed. This resulted in some success.

   Few researchers who built prototype expert systems have reported on actually eliciting knowledge from human experts, but they have noted that rules can be appropriately interpreted only by expert catalogers. Information contained in
cataloging rules has been shown to be insufficient by themselves for making cataloging decisions. Rather, catalogers base many decisions on information acquired through experience. Because this information is subjective and varies at different levels of expertise, cataloging results also vary. This subjectivity may result in inconsistent catalog records and calls for greater standardization.

One approach toward improving consistency could be the standardization of cataloging procedures for particular types of documents. It seems reasonable to believe that some types of documents consistently require the use of certain procedures. For example, some items require the use of certain sequences of rules to be consistently applied. These procedures may involve use of some external knowledge. If these cases could have the external knowledge captured, they might be amenable to formalization in a knowledge base. Several benefits could result. First, more consistent decisions and more homogeneous catalog records could be provided. Second, new catalogers could study and understand in detail all the steps and details in the process. A third benefit would be the storage and documentation of the expertise.

Problems using AACR2 as the source of knowledge have lead to the investigation of cataloging rules. The structure of rules, their relationships, and the function and use of individual rules and chapters have been studied by a number of researchers. Each has concluded that cataloging rules were not particularly amenable for immediate computerization but could, and should, be made more adaptable for implementation in a knowledge base. This becomes particularly important as the use of computers to enhance productivity and accessibility increases and demands to reduce processing costs rise.

The need to implement cataloging rules in a knowledge base is not the only reason to improve the rules. It is clear from prototype system development that there is some subjective interpretation of cataloging rules because of missing information. More systematic and consistent rules could result in easier work for catalogers, who could spend less effort on inferring the missing or contradictory information. More consistent catalog records could result if there are fewer individual decisions. These are also areas that would require more emphasis in the education of new catalogers. Instruction in the use of cataloging rules should be complemented with an examination of inconsistencies and gaps in the rules. Strategies commonly used by expert catalogers could be used to illustrate problems and help novices understand the kind of external knowledge they need to employ in certain situations. Library and information science teachers might then better prepare students to catalog in real environments.

Some researchers have studied the characteristics of documents and experimented with automatic recognition of data from the document. In their work, they all made the assumption that future OCR technology will be able to correctly read title pages regardless of graphic design variations. Some also suggested that the electronic version of the printed title page should be accepted as the primary source of information. Results in this area have been quite encouraging. More than half of personal names, corporate names, and title access points could be automatically assigned in the English-language monographs that were tested.

Two major problems exist with the accurate automatic identification of names: (1) names need to be included in the knowledge base and (2) types of responsibilities need to be distinguished. These two areas also need to be mastered by beginning catalogers. While knowledge of names mostly comes from the cataloger's personal knowledge, types of responsibilities are learned in cataloging courses and through experience. Knowledge of which names are associated with the creation of a document and where they appear in a document is one of the basic skills catalogers need to master. It is one closely associated with the knowledge of visual characteristics of documents.

Understanding visual document characteristics might help with cataloging when catalogers are not familiar with the culture and language of the document, provided
that these visual clues are not culturally based. Identification of these characteristics could help catalogers find and interpret the necessary bibliographic data.

The research we have reviewed here has covered cataloging expertise and the cataloging process. Understanding the decision process is as important as understanding sources, rules, or building cataloger-friendly interfaces on workstations. Unfortunately, these researchers seem to have contributed little to the real working environments and procedures of catalogers. While significant progress has been made in the development of cataloging workstations, which make multiple cataloging tools available for simultaneous consultation on the computer screens, they continue to leave the challenging and difficult intellectual work entirely to the cataloger.

**Conclusions**

From the studies presented one can conclude that catalogers invest significant intellectual effort in the interpretation of cataloging rules and in the interpretation of document characteristics. They learn from their own experience and from experienced colleagues how to master these two tasks. The interpretation of the rules likely depends both on the general knowledge and the expertise of the cataloger. Expert knowledge comes from formal education and interaction with colleagues. A culture of sharing expertise among colleagues has great value here. The general knowledge of catalogers is the knowledge that is not directly connected to their profession. It is gained through education and social interactions in the community. Although these aspects were not studied, a systematic approach to the individual interpretations of the rules might result in easier work and in more consistent decisions. Consistency and clarity of catalog records is, after all, the main goal of all catalogers, as they try to provide the most relevant and accurate information about sources to the users.

What must a student interested in entering the cataloging specialty learn to begin the journey to becoming an expert cataloger? The research in prototype expert systems has demonstrated that knowledge of the cataloging rules is not sufficient. It was shown that there are several missing pebbles from the mosaic of cataloging expertise. The ability to interpret cataloging rules is one important area. An example of this form of interpretation is represented to a limited degree by *Library of Congress Rule Interpretations (LCRI)*. In addition, a high degree of experience and common sense has to be employed to make the necessary judgments called for by the rules.

Another area to be addressed is the sequence of the rules in the system. Certainly there are many documents that follow the same rule pattern. Some of these patterns are taught in cataloging classes. Others, less common, are only mastered with experience, while a few always require innovative approaches to rule application.

Cataloging workstations currently utilize relatively rudimentary knowledge technology. For documents, which follow certain patterns, these workstations could assist by providing templates, such as inferring the publisher from the ISBN, as suggested by Davies (1987). Molto and Svenonius (1998) proposed an interface for the online version of *AACR2R* which would make the use of this basic tool easier. Perhaps the electronic version of *AACR2R* will begin realize this hope. Certainly linking the *LCRI* and *AACR2R* within these workstations would be a useful tool for catalogers. Again, the Cataloger's Desktop from LC offer some hope for optimism.

Access points are essential in the efficient retrieval of documents. The identification of names, titles, and functions as well as ranking their importance is necessary. Experiments have demonstrated that many access points and their proper headings for personal names could be automatically generated. These decision criteria do not come easily. Instead they are the result of multiple encounters with cataloging rules. Some criteria are suggested in *AACR2R* and *LCRI*, while others are not.

Essentially, students aspiring to the cataloging specialty cannot reasonably be expected to be taught all the solutions to
problems they may encounter. Rather, their instruction, in addition to providing the basic building blocks of cataloging (i.e., rules), must concentrate on developing problem solving skills which will allow them to enhance their cataloging experience base.

Catalogers have demonstrated that they can, and want to, use computers in the best ways possible. At the moment, however, some aspects of their work are more amenable than others to the employment of this tool. This does not mean, however, that we should not continue to try to understand and formalize the cataloging process further. Knowledge acquisition methods offer a systematic inquiry into the cataloging expertise. A thorough understanding of the mental processes involved in cataloging offers not only a tool for developing expert systems, but also for rationalization of our own work. Systematic processes are easier to learn and the transition from novice to an expert is therefore smoother. Such approach should appeal to the catalogers, who are traditionally regarded as systematic organizers of materials and information, and to all those who are concerned with the quality of library catalogs.

Considering the research and developments to date, it is apparent that the expert knowledge base continues—and will likely continue—to reside with the cataloger. The new tools that might reduce the cataloger’s other burdens are not likely to reduce the need for human expert intervention in the cataloging process. This expert knowledge continues to be the domain of the cataloger, who provides the missing pebbles for the mosaic of cataloging expertise.

Works Cited


———. 1987. The title page as the source of information for bibliographic description: An analysis of its visual and linguistic characteristics. Ph.D. diss., University of Texas, Austin.


Problems in the Establishment of Nonunique Chinese Personal Headings with Special Reference to NACO Guidelines and Vendor-Supplied Authority Control

Hsi-chu Bolick

Current vendor software for authority control is found to generate negative results for nonunique Chinese headings in the local catalog. After authority control, nonunique Chinese names in the bibliographic records are found routinely altered to match authority headings established in the Library of Congress (LC) authority file that are unrelated to the Chinese script of the bibliographic record. Headings with standard diacritics in the Wade-Giles romanization scheme are the most problematic. The name headings that are negatively affected in the authority control services need to be corrected before the LC Pinyin conversion project takes place around the year 2000 to prevent headings from being distorted further in the Pinyin conversion process. In this paper, I examine the reasons why vendor software produces negative results for nonunique Chinese names, and suggest measures for Chinese-Japanese-Korean (CJK) libraries in North America and vendors who supply authority-control services that include CJK data to improve the situation. These include vendor software upgrades, modifications to CJK Name Authority Cooperative procedures, etc. The authority control service at the University of North Carolina at Chapel Hill is used to illustrate the problem.

In the fall of 1997, the Library of Congress (LC) announced to the East Asian library community its decision to adopt the Pinyin system for the romanization of Chinese. The planning for this adoption includes converting LC's current Chinese bibliographic data from the present Wade-Giles transliteration system to the Pinyin system around the year 2000. The conversion will use an approach similar to the one the National Library of Australia took in converting its CJK (Chinese-

HSI-CHU BOLICK (bolick@email.unc.edu) is acting East Asian Bibliographer/Cataloger, Collection Development Department, Davis Library, University of North Carolina at Chapel Hill. The author wishes to thank Margaretta Yarborough, Joe Collin, and Karl Lo for reading an earlier version of this article. Manuscript received September 14, 1998; accepted for publication November 17, 1998.

While the East Asian library community is planning for the adoption of the new standard, I will address some problems in the conversion of nonunique personal name headings that should be taken into consideration in planning for Pinyin conversion in order to ensure Chinese bibliographic data integrity.

The use of a vendor-supplied authority control service at the University of North Carolina at Chapel Hill (UNC), under the currently available software program, generates negative results for nonunique Chinese transliterated name headings in bibliographic records. Nonunique names in bibliographic records routinely are altered to the authority headings established in the LC authority file that are unrelated to the bibliographic records (see figure 1). Headings with standard dia
critics in the Wade-Giles system are the most problematic. Although, in a technical sense, the retrieval of such headings is not affected, the presence of multiple authors under one romanized form presents two problems: retrieval sets high in recall but low in precision; and mispresented names in the local catalog that, if not corrected manually, will be distorted further in the process of Pinyin conversion when it takes place in a few years.

**Nonunique Chinese Name Headings**

The term “nonunique name” in the LC authority file refers to the use of one established heading for multiple people who cannot be distinguished from one another by the form of their names alone. Nonunique name authority headings occur when people share identical names, and additional personal data (such as birth and death dates or other qualifiers) are not available at the time the heading is established. It is common to encounter this type of heading in LC’s authority records for all languages.

Without qualifiers to the personal name, every Chinese name transliterated under the Wade-Giles scheme is subject to becoming a nonunique heading because of at least three factors: the nonapplication of the tonal value in transliteration practice; the use of diacritics in the Wade-Giles transliteration scheme; and Name Authority Cooperative (NACO) authority practice and normalization.

