

**PRINCIPIA BIBLIOGRAPHICA? BALANCING PRINCIPLES, PRACTICE, AND
PRAGMATICS IN A CHANGING DIGITAL ENVIRONMENT**

Dick R. Miller

SUMMARY: This article explores the emphasis on control in cataloging versus the chaos found on the Web. It delves into adversarial relationships between catalogers and technologists. It seeks commonalities and suggests that new alternatives addressing both perspectives might offer superior and more satisfying results. A series of examples juxtapose current practices, enrichment possibilities, and flaws in current digital solutions to suggest potential opportunities where catalogers might excel. Speculation on ways to promote cataloging principles and values via more direct participation in the unruly digital environment hints at a more promising future for our profession.

KEYWORDS: Cataloging, Relationships, Principles, Transparency, Balance, Cooperation, Control, Time, Open Access, Tradition, Innovation, MeSH, XML

Dick R. Miller, MLS, is Associate Director for Resource Management, Lane Medical Library and Knowledge Management Center, Information Resources and Technology, Stanford University Medical Center, Stanford, CA 94305 (E-mail: dick@stanford.edu).

The author is indebted to Ruth Carter for always being open to publishing new ideas, even controversial ones, and to his many supportive colleagues at home and abroad. Lane

Medical Library, celebrating its centennial in 2006, continues to provide an environment for exploring new ideas and pursuing excellence.

“From the principle of vanity that holds a prominent place in every man’s heart, there is too great a tendency to dwarf the past of our science, to exaggerate its present, and to greatly magnify its future.”

--Levi Cooper Lane, surgeon and namesake of Lane Medical Library.¹

Periods of rapid change make it challenging to maintain information management models that may have served well for many years. The nature of change itself invites controversy, and even confrontation, as interested parties seek to adjust past practices or to establish new models better suited to the changing conditions. Rather than lamenting loss or dismissively forging ahead, it is useful to explore the balance between the two perspectives, old and new. Inevitably, there will be lessons to be learned from traditions, which underwent their own controversial developments, and from the bold promise of emerging paradigms that may falter from lack of universal adoption and the validation of long practice.

The beginning of the twenty-first century finds libraries, and cataloging in particular, embroiled in such conundrums. Forging the future through the lens of the past may well augur greater success and stability for libraries than emulating the wasteful competition found in the business world. Open and cooperative endeavors, designed to enable and foster catalogers’ traditional talents, could flourish in Web environments--too often viewed by this constituency as threatening.

BALANCING THE BOOKS?

At the 2005 LITA National Forum, Michael Gorman observed that today’s plethora of metadata schemes parallel the early history of cataloging when libraries had their own cataloging

rules, classification schemes, etc. and operated largely independently of one another. However, instead of the apparent lone option to continue establishmentarian practice as is, catalogers need to seek ways to demonstrate the value of cataloging principles in an open Web environment. Are the nouveau information organizers reinventing the wheel under our noses? What could they learn from cataloging, and what could catalogers learn that might foster greater respect all around and provide improved library services?

Catalogers know the value of clearly identifying works to distinguish between them and recognize the growing importance of this in anchoring relationships. They appreciate the value of authority control and controlled vocabularies, the value of details that may not interest one user but are crucial to another or to enable automated retrieval, the value of reliable and persistent information, the complexity of representing the actual facts, the importance of context in interpreting bibliographic data, and the fundamental value of doing a social good. Wouldn't we and our users be better off if the world were more aware of these qualities and the difference they can make in information retrieval? As metadata troves grow, the value of cataloging skills will become more apparent. We should be stretching our thinking to be ready for the inevitable opportunities that will arise and that would recognize and take advantage of our unique skills.

To explore areas that may offer possibilities for "building a better catalog" and beyond, the following broad strokes and selected details present some problems and possibilities to ponder in thinking about the role of catalogers in the development of more robust and flexible information systems. The reality is a very competitive Web environment with commercial and specialist interests often ignoring the library, other than perhaps LC, OCLC, and RLIN. In this ramble through the information landscape, consider the potential for libraries' less restrained participation in an open environment--without compromising our values.

A SEMI-OPEN UNION CATALOG?

Is the big picture in focus? Could a web-based Metalog, an ersatz combination of OpenWorldCat and Wikipedia elements, provide a more robust shared authority file and union catalog? It is not so radical to consider combining libraries' dispersed and diluted efforts to provide a semi-open, yet structured, framework for organizing information. Rather than being a source for copy cataloging, local systems could link to it for transient bibliographic needs and use Google subsets of it to identify local holdings. "Accredited" catalogers could create and edit records freely by fixing errors or adding information on the spot. (OpenWorldCat isn't really open; it only permits appending notes.) By using an "OpenID" structure, terminology could change; record enhancement, without breaking external links, could flourish. Although reminiscent of the CataWiki prototyped at Access 2005 in Edmonton to permit group editing of MARC records and to "break the cataloguing cartel," this is not the intention. There is need for responsibility, detail, thoroughness, erudition, etc., and catalogers do have a certain "je ne sais quoi" in these regards.