**THE NONAPPLICATION OF THE TONAL VALUE IN TRANSLITERATION PRACTICE**

National standard Chinese is rich in homonyms, and is a tonal language. By applying one of the basic four different tones to a Chinese syllable, a variety of unique Chinese characters is generated. But in North American cataloging practice, according to the current version of the ALA-LC romanization tables, the tonal values are ignored in transliteration of Chinese scripts (Library of Congress 1991). This practice groups Chinese characters that are pronounced with the same sound but different tones under one transliterated form. For example, combinations of tonal numerical marks 1, 2, 3, 4 and the roman letters ta in ta1 ta2 ta3 ta4 represent at least four different Chinese characters. In current LC practice, however, these four characters are transliterated into one ‘ta’ in the authority and bibliographic records.

**Diacritics in the Wade-Giles Transliteration Scheme**

Structurally, the Wade-Giles transliteration scheme doesn’t employ a unique roman letter to correspond to each individual Chinese phonetic sound. Instead, it uses combinations of roman letters and diacritics to stand for some distinctive phonetic values. There are four diacritics—a, umlaut, circumflex, and breve—in the original Wade-Giles scheme (Lu 1995). The modified Wade-Giles transliteration scheme currently used by LC and East Asian libraries in the U.S. uses two diacritics, ayn (’) and umlaut ("), to indicate aspiration and the middle vowel sound. For example, p’t’k’ represent aspiration of p t k, while u is the middle vowel sound to be distinguished from u. Chinese characters that are transliterated into identical roman letters but
LC Authority Record for Chang, Y"u

ARN: 768061
Rec stat: c Entered: 19820520
Type: z Upd status: a Enc lvl: n Source:
Roman: ■ Ref status: a Mod rec: Name use: a
Govt agn: ■ Auth status: a Subj: a Subj use: a
Series: n Auth/ref: a Geo subd: n Ser use: b
Ser num: n Name: b Subdiv tp: Rules: c

1 010 n 82071276
2 040 DLC +c DLC +d DLC =d DLC-R
3 005 19940428061253.4
4 100 10 Chang, Yu
5 400 10 Zhang, Yu
6 400 10 Zhang, You
7 400 10 Chang, Kung-li
8 500 10 Chou, Heng
9 670 [Editor of Y"un Tai-y"ing lai hung ... ]
10 670 Y"un, T.Y. Y"un Tai-y"ing lai hung ch'u yen lu, 1981
11 670 [Author of Hsiao tang chia]
12 670 His Hsiao tang chia, 1969: +b t.p. (Chang Y"u)
13 670 [Author of Wen i ti jen wu chi ch'i t'a]
14 670 His Wen i ti jen wu chi ch'i t'a, 1953: +b t.p. (Chang Y"u)
15 670 [Author of Hsin k'an ching pan ...
16 670 His Hsin k'an ching pan pi pi p'ing cheng pai chang chuan, 1643: +b caption (Chang Y"u)
17 670 [Author of Feng hsia]
18 670 His Feng hsia, 1975: +b t.p. (Chang Y"u)
19 670 [Joint author of T'ai-tzu-t'an]
20 670 Wei, M. T'ai-tzu-t'an, 1954 (a.e.) +b t.p. (Chang Y"u)
21 670 [Author of Huan ching mi li]
22 670 [Author of Chieh chung]
23 670 Her Chieh chung, 1985: +b t.p. (Chang Y"u)
24 670 [Author of Chang Y"u hsiao shuo hs"uan]
25 670 His Chang Y"u hsiao shuo hs"uan, 1985: +b t.p. (Chang Y"u)
26 670 Her Huan ching mi li, 1983: +b t.p. (Chang Y"u) vita (Her real name is Chou Heng)

Bibliographic Record in DRA Database Before Vendor-Supplied Authority Control


Type: a Bib l: m Enc l: I Desc: a Ctry: hk Lang: chi Mod:
Srce: d Ill: Audience: Form: Cont: Gvt: Cnf: 0 Fst: 0
Ind: 0 Fic: 1 Bio: Dat tp: s Dates: 1995 Control:
005; ;a 19960410171011.0 $ 
010; ;c 32376017 $ 
020; ;a 9624474249 (pbk.) : $c HK$38.00 $ 
040; ;a SFR $c SFR $d JTW $d NOC $ 
066; ;c $1 $ 
090; ;a PL2837.Y76 $b C48 1995 $ 
092; ;a P $ 
100; 1:6 880-01 $a Chang, Y"u. $ 
245; 10;6 880-02 $a Chuang feng pan sha / $c Chang Y"u chu. $ 
250; 6 880-03 $a Ti 2 pan. $ 
260; ;j 880-04 $a Hsiang-kang : $b Ch"in in yuan ch'u pan she, $c C 1995. $ 
300; ;a 207 p.; ;c 17 cm. $ 
490; 1:6 880-05 $a Chang Y"u hs"uan huan hsi lieh ; $v 18 $ 
500; ;a Short stories. $ 

Figure 1. Example of Authority Verification Process on a Nonunique Chinese Heading.
(Continued on next page)
with different diacritics or without diacritics, such as p and pi, t and ti, k and ko, and so on, stand not only for different pronunciations within each pair but different meanings as well. This design in the Wade-Giles scheme generates a high volume of unrelated Chinese characters that have one identical roman form except for the distinguishing diacritical marks; for instance, ch’un, chin, chun; chu’an and chuan; yu and yü, etc.

NACO AUTHORITY PRACTICE AND NORMALIZATION

Authority records contributed to the LC authority file are created by LC and NACO participating institutions. When creating authority records for the LC authority file, NACO catalogers must follow a set of LC guidelines to ensure the integrity and quality of the large shared file. Rules of normalization are one set of the guidelines that catalogers apply to avoid creating duplicate headings.

Normalization is a process that is designed to eliminate all but the essential characters of a heading for the purpose of comparison (Cataloger’s Desktop 1998). Normalization programs do this by removing all diacritics and punctuation and converting all letters to upper case (see figure 2).

Under this normalization rule and
NACO guidelines, a new heading is created only when the normalized form does not match the normalized form of an already established heading in the LC authority file. If its normalized form is found in the authority file, the heading is considered established. When applying normalization to Chinese transliterated names, diacritics, which are essential for distinguishing Chinese syllables under the Wade-Giles system, are considered not essential and therefore are eliminated from the heading comparison process. If headings are found normalized to the same form, they are put together in one authority record (see figure 3).

Under NACO practice, the Chinese names in example 1 of figure 3 share one authority record. The authorized heading for these two names will either be Cheng, Ch'en, or Ch'eng, Chen, depending upon which name gets created first in the authority file. Likewise, the Chinese names in example 2 of figure 3 have one authority record, and the established heading takes the diacritics associated with the name first created in the authority file.

When dealing with a new nonunique name whose normalized form matches an established heading in the authority file, NACO catalogers do nothing to the established heading but revise the existing authority record in the note area (670 fields in the authority record) to include new information on additional names and their related citations. New reference headings for different names are added to the authority record when appropriate. If the existing authority record is coded as a unique name, it is changed to a nonunique name record when new names are found that share the same normalized form.

For instance, in the above examples, when the new name Cheng, Ch'en appears in the cataloging source and Ch'eng, Chen is found already established as an authorized heading in the LC authority file, Cheng, Ch'en is only added to the note area. Ch'eng, Chen remains as the established heading for this record. Cross-references related to Cheng, Ch'en can be added to the record as long as they do not normalize to the established form.

If unaware of this practice for nonunique names, one is likely to interpret Ch'eng, Chen as it appears in the authorized heading field of the authority record, as the authoritative form for Cheng, Chen. Likewise, one might take Chang, Ch'uan, if established first, as the established authority heading for Chang, Chuan and Cheng, Chuan (see figures 1 and 4).

CATEGORIC FOR NONUNIQUE CHINESE NAMES

As a rule, catalogers choose LC-established headings over name forms found in the cataloging source for heading forms in bibliographic records to control quality and uniformity of access points; yet in creating bibliographic heading forms for Chinese nonunique names, this is not always the case. In the example of the heading Ch'eng, Chen, when the new name Cheng, Ch'en appears in the cataloging source, for which a nonunique authority heading is found established under Ch'eng, Chen, instead of using the established heading Ch'eng, Chen, the cataloger enters Cheng, Ch'en as the author's name in the bibliographic record to

---

1. a. Catalog form: Cheng, Ch'en
   Normalized form: CHENG, CHEN

b. Catalog form: Ch'eng, Chen
   Normalized form: CHENG, CHEN

c. Catalog form: Chang, Ch'uan
   Normalized form: CHANG, CHUAN

2. a. Catalog form: Chang, Ch'uan
   Normalized form: CHANG, CHUAN

b. Catalog form: Ch'ang, Chuan
   Normalized form: CHANG, CHUAN

c. Catalog form: Chang, Ch'uan
   Normalized form: CHANG, CHUAN

---

Figure 3. Normalization in Chinese Name Headings.
reflect the Chinese characters.

This practice is necessary in cataloging Chinese materials for the purpose of appropriately matching roman data to the corresponding vernacular data in the Machine-Readable Cataloging (MARC) record. The current negative authority control result found on nonunique Chinese names occurs because available vendor software does not take into account this special cataloging practice for nonunique Chinese names.

VENDOR-SUPPLIED AUTHORITY CONTROL SERVICES

In 1995, three vendors, Library Technology Inc. (LTI), Blackwell North America (BNA) (now OCLC Authority Control Services), and the Washington Libraries Network (WLN), were invited to UNC to demonstrate their authority services. During their presentations, LTI, BNA, and WLN displayed similar approaches and software designs in processing authority control for local files.

LTI was chosen to provide vendor authority control for UNC. Its procedure for authority control is as follows (LTI 1998):

First, headings are “normalized” for spacing and punctuation to increase the probability of a link between a catalog record heading and an authority record heading. Next, the normalized headings are compared against a comprehensive index of authorized and variant-form headings generated from authority records. When a match occurs, a link is created between the authority record heading and the catalog record heading. If the catalog record heading matches a see reference in an authority record, the catalog record heading is replaced by the heading in the LC authority record.

Existing authority control software lacks special instructions that would account for the special cataloging practice for nonunique Chinese names. All nonunique Chinese names in bibliographic records that are normalized in the same way as authority headings (but that differ from them in diacritics) are treated as incorrect forms, and thus are changed to authority headings. As a result, after vendor authority control processing, one finds the roman data doesn’t match the vernacular data for nonunique names in the bibliographic record.

Taking again the example of Ch‘eng, Chen, under the current vendor software process, because Ch‘eng, Chen is the established heading, all occurrences of Cheng, Ch‘en will be changed to the authority heading Ch‘eng, Chen in the catalog after the authority control service.

LTÍ, BNA, and WLN all agreed that they have this problem in their services. It appears that on the current market there are no available commercial vendors who offer authority control services that take this situation into consideration.

Figure 4. The Unique Name Authority Record (See Name: a in the Fixed Field).
Normalization, the current industry standard for database management—which by design aims for effective data storage and management—requires that the relationship between the related elements (the primary key and the data) is one-to-one or many-to-one, but never one-to-many or many-to-many (Dutka 1989). That is to say between personal names and authority records, normalization functions effectively when one authority record relates to one unique personal name, or one authority record stands for one unique person with multiple names, but not vice versa when multiple unique personal names are represented by one authority record.

The application of the Wade-Giles scheme in current cataloging and NACO practice for nonunique names shows a breakdown with the principle of normalization. Diacritics used in the Wade-Giles scheme to represent the uniqueness of sounds and words are considered not essential in the existing normalization rules and thus are excluded from the process. This explains why procedures that occur after the normalization process in authority-control services cannot process nonunique Chinese transliterated names properly.

The Wade-Giles scheme will soon be dropped for cataloging work, and the negative effect associated with it under normalization will be eased substantially when the new Pinyin scheme is adopted. Nevertheless, it will not disappear. Compared to the Wade-Giles scheme, the Pinyin scheme reduces the use of diacritics down to the umlaut for distinguishing certain unique Chinese characters. For instance, the Chinese character “green” is transliterated in Pinyin as lǐ, while the Chinese character “furnace, stove” is transliterated in Pinyin as lú. Under current normalization, lǐ and lú again appear as the same form. Therefore without software upgrades or modification in the transliteration scheme, the nonunique name syndrome in the authority control process will continue to happen with Pinyin data.

Before the year 2000 (when LC will start converting its database to the Pinyin scheme), libraries preparing to follow LC in converting their local files to Pinyin not only should undertake correction of all nonunique Chinese names that are affected negatively by the authority control process as described above but also should work on correcting diacritics in the bibliographic records to prevent data distortion in the conversion process.

The National Library of Australia (NLA) in 1996 successfully converted its CJK data from the Wade-Giles scheme to the Pinyin system with computer technology. After closely observing the process and the conversion result at NLA, LC announced that its Pinyin conversion project would be carried out using the same approach as NLA's.

Describing the CJK conversion project in detail, Groom (1997) wrote that one key element in designing the software program for Australia's Pinyin conversion project was the setting up of a conversion table that listed and correlated all valid Wade-Giles words with their Pinyin counterparts. With the conversion table in place, the conversion program was thus able to identify and subsequently convert all Wade-Giles words to their Pinyin counterparts when run against the local bibliographic file.

The conversion table was written precisely based on the Wade-Giles rules to map each Chinese character to its Pinyin counterpart, taking into account the exact use of Wade-Giles diacritics and their position in the syllables, etc. The conversion program carries out the conversion only when the bibliographic data match correctly to the conversion table. This is to say local bibliographic records submitted for conversion are required to have the same Wade-Giles forms and diacritics as the conversion table to guarantee accurate data conversion. Any spelling errors in roman forms or misused diacritics will cause incorrect conversion or no conversion at all.

Groom (1997) reported that conversion errors were found when records con-
tained unexpected diacritics. In some cases, diacritics were not input as prescribed in the Wade-Giles scheme, with the result that no conversion took place because a diacritic discrepancy occurred between the catalog entries and the conversion table. In other cases, diacritics were omitted in the NLA catalog entry, resulting in an erroneous conversion to another heading.

**Suggestions**

The inadequacy of vendor-supplied authority control of Chinese name headings demonstrates an immediate need for the East Asian library community to find solutions for the incompatibility between the normalization rules governing the setup of Chinese personal headings in the NACO program and the design of the standard transliteration scheme used in North America for transcribing Chinese. Through examination it appears that a direct method for resolving such incompatibility is simply to make all Chinese personal headings unique in the LC authority file. Use of unique headings not only allows Chinese names to be processed correctly under normalization rules but is effective for authority control. As a means of doing this and solving the current authority control problem, several approaches are possible: improving and developing new system software; completing NACO headings with dates; providing clear instructions regarding the use of Roman forms for authorized headings; and eliminating diacritics from the Pinyin scheme.

**Improving and Developing New System Software**

Before a permanent way is found to deal effectively with romanized Chinese names in the LC authority file, in the short term we need to request that a new algorithm be developed immediately in vendor software that can halt the incorrect replacement of nonunique Chinese names in the bibliographic records with authority heading forms. One way of updating the current vendor software for this purpose is to incorporate the NAN element in the fixed field of the USMARC Authority Format record into the new algorithm so that the software will perform global replacement of catalog forms only to those headings that are linked to unique name records. Headings that are linked to nonunique authority records should be marked and stored in a file for further processing. The review of unprocessed nonunique names should not be a heavy burden for catalog maintenance because most of these names deliberately are kept in those forms by human decision at the time of cataloging. Thus, the concern over the error rate for those names is small.

In the longer term, software must be developed that incorporates vernacular data in the normalization process. Adding vernaculars to authority records has been advocated by the East Asian library community to solve problems of differentiating Chinese names given the high volume of homonyms found in transliterated data in authority records (Yu 1996; Morimoto 1996).

To increase the uniqueness of Chinese names in the LC authority file, the new software should not only have the capability of storing and displaying vernacular data, but should also include a mechanism that takes vernacular data into consideration in the normalization process. That is, an additional procedure should be devised that would involve comparing vernacular data for headings that are identical when normalized. Headings that would otherwise be conflated could be kept separate when the vernacular data do not match. Authority control software would follow the same design for heading comparison and correction. This extra procedure would substantially improve effectiveness in controlling Chinese names and prevent false conversions during the authority control process.

A more cooperative effort must be made through the NACO program to improve Chinese name heading quality. Before any new updates are developed in the software, NACO catalogers and the NACO program can reduce or prevent
the normalization impact to Chinese names in the meantime by making the heading unique in ways such as: completing NACO headings with dates; providing clear instructions regarding the use of roman forms for authorized headings; and modifying the Pinyin scheme to eliminate the use of diacritics.

Completing NACO Headings with Dates

Adding birth and death dates to Chinese romanized names is very effective for establishing unique headings in the LC authority file. All NACO catalogers should be encouraged or required to provide such data to Chinese headings when available. To make this a common and practical practice in NACO work, not only should the NACO program be expanded to include more East Asian catalogers to share the burden and expense of such authority work, but a global effort should be made to solicit information from libraries, publishers that can provide such information through Internet, or author introductions in published books (Eastlib Discussion List 1998).

Providing Clear Instructions Regarding the Use of Roman Forms for Authorized Headings

Instructions must be provided about whether to use roman forms that appear in the source or author-supplied roman forms as authoritative forms instead of using systematic transliterated forms from the Chinese scripts. The Anglo-American Cataloguing Rules, 2d ed., 1988 revision rule 22.3C2, and its alternative rule, which direct the setup of Chinese headings in NACO practice are insufficient for catalogers to handle the increasingly common situation where, in the cataloging source, the romanized form of a name and its Chinese script coexist on the title page.

Due to a lack of guidelines in the NACO procedures for such circumstances, the choice of forms for these names varies among catalogers. Some headings are established under 22.3C2 and its alternative rule, using the most popular form of authors as known by readers and in reference sources. For some headings lacking reference sources or guidance in the NACO procedure, it is difficult to comply fully with chapter 22 to establish unique Chinese headings, and catalogers refer to Rule 1.0 for guidance, which prescribes transcribing the statement of responsibility of a work in the language and script in which it appears. As a result, the former rule usually brings the roman forms already supplied by the source as authority headings, while the latter rule results in headings that are systematically transliterated from the original scripts.

Roman forms systematically transliterated from Chinese scripts, as discussed earlier, frequently result in a high volume of homonyms, thus producing nonunique headings. If NACO guidelines clearly prescribe 22.3C2 and its alternative to take precedence over rule 1.0E in establishing authority headings in such situations, more unique name headings will result.

The author or publisher-supplied roman forms appearing in the source not only indicate the authors' choice of forms in their names, but are usually formulated less systematically and predictably than those generated according to the transliteration scheme. Therefore, they are less likely to become nonunique names in the authority file than names formulated by the transliteration scheme. Such practices also reduce the confusion of public catalog users who recognize authors' names in their roman forms from publications rather than the forms from the prescribed transliteration scheme. Figure 5 displays the difference between names transliterated with the Wade-Giles scheme and those supplied by authors or publishers.

Under the Wade-Giles scheme, Tung Chien-hua represents hundreds of different Chinese characters for names, but the combination of Tung Chee Hwa together, so far, only represents one personal name. Likewise, Ng Chi-sum, formulated under the Cantonese pronunciation, is drastically different from Wu Chih-sen, formulated under the standard Mandarin sound on which most transliteration schemes for Chinese are based.
ELIMINATING DIACRITICS FROM THE PINYIN SCHEME

The Pinyin scheme will no doubt work much better in normalization for Chinese name headings than the current Wade-Giles scheme. Not only does Pinyin allow each sound to be associated with a unique roman symbol, but it also uses fewer diacritics. The use of diacritics to distinguish a pair of vowels or aspiration in Chinese has made Wade-Giles very ineffective in normalization. Thus, the only diacritic introduced for use in the incoming Pinyin scheme—the umlaut—would need to be eliminated and replaced with an alternative expression in the scheme. Repeating the roman letter that the umlaut attaches to is one alternative to representing the sound with umlaut. For instance, \( \text{\textit{liu}} \) would be presented as \( \text{\textit{luu}} \). In the pair of words \( \text{\textit{liu}} \) (meaning “green” in Chinese) and \( \text{\textit{lu}} \) (meaning “stove” in Chinese) can thus be distinguished as \( \text{\textit{luu}} \) and \( \text{\textit{lu}} \) in normalized form. This practice is common in publications and in other transliteration schemes used in Taiwan, Japan, and in the Macintosh Chinese language kit, for example. To allow the access of Pinyin headings where standard umlauts are used, Pinyin entries with standard umlauts should be entered as cross-references in authority records.

Data processing for nonroman languages, especially Chinese, in areas of library services is by no means easy. But with careful design in software to incorporate Chinese vernacular data, adjustment to transliteration schemes, and cooperation among NACO catalogers to set up unique personal headings, I am confident that Chinese names can be processed as effectively as names in other languages in the online environment.

WORKS CITED


Notes on Operations

Classifying Newspapers Using Dewey Decimal Classification

Tamara J. Kuhn

Librarians facing the decision of whether or not to classify a newspaper collection must consider the basic research needs of patrons. Users primarily access newspapers by geographic location (more specifically, by city of publication), and by date. Librarians also must consider that newspapers are serials and that classifying newspapers presents the same challenges as classifying general serials. Multiple title changes and unique relationships among newspapers can complicate the application of a classification system to a newspaper collection. Keeping newspaper families together chronologically on shelves is important when locating titles. The Dewey Decimal Classification system, with a few modifications, can be used optimally to classify a newspaper collection. Knowing that classification can function as an efficient organization method for a newspaper collection makes a library’s decision of whether or not to classify easier.

The Illinois Newspaper Project (INP) moved to the University of Illinois at Urbana–Champaign (UIUC) in January 1996 to begin cataloging the library’s United States newspaper collection. As a part of the United States Newspaper Program (USNP) funded by the National Endowment for the Humanities, INP shifted its focus from cataloging the extensive Illinois newspaper collection at the Illinois State Historical Library in Springfield, Illinois, to cataloging a much broader United States collection in one of the largest academic libraries in the nation.

One major obstacle discovered while evaluating the UIUC newspaper collection was the need to classify newspapers as well as provide description and subject analysis. Existing rules and guidelines provided by USNP and CONSER governed the latter two elements, but classifying newspapers was an entirely new realm. Because INP is only cataloging United States newspapers, I will focus on classifying U.S. titles.

Newspapers as Serials

Newspapers are a specific form of serial, and the question of whether or not to classify a serials collection applies directly to a newspaper collection. Much has been written on classifying serials and periodicals, and authors debate the pros and cons of arranging a serials collection in call number order. In an article discussing...
whether or not to classify serials, Smith (1978, 372) wrote, "Classification is indeed a very basic library approach to materials, and it cannot be overlooked. Serials collections and their use are much too important to be given the short shrift. Classification is imperative." Newspapers need to be given the same consideration in library collections as other materials, so a library's decision to classify serials should also include newspapers.

Smith (1978) argued that serials should be classified for seven reasons, many of which are still relevant today:

1. Materials on the same subject should be kept together, regardless of type, to facilitate subject access and browsing.
2. Classification encourages broader use of serials by treating them like other materials.
3. Serials reference work will be aided by classification.
4. Sorting call numbers can more easily produce subject aids and guides.
5. Reshelving becomes a mechanical task, and shelvers are not required to know special organization rules.
6. Materials that have undergone title changes can remain on the shelves together.
7. Classification can help prevent overcrowding in the subject catalog due to successive entry cataloging.

Bross (1990) listed some benefits of classifying a periodical collection (178): "A classification system enables a library to keep volumes of the same periodical together on the shelves in spite of title changes." Newspapers are notorious for changing titles many times within a lifespan, and using a classification system to organize a newspaper collection is one way to manage title changes on shelves and in local catalogs. Bross also stated (178) that "another benefit is the assistance a classification number can give staff/patrons who may be faced with the problems of correlating catalog entries, spine titles, filing rules, and new cataloging codes." Newspapers often merge to become new bibliographic entities or combine in various ways to form different manifestations of an earlier title. Classifying newspapers can help staff and patrons navigate through shelves without having constantly to consult the catalog for exact titles and title changes. Finding newspapers becomes a matter of following a base call number instead of remembering all of the various titles within a newspaper's lifespan.

### Classifying Newspapers at UIUC

Patrons access newspapers primarily by place of publication. Many times a researcher does not have a specific newspaper title in mind to fill an information need, but instead knows that a newspaper from a given area likely will contain the needed information. Historians, for example, can utilize newspapers from different geographic locations to compare how an event was covered at the time that it occurred. Location and date are most important in this type of reference request. Basing classification on geographic location can play a vital role in organizing and accessing a newspaper collection to expedite a researcher's information need.

Arranging a newspaper collection geographically is logical in relation to the manner in which newspapers are requested by patrons. In many catalogs, the call number serves as an additional access point. When accessing the UIUC newspaper collection, searching the catalog by call number is a very important function because most materials are housed in closed stacks areas. Patrons who know a general call number can search that number and browse the catalog by shelf arrangement. This serves them as if they were browsing the shelves in search of other newspaper titles from the same geographic area. In collections that allow patrons to retrieve their own materials, arranging newspapers geographically expressed through classification will fulfill a basic research need.

Organizing title changes through classification facilitates browsing a newspaper collection. Patrons can find the classification number of a title in the middle of a newspaper family and use that number as an entry point to the shelves. Once patrons locate that call number on the
shelves, they can easily move from title to title within that family without having to know the intricacies of title changes.

The decision to classify the newspaper collection at the UIUC library was compounded by the mechanics of the local online catalog. The Full Bibliographic Record/Library Computer System (FBR/LCS), the online catalog that UIUC used at the time this article was written, required a call number to function properly. The catalog consisted of two separate databases and software programs; one for bibliographic records (FBR) and one for holdings records (LCS). The linking component in this system was the classification number, and without one, patrons could not move easily between bibliographic and holdings records. Because of this (in addition to the need for geographic arrangement), it became clear that classification was essential in providing local access to the UIUC newspaper collection.

**USING THE DEWEY DECIMAL CLASSIFICATION SYSTEM**

Librarians at the UIUC library classify the general collection using Dewey Decimal Classification (DDC), and the newspaper collection needed to fit into this scheme. Before evaluating the classification scheme in relation to organizing newspapers, a literature search revealed that only one article had been written concerning newspaper classification, and it dealt only with Library of Congress classification (Krissiep 1991). There were no previously published articles on using DDC to classify a newspaper collection to help guide the INP. Few articles have been written concerning newspaper classification because newspaper cataloging is relatively new to librarianship and few newspaper collections are currently classified.

The USNP, formalized in 1984, was the first coordinated nationwide effort to catalog newspaper collections. Rules specific to newspaper description and subject analysis were established to facilitate this program (Field 1986). Classification, however, was not a formal part of the USNP.

Before examining DDC’s application to newspapers, it was important to look at the existing organization of the UIUC newspaper collection. Like many newspaper collections, UIUC’s collection was arranged primarily by geographic location. The first level of geographic organization was by country. Then the newspapers were organized alphabetically by state, and then alphabetically by city within each state. The next element of organization was by title. A common title, which is a shortened form in popular use by the public that does not reflect title changes, had been assigned to most newspapers and alphabetized within each city. For example, the Chicago Sun-Times underwent six title changes (including one merger) between 1941 and the present; however, the paper is commonly known as the Chicago Sun-Times or the Sun-Times. All boxes of microfilm for the Sun-Times held in the UIUC newspaper collection were labeled as Chicago Sun-Times, without acknowledging the numerous title changes.

After the alphabetization of each common title within its city, newspapers were arranged chronologically. This organization system became the basis for evaluating DDC to see how useful the classification scheme could be in organizing a newspaper collection.

**EVALUATING DDC IN RELATION TO A NEWSPAPER COLLECTION**

The evaluation uncovered both strengths and limitations. The main strength of the DDC to classify a newspaper collection is its geographic organization capability in applying Table 2: Geographic Areas, Periods, Persons. It is possible to build a number expressing country, state, region, and county. Figures 1–4 illustrate the application of Table 2. Another DDC strength is that all general newspapers can be found within a small range of numbers. General newspapers from the United States are found in the 071 classification number. In an open stacks collection, this will facilitate browsing.

One limitation was found, but it proved to be crucial in applying DDC in a strict manner. In most cases, geographic
arrangement in the scheme only allows access as far as the county level within the United States. Larger cities such as Chicago or New York have their own number in Table 2, but almost all average- to small-sized cities do not. Because of this, the city of publication would be eliminated from the organization of the collection, and the final level of geographic arrangement for a U.S. newspaper would be the county. Patrons very rarely request newspapers by county, and frequently request them by city. For example, a patron is more likely to ask for newspapers from Springfield, Illinois than from Sangamon County, Illinois.

Traditionally, the cutter number is created based on the main entry of a catalog record, and for newspaper records the main entry is always the title, either in the form of the title proper or a uniform title. Using DDC in a strict manner and cuttinger traditionally would make the title more important in the arrangement than the city. For example, if a strict and traditional method of applying DDC and cuttinger were to be used to classify a newspaper collection, all newspapers from Cook County, Illinois outside of Chicago would be arranged by title with no city arrangement whatsoever. Hundreds of Chicago suburban newspapers would file alphabetically by title beneath the county arrangement. Skokie newspapers would be mixed alphabetically by title with Calumet City papers and so forth. However, newspapers from Chicago would be separated from other papers from Cook County, because Chicago has its own number in Table 2. No city access is uniformly available through strict use of DDC and traditional cuttinger methods.

MODIFYING THE TRADITIONAL

After finding this major stumbling block, it became apparent that DDC had to be modified to make it a sensible choice for classifying a newspaper collection. It was possible to use certain elements in a strict manner, though. The subject number 071 and the use of Table 2 were applied as instructed in the DDC schedules. As stated above, this application generally only allows state, region, and county arrangement.

The cuttinger method was modified to
make certain that the city of publication was included in the classification arrangement. Choosing a cutter number based on the newspaper’s place of publication instead of main entry (title for newspaper records) ensures that the city is a standard part of the classification number. This provides the city level of geographic arrangement that is so important to newspaper organization. With this modification, the classification number parallels the organization of the hierarchical place name added entry, which is a core element in newspaper cataloging. The first geographic area in this data element is country. The second area is state. The third area is county, and the last area is city.

Generally, the cutter number is based on the place of publication found in the imprint. A newspaper can change places of publication within its lifetime, in which case it is up to the cataloger to decide which place of publication to use for the basis of the cutter number. For example, the title *Times of Havana* began publication in Havana, Cuba, on February 4, 1957, but by April 20, 1961, it had moved to Miami, Florida. The UIUC library holds issues beginning on April 20, 1961, but nothing before that date. It was logical to cutter this title based on Miami, Florida, rather than Havana, Cuba, given that none of the library’s holdings represent the time period in which this title was published in Havana.

The next element in the classification number remained the same as in a traditional application. The workmark is assigned from the title of the newspaper. This provides the alphabetical arrangement by title within the city of publication that is generally expected in newspaper organization. Because so many newspaper titles begin with a city name (*Champaign News-Gazette* from Champaign, Ill.) or frequency (*Chicago’s Daily Inter Ocean*), a guideline was established to keep workmarks from becoming too long:
Generally, if a title begins with a city name or a frequency, we use the first letter of the first word then skip to the first letter of the second word. A workmark should not extend beyond three characters, if possible. Figures 1–2 provide examples of distinguishing workmarks.

Because newspapers change titles frequently, the last element in the classification number was developed to handle title change situations and to make sure that newspapers remained in chronological order. When a title change occurs, a numeric extension is added to the classification number of the earlier title following the workmark. The first time a title changes, a "1" is added to the previous classification number. The next time the title changes, a "2" is added, and so forth. This keeps titles together in chronological order even when publishers decide to change the title on the newspaper masthead.

In cases of mergers, absorptions, or other complex relationships, the designation or numbering system can act as a pointer. For example, when two newspapers merge to form another newspaper, a cataloger has three possibilities for classifying the later title. An extension can be added to the classification number of the first merger title, an extension can be added to the classification number of the second merger title, or a completely new classification number can be created for the newly formed title. In this situation, the numbering system can offer clues. If one designation is continued by the later title, and the other designation dies with its title, then it is best to continue the classification number corresponding to the title whose designation continues. Figures 1–3 illustrate this concept. On the other hand, if neither designation from the merging titles continue and a new numbering system emerges, it might be best to create a new classification number altogether based on the new title. Often these tough situations are best resolved based on the cataloger's judgment and what works best within a collection.

---

**Figure 5.** Family Tree for the News-Gazette from Champaign, Illinois.
Figure 6. Application of DDC, Modified Cutter and Workmark to Genre Newspapers.

Figure 5 shows a portion of the newspaper family tree for the News-Gazette from Champaign, Illinois. Figures 1 through 4 illustrate each title and its assigned classification number. Together they demonstrate the use of DDC, the modified cuttering method, workmark, and numeric extension to classify a newspaper set.

**GENRE NEWSPAPERS**

Not all newspapers carry only general news and serve a geographic area. Some newspapers cover specific topics and represent certain genres, such as religious newspapers or labor newspapers. In cases like this, the genre or subject orientation of the newspaper is often more important to researchers and should supersede the place of publication as the primary focus of organization. It is best to classify genre newspapers within the appropriate subject area from the DDC schedules.

Geographic coverage is still important in most topical newspapers, but the subject of the newspaper needs to be expressed through the base classification number. Whenever possible, use Table 2 to make sure that any geographic coverage is also expressed. This provides geographic organization to a genre newspaper collection. Using the same modified cutter, workmark, and numeric extension guidelines that were developed for general newspapers will keep the classification method consistent. It will also provide the basic geographic organization that a newspaper collection needs. Figures 6 and 7 illustrate two genre titles and their assigned classification numbers.

**OBSERVATIONS OF IMPLEMENTATION**

After the INP began classifying the UIUC newspaper collection, the Newspaper Library started shifting the microfilm collection into call number order. It would have taken much effort and resources to shift close to 75,000 reels of microfilm at one time, so the library staff decided to shift the newspaper microfilm reels for each state as INP finished cataloging them. This distributed the work effort of this formidable task and broke up the process into logical phases. Each state was organized into call number order, but the shelving locations for each state remained in alphabetical order. It is possible that the collection can be later shifted into strict call number order.

Because the Newspaper Library microfilm collection is housed in a closed stacks area, staff and student employees must retrieve all requested film for patrons. Now that the U.S. collection is cataloged and classified, patrons are able to find titles and call numbers more independently than before. Patrons can

**Jewish World (Fairview Heights, Ill.)**

| 296 | Judaism |
| .09 | Historical, geographic, persons treatment |
| 77389 | Number from Table 2 representing Saint Clair County, Illinois |
| F168 | Cutter number representing Fairview Heights (city of publication) |

Figure 7. Application of DDC, Modified Cutter and Workmark to Genre Newspapers.
provide the Newspaper Library staff with a call number to expedite film retrieval. By relying on call numbers, there is not as much room for error in retrieval as when microfilm was completely arranged alphabetically by geographic location. Staff, however, must know a title's state of publication before the call number can be used for locating the film. This sometimes slows the retrieval process. Arranging a newspaper collection completely by call number would alleviate this additional step.

Patrons do not always understand the concept of a newspaper title change and provide the Newspaper Library staff with a call number of a preceding or succeeding title to the title they want specifically. Staff and student employees know, however, that the call number can provide an entry point to the shelves for that newspaper family. They can rely on the extension numbers to move easily through title changes on the shelves without knowing the titles themselves.

Staff members have noticed a difference in the time it takes for student employees to reshelve microfilm reels. Employees do not have to pre-arrange microfilm geographically as much as before, because there is a more systematic arrangement provided by call number organization. There are also fewer incidents of microfilm being shelved incorrectly since the collection has been shifted into call number order, according to staff. Student employees are instructed to match the call number to the correct shelf location and then check the exact title against neighboring film reels. This provides a mechanism for accuracy when shelving microfilm.

CONCLUSION

Librarians facing the decision of whether or not to classify a newspaper collection must consider patrons' basic research needs. Users primarily access newspapers by geographic location, more specifically by city of publication, and by date. Geographic organization was the primary focus when evaluating DDC as the classification scheme to use for the UIUC newspaper collection. Applying DDC in a strict manner cannot express consistently the city of publication of a newspaper title, which is so important in meeting users' needs. Modifying the traditional use of the cutter number can overcome DDC's limited geographic expression.

Librarians also must consider that newspapers are serials and that classifying newspapers presents the same challenges as classifying general serials. Multiple title changes and unique relationships among newspapers can complicate the application of a classification scheme to a newspaper collection. Keeping newspaper families together chronologically on shelves is important when locating titles. It is crucial that patrons and staff have the ability to browse the shelves without knowing the intricacies of every title change that occurs in the lifespan of a newspaper family. The addition of a numeric extension to a base call number provides chronological order to the classification scheme.

DDC, with a few modifications, can be used optimally to classify a newspaper collection. Knowing that classification can function as an organization method for a newspaper collection makes a library's decision of whether or not to classify easier.

WORKS CITED

Smith, Lynn S. 1978. To classify or not to classify... Serials librarian 2: 371-85.
Library Gateway: Project Design, Teams, and Cycle Time

Karen Calhoun, Zsuzsa Koltay, and Edward Weissman

Librarians at Cornell University Library (CUL) launched the first systemwide integrated gateway to networked resources, services, and library information in January 1998. The system was created and introduced in just 17 weeks. The Library Gateway was launched in response to a confusing CUL online presence; the library was spending half a million dollars a year on online resources and devoting a substantial amount of staff time to provide access to them, but systemwide efforts were only loosely coordinated. In this article, we examine the design and implementation process that helped the project succeed, rather than focusing on the technology or the vision behind the gateway. Examining the gateway project as a case study of new product development, we identify and discuss critical success factors, including: starting with a clearly defined concept, the importance of buy-in, simultaneous development of different portions of the project, employing cross-functional teams, and seeking continual feedback. Working together on a project of this complexity and on such a fast time line was an important organizational learning experience that moved CUL a step closer to mastering the process of innovation.

The Cornell University Library Gateway (http://campusgw.library.cornell.edu/) (figure 1) represents the first integrated approach to providing Web-based information services to the Cornell community, and as such its introduction in January 1998 was an important step forward. Yet the process by which the gateway was implemented was equally important to the library, whose managers and staff continually are striving to deliver new services as quickly and cost-effectively as possible. It has become critical to know, from an organizational perspective, what enhances the ability to deliver new services in time, on time.

In this paper, we examine the gateway project as a case study of project design, teams, and rapid cycle time. While the technology of the gateway might not be transferable to other environments, the cross-functional process employed to introduce it can serve as a model for others. We favor not merely reconciling, but tightly integrating the work of technical service librarians, public service librarians, collection development librarians, and information technology specialists to solve mission-critical problems and to ensure the continuing success of the library.

Karen Calhoun (kscl0@cornell.edu) is Head of Mediation Services, Central Technical Services, Cornell University Library, Ithaca, New York. Zsuzsa Koltay is Public Services Librarian, Mann Library, Cornell University Library. Edward Weissman is Assistant to the University Librarian, Cornell University Library. Manuscript received October 22, 1998; accepted for publication December 11, 1998.
We are in a period of turmoil. Electronic resources have turned library practices upside down. The tried-and-true practice of organizing libraries into fairly independent functional groups (collection development, public services, technical services, and information technology) no longer yields acceptable results. For example, Duranceau (1997) compares the workflows of print and electronic, networkedserials and concludes, "The print purchase process is a short, straight garden path. The digital world is cyclical... communication, coordination, and team effort is required at almost every stage... [and] the purchase process is a long, complex, winding dirt road filled with potholes.” Schroeder (1997) describes the potential benefits of blurring territorial lines between public and technical service librarians. Martin (1996) emphasizes the importance of healing the schism between technical and public services, and argues that the electronic library provides a tremendous opportunity for public and technical service librarians to begin to share leadership roles.

**CONTEXT OF THE GATEWAY PROJECT**

In 1868 Ezra Cornell, a plainspoken inventor inspired by egalitarian ideals, founded “an institution where any person can find instruction in any study,” including not only the arts and sciences, but also applied technology and agriculture (Cornell 1999). Today, the 19 libraries of the Cornell University Library (CUL) system serve 7 undergraduate colleges and schools, 4 graduate and professional units, and 2 medical units. The hallmarks of this complex library system, which employs more than 500 people, are diversity, excellent collections, and a history of semiautonomy for unit libraries.

Until the late 1980s, CUL had a largely captive audience for its print collections of more than 6,000,000 printed volumes and more than 60,000 journals. Since then, the Internet has altered dramatically library users’ perceptions and preferences for obtaining information of all kinds. Based on the experience of many reference librarians and the findings of a campus survey, an important shift in atti-
tude has occurred: students tend to perceive the effort of using print resources as higher than finding the information online.

In the present dynamic environment, librarians in CUL unit libraries have committed themselves to keeping pace with technological innovations and with their users’ preferences for finding information. By early 1997, the 19 CUL libraries had mounted numerous separate “official” library Web sites to provide Web or telnet access to more than 900 online resources. The library was spending half a million dollars a year on online resources, and staff devoted a substantial amount of time to providing access to them, but system-wide efforts were only loosely coordinated.

Organizationally, CUL staff have tended to interpret their roles and responsibilities in two dimensions: (1) with respect to their function (collection development, technical services, public services, information technology) and (2) with respect to the library unit for which they work. The conventional framework for accomplishing large systemwide projects, such as the selection of a new library management system, has been to organize functional committees, with representatives from various library units, to look after each function’s and unit’s needs. Coordination and conflict management across functions and units have tended to occur at the administrative level.

THE PROBLEM

The result of the technical, service, and organizational environment was a confusing CUL online presence. Networked databases, numeric files, and full-text resources were available but often difficult for users to find. The library was providing three different approaches to these materials.

The first of these was Bear Access, a site based on Mandarin technology developed by Cornell Information Technologies, which presented a set of “launch pads” and “buttons.” The buttons represented both specific networked resources (which opened specific sessions or resources), and additional compilations of resources—either more launch pads and buttons or Web pages that provided links to specific networked resources. To discover and access a particular resource, however, users needed to know either the genre (e.g., abstract, index, electronic journal, catalog) or where in the hierarchy the specific pointer to the resource existed. Library staff in the unit libraries supporting the humanities, social sciences, and area studies tended to rely most on the Bear Access model for connecting users to online resources.

The second approach was the Mann Library Gateway, which was developed by the staff at the Albert R. Mann Library, the unit library that supports agriculture, biology, nutrition, human ecology, and related fields. This gateway began as a text-based product but evolved into a Web-based product. Its initial goal was to provide a search engine that allowed users to key in either the name of a resource or keyword terms in order to retrieve descriptions of matching networked resources. Connections to the resources were embedded in the descriptions.

Finally, unit library Web pages gave several unit libraries strong Web presences of their own, among them the hotel management school, the industrial and labor relations school, and the graduate school of business. Here again, to find and use an online resource, the user needed to know something about the cubbyhole in which the resource was located.

Importantly, none of these approaches encompassed all of CUL’s networked resources. Users at Mann Library were presented with the Mann Gateway on the public access workstations. Users at most of the other libraries were presented with Bear Access menus. Although the library staff realized that a unified approach—a single point of entry to CUL’s networked resources, or a “common entryway”—was needed, there was no consensus on the best way to achieve it.

There was at least one point of agreement however. With the rapid and continuing growth of networked resources, everyone felt that users should have the option to search as well as browse for networked resources. Library managers and
staff therefore desired a catalog that incorporated records for networked resources with embedded Uniform Resource Locators (URLs), but they supported two different options. One was to adapt the Mann Gateway technology into a CUL-wide service. The second was to develop a robust and flexible Web front end to the CUL online catalog while at the same time creating the capability of searching that slice of the library catalog that represented networked resources.

In June 1997, the library administration appointed a small committee—the Common Entryway Committee—to review the two common entryway models; evaluate the functions, features, and implementation details for each option; and to recommend a model for CUL. By the end of July 1997, with the assistance of a number of library staff from across CUL, the committee submitted its recommendation to the library administration for the adoption of the Mann Gateway technology and the creation of a CUL Library Gateway.

**THE SOLUTION**

In late summer 1997, the library administration accepted the recommendations of the Common Entryway Committee and formed three committees to build the Library Gateway over the course of the fall semester. In addition, two programmers—a specialist in HyperText Markup Language (HTML), and a specialist in database work, CGI scripts, Java scripts, connection scripts, and staff interfaces—were assigned full time to the project.

A Public Services Design Committee was charged with designing the CUL common entryway and associated pages. A Technical Services Implementation Committee was formed to provide an approach for browsing networked resources by subject, to develop procedures for gateway record creation and maintenance, and to prepare the Library Gateway database. To coordinate the work of the two implementation committees and provide project oversight, the Library Gateway Steering Committee was appointed.

The steering committee was a cross-functional group that included the chairs of the two implementation committees, the heads of two information technology departments in the library (the Library Technology Department, which served many of the unit libraries, and Mann Library’s Information Technology Section), the associate university librarian for information technology, and the head of the reference department in the largest CUL unit library. While the contributions of all committee members were important to the process, the core project team consisted of the chair of the steering committee, the chairs of the two implementation committees, and the two programmers assigned to the project.

The project was intended to accomplish five objectives:

- To create a unified, identifiable online presence for CUL
- To create a common entryway to CUL’s networked resources and services
- To develop and elicit the use of a common set of processes and procedures for cataloging networked resources
- To develop and elicit the use of a common set of processes and procedures for user training and support
- To improve coordination between CUL’s two information technology departments

The first meeting of the steering committee was in early September 1997. The project team adopted an extremely ambitious implementation time frame, and committed themselves to bringing up the Library Gateway in 17 weeks (that is, by early 1998). All of the subgroups worked in parallel, with the steering committee providing oversight. The milestones are given in table 1.

The team introduced the Library Gateway on schedule on January 5, 1998, several weeks before the start of the second semester of the academic year.

**PROJECT COSTS**

Library Gateway development from the start of the fall semester to the launch required about 2,900 hours of staff time, which are listed in table 2.

The gateway had a strong impact on li-
TABLE 1
Gateway Project Milestones

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Milestones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Technology</td>
<td>Determined hardware, software, and manpower requirements; created a Web interface to the MySQL database, which supports searching and display of CUL networked resources; installed the hardware; produced software documentation; loaded the Gateway software onto the server; trained technical support staff; and worked with the two implementation committees to create the necessary Web pages, CGI scripts, PERL scripts, a staff interface to support database creation and maintenance, and other developments.</td>
</tr>
<tr>
<td>Technical Services Implementation</td>
<td>Developed a scheme for browsing Gateway resources by subject; prepared guidelines for cataloging Gateway resources and trained CUL catalogers to use them; worked with the information technology staff to test the staff interface to the Gateway; and created catalog records for the 1,100 Gateway resources.</td>
</tr>
<tr>
<td>Public Services Design Committee</td>
<td>Decided on the specific services to be supported on the Gateway and created the design for the screens; implemented the Web pages and designed the pages that are generated by the PERL scripts; and worked closely with the information technology staff to define the final look and feel of the Gateway; wrote the content for help and informational pages.</td>
</tr>
<tr>
<td>Steering Committee</td>
<td>Prepared recommendations with regard to Cornell and CUL systems, hardware and software that would be affected by the Gateway (e.g., Bear Access); gave numerous presentations on the Gateway to solicit feedback from CUL functional groups; oversaw the incorporation of staff feedback into the Gateway design; drew up plans for public services staff orientation to the Gateway; and completed a plan for ongoing Gateway maintenance and enhancement.</td>
</tr>
</tbody>
</table>

Library staff, who had to support the new system, which was not a trivial addition to their usual tasks. Training library staff meant involving most of the organization in maintaining the gateway on some level. Prior to the launch, a task force for gateway training and user support was formed, and the chair became a member of the steering committee. This group put in place a structure and process for handling questions that arrived at reference desks and via e-mail using the "Reference Question?", "Problem Connecting?", and "Comments about the Gateway?" links at the bottom of most gateway pages. The task force also designed and carried out numerous library staff and user training sessions. This group's work had the dual benefits of preparing staff and users for the gateway and of widening the circle of stakeholders in the gateway's future.

**The Process: What Was Learned**

To maintain the library's attractiveness to readers and its centrality to the university, staff at CUL have committed themselves to offering efficient, quality access to online resources of interest to their users. To that end, a surprising new organizational competency is required: CUL must be the sort of organization that can design and bring up new online systems and services swiftly and effectively. Further, CUL is faced with mastering the process of innovation and new product development in general—not unlike organizations in the commercial sector.

Researchers in the area of new product management (see, for example, Crawford 1997; Cooper 1993; Cooper and Kleinschmidt 1993) suggest that the ability to innovate and deliver a steady
stream of new products is essential to an organization’s growth and long-term viability. Along the same lines, reduced cycle time—speed to market—appears to be central to an organization’s continued success in its market. The experience with the Library Gateway project indicates that there may be some key ideas to success for introducing new products in large, organizationally complex organizations such as CUL.

The first key idea is to start with a well-developed, clearly defined concept of the new product (which could be a new system or service, like the gateway). Make sure that management and those who must implement the new product understand and buy into the concept. The Library Gateway group was fortunate to begin its work with a well-developed prototype based on the Mann Library Gateway. And, thanks to decisions by the library administration in response to the work of the Common Entryway Committee during the summer of 1997, the implementation groups knew they had the unqualified support of CUL’s leadership.

The gateway prototype and its later iterations were valuable communication tools and guides for various individuals and functional groups in the CUL system that needed to buy into the product concept. As a result, everyone involved was reasonably clear about what the Library Gateway would be from the beginning of the project. Without such sharp, early definition of the gateway, the implementation groups might have faced vague or moving targets, or failed to gain the necessary acceptance, and the project might have taken too long—or worse, have not been completed at all.

The second key idea is to undertake a new product development project in a set of simultaneous, overlapping activities, rather than developing in phased, sequential stages. After the selection of a model for the Library Gateway, all of the implementation groups were formed at once and charged with various aspects of the work. Interface design occurred concurrently with gateway database development, technical development, staff training, and marketing (in the sense of gaining awareness and support from those who would be affected by the gateway). The process might have looked chaotic from the outside but it was not. It required a great deal of cross-functional communication, tolerance for ambiguity, management support, and project team commitment to the project, but the concurrent operations enabled the implementation groups to deliver the gateway to library users in only 17 weeks.

The third key idea is to break down organizational barriers by using cross-functional teams that are interdependent and accountable for the project. The Library Gateway Steering Committee was a cross-functional group that included administrators and staff from various functional areas of the library system (information technology, public services, and technical services). The chairs of the two implementation subcommittees (which were not cross-functional in nature) were part of the steering committee. The steering committee was effective because members were committed to the project; diverse ideas and open communication were encouraged, several team members shared initiative, and team resources were identified and used to good

<table>
<thead>
<tr>
<th>Group</th>
<th>Staff Hours (Estimate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Technology Staff</td>
<td>1,350</td>
</tr>
<tr>
<td>Technical Services Implementation Committee</td>
<td>650</td>
</tr>
<tr>
<td>Public Services Design Committee</td>
<td>625</td>
</tr>
<tr>
<td>Steering Committee</td>
<td>275</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2,900</strong></td>
</tr>
</tbody>
</table>
purpose. Without this cross-functional structure, the project could have experienced lengthy delays and foul-ups due to the inherent difficulties of passing off responsibility in sequential phases from one compartmentalized functional group to the next.

The final key idea is to seek continual feedback from stakeholders, be responsive to their concerns, and build organizational consensus as the project progresses. The gateway group did its work in the open. While speed was one of the requirements of the project, the team did not allow the gateway project to gather uncontrolled momentum. In other words, as the gateway was developed, team members proactively identified stakeholders—people who needed to know about the project, people who would have to support the gateway once implemented, and people whose approval was needed—and organized demonstration sessions to gather their feedback and advice, uncover and deal with resistance, and gain stakeholders’ acceptance of the system. Several times during the course of the project, developers were able to prototype changes to the system within a day or two of receiving a suggestion in one of the demonstration sessions. This responsiveness did a great deal to enhance the project team’s credibility and build confidence in the product.

There is evidence that cross-functional teams can produce dramatic benefits, such as rapid project completion, more innovative services, better quality, and improvements in productivity. Unfortunately, simply creating a team and promoting teamwork does not appear to be sufficient to reap these benefits. For true teamwork to materialize, the team should be the right size, its members must be committed to (and accountable for) the project, and the team must have sufficient autonomy and adequate resources to accomplish its work. Further, there is evidence that effective teams rely on shared decision making, consensus building, and frequent communications with stakeholders to manage organizational boundaries and overcome obstacles (see, for example, Ranney and Deck 1995).

Except as stakeholders, CUL collection development and acquisitions staff did not participate in the design and implementation of the gateway. This turned out to be a weakness in the project design. Toward the end of the project, the steering committee conducted a post-project evaluation, and recommended the creation of a cross-functional Electronic Resources Committee with responsibility for policy and procedural decisions associated with selecting and incorporating new networked resources into the gateway.

By summer 1998, the initial steering committee, the two implementation committees, the task force on training and user support, and the rest of the project team had disbanded. The steering committee’s final action was to produce a “mainstreaming document” in which, among other things, it was recommended that a successor group be appointed—the CUL Gateway Committee—that complements the work of the Electronic Resources Committee. The Gateway Committee is made up of a coordinator, the gateway Web editor, the editor of the guidelines for cataloging networked resources, the associate university librarian for library information technology, a technical specialist, a public services librarian, and the chair of the Electronic Resources Committee. The CUL Gateway Committee’s charge is to ensure that the gateway continues to evolve to best serve the needs of CUL and its users.

**Benefits of the Gateway**

**Benefits for Users**

With the creation of the gateway, Cornell University built a virtual branch library on the Internet. By pulling all network accessible resources, services, and library information together into a single interface, CUL created an all-inclusive Web presence. Users have a single system to go to instead of having to know ahead of time what they need and where they can find it. Using the library has become easier, and networked resources have become more visible. At the same time, the gateway
supports and points to the Web sites of unit libraries, thus highlighting and taking advantage of the talent, initiative, and diversity of these libraries.

The gateway was released in early January 1998, in the middle of the academic year. To make the transition to the new system easier the project team decided to maintain the old-style access for a semester, in tandem with the gateway. Despite this dual environment, the gateway received heavy use all over campus. In January 1998, there were 12,106 connections; one month later, total connections had jumped to 26,610. In September 1998, total connections were 48,105.

Another major benefit is the fact that the gateway has made remote use of CUL resources possible. Because of the authentication and authorization module of the system, it is not necessary to limit access to most of the CUL licensed databases to Cornell-based computers only. Now there is a way to screen users coming from non-Cornell machines and determine who is authorized to have access. This is a major benefit for travelling Cornellians.

**Organizational Benefits for CUL**

Working on the implementation of the gateway was a shared project across the unit libraries and across functional groups. Working together on a common project of this complexity and on such a short timeline was a very important step towards creating greater understanding and cooperation between functional groups and among the unit libraries. The process affected not only the key players; along the way the gateway project team had to communicate with the wider library community regularly and build its feedback and ideas into the system. The openness and responsiveness of the process helped build trust and acceptance for the gateway in library staff, some of whom were quite reluctant and even resentful of the change at the beginning of the project.

Even after the implementation, the gateway project continues to foster dialog and cooperation. The maintenance of the system requires constant input from and dialog between different functional units of the library (collection development, acquisitions, cataloging, public services, and information technology) and between the main and unit libraries. Openness to suggested improvements and criticism continues to be important.

The library's investment in its online resources is half a million dollars a year. Making these resources more readily accessible and more visible maximizes the benefits of this substantial investment.

**The Future**

The library administration has funded a series of focus groups to provide CUL managers and staff with more feedback on the system and more general information about what features the users value and need. The results of this study should prove useful for the development of future generations of the gateway as well.

**Summary**

While the Library Gateway is an important technological achievement, what might be most helpful to other institutions is an understanding of the process used to build it. First, the experience with using teams was an excellent one. The cross-functional steering group and its two subgroups were able to cut across traditional organizational boundaries, to work and respond quickly, and to build consensus continuously. Second, but equally important, was the library management's support. By providing a clear mandate and strong sponsorship while still giving the team sufficient autonomy, CUL's leaders established the conditions in which innovation, teamwork, and excellent service could emerge. Third, the critical importance of involving all functional groups in the task of organizing electronic resources in libraries was demonstrated. In the face of the dramatic shift brought about by the Web, it is essential to integrate library functions effectively, and to make the most of what each group knows and does best.
WORKS CITED


INDEX TO ADVERTISERS

Library Technologies 2d cover
Archival Products 105
ALCTS 3rd cover
OCLC 4th cover

Why is the DDC (Dewey Decimal Classification) structured the way it is? What is its relationship to theories of classification of knowledge and of library classification? How has it evolved over the years, and what are its prospects as we enter the age of the "post-modern library"? These are some of the questions Francis Miksa discusses in this interesting study, an expansion of his DDC Anniversary Lecture at the Fourth International ISKO (International Society for Knowledge Organization) Conference in July 1996.

The DDC is one of the most successful library classification systems in the world, as evidenced by its widespread use in libraries in the United States and in many other countries and by its growth and evolution over more than 120 years. But the history of the DDC has not followed a straight line. Its development has been affected by the personal beliefs of founder Melvil Dewey and of the various editors who came after him, by changing theories of library classification, and by practical considerations. Numerous controversies have arisen along the way; for example, the radical changes in the fifteenth edition in 1951 are blamed in part for the move by many academic libraries from the DDC to the Library of Congress Classification.

The DDC was a product of its age. Dewey developed it in the 1870s, when the larger scientific and philosophical community was greatly interested in the classification of all knowledge. However, library classification always has had a practical aspect—the DDC is not just a classification of knowledge; it is used to classify actual books in libraries of various sizes and types. There always have been conflicts over how many levels of hierarchy to include in the Dewey classification scheme and over how much it should be revised to keep up with changing knowledge, given the need for integrity of class numbers to reduce the work of reclassifying books in libraries.

In 1899, the Library of Congress wanted to adopt the DDC, but with fundamental changes that Dewey refused. LC then went on to develop its own very different classification scheme. Also at the turn of the century, the new Institut International de Bibliographie in Brussels set out to catalog and classify scientific literature. Paul Otlet and Henri LaFontaine began by expanding on the DDC, but the specialized nature of the materials with which they were dealing and the particular needs of the scientists and scholars they served led to the development of the Universal Decimal Classification (UDC). Their work led to a new idea of subjects, one that was much more complex and more difficult to contain in a strict hierarchy. Miksa also details the contributions of Ernest Cushing Richardson, Henry Evelyn Bliss, W.C. Berwick Sayers, and S.R. Ranganathan to library classification theory in the early parts of the twentieth century, when the old view of the universe of knowledge as a one-dimensional, hierarchical structure was replaced by a more complex, modular, and faceted view.

Since the 1950s, these theoretical developments have had an impact on the DDC in its major efforts to keep the ter-
minology up-to-date and to improve subject collocation, as well as in increasing use of subject faceting and notational synthesis. However, Miksa finds that the DDC is still based on three assumptions that, if unexamined, could have a negative effect on its future and on the future of library classification in general. These assumptions are: “knowledge categories are by nature hierarchical and logical; one best classification system is achievable; and document retrieval is the main purpose of library classification” (p. 82).

Only in his conclusion does Miksa consider the “post-modern library,” by which he means both an evolving concept of the library as a personal space, made possible by electronic information and modern telecommunications, and the library as a product of the post-modern age, in which knowledge and truth are no longer absolute and everything is relative. The best hope for the survival and continued relevance of the DDC in this age lies in a thorough re-conception of the system, including variable levels of specification and alternative arrangements for locating information.

This slim book will interest readers who are curious about the history of libraries and the classification of knowledge. Its relevance is not limited to libraries using the DDC or to classifiers. It is very well organized, though it may provide a little too much detail on the history of the various editions of the DDC and not enough insight into possible future developments. A nine-page reference list is included.

—John Hostage (hostage@law.harvard.edu), Authorities Librarian, Harvard Law School Library, Cambridge, Mass.


In Communicating Research, A. J. Meadows begins by stating that “Communication lies at the heart of research ... for research cannot properly claim that name until it has been scrutinized and accepted by colleagues” (p. ix). This is obvious, but perhaps too easily taken for granted. Throughout this book, Meadows reiterates that understanding results from communication—from the interaction between the researcher and the receiver of research information. Research communication continues to evolve, most recently with the introduction of electronic networks—Meadows summarizes his main theme as “change and diversity” (p. 239). In his view, changes in the research communication process are not driven solely by technology, but also by the needs of the research community. He recognizes the effects of the technology, but considers it in its proper place as a tool of the researcher.

This book provides a good overview of the history and evolution of research communication, with a concise timeline that proceeds from Aristotle’s symposia to the invention of printing, the development of postal systems, and the rise of the modern research journal as learned societies found meetings, personal correspondence, and books inadequate to keep a growing audience abreast of expanding research.

The emphasis in Communicating Research is on academic research, though private industry and government-funded research are included as a basis for comparison. Meadows focuses mainly on research in the natural sciences and includes humanities and social sciences research when he touches on the evolution of these distinct fields and the difficulty in defining the boundaries between them. A general characterization of the differences between the humanities and social sciences versus the natural sciences is that in the former, the book is more widely used as a tool of communication, while the journal article is most common in the latter. In addition, the emphasis in the natural sciences is almost always on the most current research, while older knowledge is read and cited more often in the humanities and social sciences. The structure of the journal article contributes to the efficiency of the communication process by providing the expected title, author, date of receipt (essential for establishing “first” discoveries, especially in the natural sciences), abstract, body (usually introduction, methodology, results, conclusion), and list of references. Books also have a typical layout, with an
index to assist researchers in locating and absorbing information quickly.

The "professionalization" (p. 24) of research began in German universities, was well-established by the second half of the nineteenth century, and then spread to the rest of academia and eventually to industry and government. The size and education level of the research community began to grow, along with the amount of information it generated. Meadows makes the interesting point that the feeling of being overloaded with information actually has been common for a long time. To cope with the expansion of knowledge, universities were organized into departments and new disciplines were created through specialization or fusion.

The research community is subject to a set of social norms that are reflected in the communication process. When these norms are followed, they are invisible, but when they are not, plagiarism and the forging of research data may result. The community typically experiences periods of "normal science" interspersed with periods of "revolution," all documented through the recognized communication channels that record the cumulative knowledge of a discipline. Meadows discusses various aspects of research, including who performs research and why, the lower visibility of women researchers and those from developing countries, the effect of age on the researcher's output, the involvement of multiple authors in writing results, citation studies, and the roles of editors, referees, the mass media, and even amateur researchers. He highlights current trends, such as the increasing emphasis upon general, theoretical research as opposed to specific, empirical research (though this varies by discipline) and the increase in collaborative and interdisciplinary efforts.

The effects of technology on research communication are stressed throughout this book. Meadows points out that technological manipulation of bibliographic data helps researchers locate information. At first, only bibliographic citations were searchable by computer, then abstracts, and now full-text. He credits information technology with opening up new research opportunities and regular communication outside the research hierarchy, making the research process more democratic and open. Electronic networking allows wide dissemination of research in progress. On the other hand, these newer, more flexible channels bring a number of challenges: lines are blurred between formal and informal exchanges (when is something "pre-" and when is it a "print"?), questionable material can go undetected by the novice reader, data tampering can occur, and "electro-copying" can jeopardize copyright protection. There are citation difficulties, uncertainties regarding long-term archiving, and a continuing reluctance to accept electronic publishing within the academic community. Meadows endorses an ideal scenario in which print and electronic publication are parallel.

Though it is not the major focus of this work, Meadows recognizes the role of publishers and libraries in the research process. Publishers work with authors to assure that information is suitable for publication, that it is produced in an appropriate, legible format, and that it is well-disseminated. Libraries are the primary purchasers of scholarly publications; their selection decisions affect the actions of both publishers and readers, and these decisions are in turn often driven by the immediate needs of the institution's own researchers. The library's role is to archive the records of research and to make them available, a role that brings with it dilemmas regarding the financial inability to keep pace with publishing and the rising cost of physical storage. Both libraries and publishers face new challenges with electronic publishing, though Meadows believes that libraries have the most difficult task—storing and providing access to both the old and the new.

Electronic journals gained popularity first in the humanities and social sciences—articles are composed mainly of text, and publishing space was needed that was unavailable in print. The natural sciences are catching up now that technology can support the graphics required for publications in this field. Researchers at present recognize that a mix of print
and electronic documents is the reality of the mid-1990s, but Meadows advocates more retrospective conversion to electronic format in order to expedite researchers' work. Paradoxically, information technology has made the researcher's task more difficult by increasing the amount of information, while at the same time making it more convenient for each researcher to access that information from a personal workstation. Meadows cautions that these electronic channels may further divide the information-rich from the information-poor, based upon the degree of access to information technology.

Many of the author's points are illustrated by tables with data on everything from "A Comparison of the Number of Articles Devoted to Astronomy/Space and Medicine at Two Epochs" (p. 73) to "The Existence of Bias in Refereeing Judgments" (p. 190). The concepts that Meadows presents are so interconnected and so often repeated that I found myself pondering a better organization of this work while I read it. Yet in the end, I realized that this is the point: the communication of research is a complex, chaotic process that is constantly changing. The choice of the term "Postscript" for what I first considered a conclusion or summary seemed odd, but it became clear from the closing sentence that the author's clever play on the word "post-script" (p. 242) was intentional. Meadows correctly observes the myriad questions raised by the electronic channels of communication now available to researchers, although he unfortunately does not seem to have any greater insight than the rest of us into how it will all turn out.—Ellen McGrath (emcgrath@acsu.buffalo.edu), Head of Cataloging, Charles B. Sears Law Library, State University of New York at Buffalo


In *Budgeting for Information Access*, Martin and Wolf define access very broadly as "finding, looking at, or using any printed or electronic information" (p. 20). All types and formats of accessible collection-related materials, from print to electronic media, are discussed and examined in a series of short chapters. These include materials that are directly retrievable on-site in traditional and electronic formats and those received through interlibrary loan or document delivery services. Both editors are associated with *Technicalities*, a library journal and forum that represents the practitioner's point of view. Martin, who died in April 1998, was the author of the "Money Matters" column in *Technicalities* and many articles and books on budgets and library collections. Milton T. Wolf is the founding editor of *Technicalities* and is currently vice president for collection programs at the Center for Research Libraries.

Most of the chapters in *Budgeting for Information Access* contain practical information about collection-related concepts such as preservation, pricing, collection organization and maintenance, and budget implications. The "Notes" section in each chapter and the selective bibliography at the end of the book contain current and relevant source material, although the mention of some of these publications and studies is repeated too often throughout the text, as are material and charts from several of Martin's previous works.

The strength of this book lies in its discussion of budgeting and of the wide range of information alternatives available to libraries in the current marketplace. These alternatives include other collection-related areas—beyond print and electronic access—such as document delivery, interlibrary loan, resource sharing, and consortial and contractual relationships. This is important reading for acquisitions librarians whose functions overlap increasingly with areas outside their traditional domain, i.e., borrowing, leasing, and procurement for individuals, rather than purchasing materials to be cataloged and processed for the library collection. The thoughtful sections on developing a resource budget and on budget
scenarios cover an important and often overlooked area of library management.

**Budgeting for Information Access** is less successful in its discussion of electronic resources. Problems associated with ownership and preservation in the electronic arena, as well as the economic consequences of acquiring and licensing these materials, are all important issues that many librarians confront. This book's heavy-handed bias toward print materials can be justified in some circumstances; nevertheless, statements that compare and contrast print and electronic resources often are confusing and overly dramatic as indicated in the following statements: "Of particular importance is the difference between accessing a book and accessing electronic media . . . because the first is a simple matter of picking up something that has already been purchased and is therefore subject to the Doctrine of First Sale, while the second is like accessing a datafile that is not in the public domain" (p. 119) and "Unless providers and users can find more common ground, the unlimited promise of the new medium may diminish or vanish" (p. 88). In addition to commercially available products, which are emphasized in the text, there are thousands of government-maintained, content-rich Internet sites, accessible from the home or library, which provide information directly to the public. Electronic resources vary greatly in format, access, usage, content, and quality; in this text, their distinctions are sometimes neglected in favor of broad editorial statements on electronic resources in general. Peggy Johnson (1998) and Ross Atkinson (1998) deal with these issues more effectively in an effort to inform and improve library practice and theory. Discussions of copyright and contract licensing are scattered throughout the text; individual sections devoted to these issues, with specific guidelines, would be helpful.

Over the last decade, in an effort to meet users' information and research demands, libraries have attempted, within budgetary constraints, to provide access to a wide range of services and materials. Martin and Wolf's book is a good starting point for a better understanding of "access" in an ever-changing library environment.—Amy Dykeman (amy_dykeman@solinet.edu), Library Products and Service Manager, Electronic Information Services, SOLINET, Atlanta, Ga.

**Works Cited**


Michael Lesk is a computer scientist who has shown his interest and expertise in library-related issues in research for the Commission on Preservation and Access and who has dedicated his work to studying problems and solutions for the electronic library. He has written several respected works on preservation and technology; but this book, a natural extension of his interests and experience, is by far the most extensive and will reach the broadest audience.

**Practical Digital Libraries** is one of the first books to treat this subject so comprehensively and in such a readable style that students, educators, librarians, and computer scientists all will find it interesting and valuable. The subtitle "Books, Bytes, and Bucks" indicates the book's attention to the economics of digital library projects. This topic is critical in a discussion of changes in technology and how to mobilize, plan, and transfer library functions and resources to a digital environment.

Lesk provides a thorough review of what a digital library is and how to build one; the book can easily serve as a reader or textbook, and it would also be useful in a study of trends in scholarly communication. In the opening chapters, Lesk focuses on the technology of conversion and construction and the needed equipment and software. He examines techniques for storing and manipulating images and for
storage of sound and multimedia, and possibilities for classification and indexing. A discussion of options for delivery of the digitized content and for information retrieval is central in this book. Lesk concludes with discussions of collection, maintenance, and preservation of digital information and a review of economic models to sustain the development of digital libraries. The practical and well-organized nature of this volume, and its numerous illustrations and graphical presentations of complicated information, make it a useful resource for anyone involved in the developing digital library environment. Valuable reader-friendly features are the extensive list of figures, a bibliography of references, and a useful index.

Lesk tackles the "whys, hows, and what decisions remain to be made" (p. ixv) as new technology makes digital libraries more a reality at all economic levels. The author underscores how library users have come to expect to be able to access, search for, copy, and store information with maximum ease and dependability, using common equipment that doesn't require complex skills. The outcome of a digital library is a new sense of organizational culture—it is this social aspect of computing and the transformed library environment that is so exciting to study and experience. This book is a success as a planning tool, an analysis of how to problem-solve in planning various aspects of a digital library. In building a case to answer the question "Why digital libraries?", Lesk presents dilemmas that can be studied and solved in a variety of ways. Examples of these problem situations include text formats; methods of transferring images and text so that they can be read and copied; new options for multimedia storage and access; methods of knowledge representation; distribution of content and security concerns; and usability and retrieval evaluation. The concluding summary in each chapter includes both a summary of the chapter's content and ideas on future developments in this dynamic environment.

The second part of the book is less technical and more application-oriented, covering collections and preservation, economics, copyright and intellectual property, the scope of international activities, and future impact on society and public policy. Lesk's almost encyclopedic approach gives direction to the technical planning process. He is candid and honest, identifying genuine problems and solutions that obviously have limitations and will need to be revisited when new technologies become available.

As a librarian, I found the chapters "Economics," "Knowledge Representation Methods," and "Collections and Preservation" most useful. The understanding of technicalities gleaned from the earlier chapters can be applied to these discussions of planning for digital access. The information is useful to planners from different professional avenues, for small and large projects, and in various higher education environments. There is a good introduction to government-funded research initiatives that nonexperts will find helpful.

This book will have a reasonable shelf-life in libraries because it fills a void. Until it became available, librarians had to study computer science and engineering manuals to learn about the technology of digitization. The 1997 imprint should not be a concern because there is little else published of comparable scope in so intelligent and readable a volume. Many earlier and later works cover aspects of this subject in greater detail, but as a single-volume work on this subject, Practical Digital Libraries stands alone. A newer and somewhat related work on the same subject, without the technical information, is The Mirage of Continuity: Reconfiguring Academic Information Resources for the 21st Century, edited by Brian Hawkins and Patricia Battin (Washington, D.C.: Council on Library and Information Resources, 1998). The World Wide Web is also a source of information about digital libraries. The D-lib Program, Research in Digital Libraries (http://www.dlib.org), provides a variety of useful resources, including D-Lib Magazine, but it is nowhere near as easy to use or as complete an overview as Practical Digital Libraries: Books, Bytes, and Bucks.—Julia Gelfand (jgelfand@sun1.lib.uci.edu), Applied Sciences Librarian, University of California, Irvine.
ALCTS 1999 Preconferences

ALCTS #1
BUSINESS OF ACQUISITIONS
Advancing Acquisitions:
Services, Standards, Skills

Thursday, June 24, 8 a.m.-5 p.m.
Friday, June 25, 8 a.m.-noon

Acquisitions departments and acquisitions librarians face a variety of new demands, such as managing electronic products, outsourcing, and responding to shrinking resources. This advanced institute examines these demands and challenges for continuing growth in combination with resources and abilities available

ALCTS Members $175
ALA Members $225
(includes ALCTS membership)
Nonmembers $275

ALCTS #2
Bringing Monographs into the Digital Library

Friday, June 25, 8 a.m.-5 p.m.

This preconference will explore issues related to the incorporation of eMonographs/eBooks into the digital library, presenters will define the eBook; identify its significance in the digital library; address eBook selection, acquisition, organization, distribution, copyright, preservation, and perpetual access from the librarian's perspective; and address key issues of importance to librarians and publishers in

ALCTS Members $115
ALA Members $165
(includes ALCTS membership)
Nonmembers $215

For more information please go to the ALCTS website at http://www.ala.org/alcts/events/preconference99.html or contact Yvonne A. McLean, 1-800-545-2433, ext. 5032 or ymclean@ala.org

3 WAYS TO REGISTER

On Line- Complete the on-line registration form located at http://www.ala.org/events/ac99/online_reg.html

By Fax- If you pay with a credit card you may fax your completed registration form 24 hours a day by dialing 312-944-7841

By Mail- Send your completed registration form with payment to:
American Library Association, Dept #77-6565, Chicago, IL 60678-6565

You may find the registration form on the ALA website or in the January or April issues of American Libraries
Developing Asian language collections is easy with the OCLC AsiaLink service. Staffed by Asian language experts, AsiaLink offers everything you need to establish, maintain and expand Chinese, Japanese, Korean and Vietnamese (CJKV) collections in your library.

- The AsiaLink Collection Development service lets you determine the specific subject matter and format of materials that AsiaLink staff will acquire on your behalf, based on your users' needs. You'll also get significant savings on the cost of books, including higher-than-average discounts from book vendors.

- AsiaLink Sets get you off to a quick start or allow you to easily expand existing collections. Books are available in sets of 25, 50 or 100 pre-selected fiction and nonfiction titles. Both juvenile and adult sets are available. Title lists are updated annually, so it's easy to add to your collection regularly.

- Full-level OCLC cataloging is provided for all titles and materials may be physically processed upon request.

For help cataloging titles not acquired through AsiaLink, turn to the OCLC TechPro service. TechPro provides expert cataloging for materials in all bibliographic formats and in many languages (including CJKV), plus optional physical processing. TechPro can help you quickly and efficiently eliminate backlogs and keep your cataloging up-to-date, giving your users prompt access to materials.

For more information on AsiaLink or TechPro, contact Marcia Stout at OCLC.

1-800-848-5878, ext. 4386  614-764-4386  Fax: 614-718-7277
E-mail: marcia_stout@oclc.org.

www.oclc.org/oclc/menu/col.htm