Some of our redundant cataloging effort could be pooled as an experiment in demonstration projects to show how groups of libraries could cooperatively create and maintain a Web-based and Web-embedded catalog and authority files far more valuable than the many separate silos. Supervised volunteers and specialists could be enlisted to extend the possibilities. Quick responses to a suggestion/contribution box would encourage the unaccredited to flag errors and submit records. Library programmers could be enlisted to make batch changes. Also, by encouraging other websites to link in at the appropriate level, a looser array of resources could cluster around the more formal managed core. The open, cooperative environment should

appeal to grant agencies, which might fund projects to map records and holdings, enrich the data with FRBR relationships, encourage international cooperation, and other similar projects. Protocols for ironing out data conflicts and other policies could develop gradually without discouraging participation, i.e. accepting that it is OK to make mistakes.

Google Book Search and others are making complete books available on the Web with little or limited metadata, but libraries are largely keeping their cataloging efforts isolated and straight-jacketed. This disconnect is not to our advantage. Libraries' interests would be better served by a convergence of interests and efforts.

Criticisms of cataloging vis-à-vis new technologies have called for the “death of MARC” and used strident phrases such as “rearranging the deck chairs on the Titanic” and “putting lipstick on a pig.” While inflammatory, they often contain kernels of truth. Efforts by the concerned parties would be more productively spent together in exploring both XML's benefits and pitfalls and cataloging's values and reticence in an open and constructive manner. The issues are fundamental and not restricted to a particular purview. Focusing the ardor on effective communication and joint involvement would likely promote more innovative solutions to the shared problems.

VALUE-ADDED METALOGING?

In addition to reviewing the big picture, detail needs scrutiny as well. Should catalogers be scribes or illuminators? Cataloging in a full text environment calls for a greater emphasis on judgment and evaluation to create the best metadata. Identifying key concepts and translating them to a controlled vocabulary enhances findability, and constructing relationships provides valuable context. These professional skills will likely retain or increase our value whereas

transcription is likely to become a mostly automated extraction taken for granted. Likewise, enrichment of authorities by supplying alternative terms and by determining and normalizing relationships will help users navigate masses of information more effectively. The value of such information need not be limited to the catalog but could support interfaces to other databases and Web resources and act as a device to automatically provide synonymy and offer alternatives before sending an enriched search request to chosen targets.

Words

Fill-in-the-box keyword searching is popular. The “Did you mean” option in Google is a useful feature; a search for “psuedo” yields a respectable 880,000 hits; however, this represents only 3% as the suggestion of “pseudo” produces over 30 million, no doubt enhanced by Google’s stemming algorithm. Misspelling is an interesting phenomenon because word processing software can automatically correct this error, and yet savvy EBay shoppers actually seek sleepy listings containing misspellings in search of bargains overlooked by compliant spellers. But, maybe you meant “fake” (43 million), “false” (132 million), “faux” (16 million), or “counterfeit” (6.5 million)? The “pseudo” retrieval is only 14% of these near synonyms; and then there’s “artificial,” “bogus,” “not genuine,” “simulated,” etc.

Roget’s has been around since 1852, but thesaural relationships have not been incorporated effectively into search retrieval software. Because word searching is so popular, help in selecting semantically related words could distinguish catalogs. Do searchers always think of dirigibles or Zeppelins when seeking blimps? Browsible keyword indexes could identify variants and related terms, e.g.:

house, abode, domicile, dwelling, habitation, residence [near synonymy]

apartment, castle, condominium/condo, digs, home, mansion, etc. [related]

casa, Haus/Häuser, maison, [other languages]

House of ... [more sophisticated options for royalty, government bodies, etc.]

Entries

Deliberate catalog entries are often touted as superior to free text. However, without building richer references, the same problems in finding the right word are likely to occur as in locating the right entry. The still brewing *Resource Description and Access* (RDA) code emphasizes relationships,² but the new title signals the traditional mantra while downplaying its Anglo-American slant. Hopefully, the new standard will balance a combination of the identification/characterization of resources and the articulation of relationships among them to improve access--not only to the resources, but to the underlying and underutilized authority structures. Whether the code is truly “designed for use in a digital environment” may lie in its answer to questions such as: Will users be able to generate standard citations from bibliographic records?

Even when structures are available, will reluctance to go beyond the traditional persist? Might catalogers adopt a more value-added orientation? The following example may stretch cataloging attitudes more than it does MARC coding. Consider a literary movement, La Pléiade. In December 2005, the LC catalog had 46 records with the heading “Pléiade” and 581 records containing the keyword, many from series. In contrast, an exact phrase search produced 188,000 results in Google, 3,560 in Google Scholar, and 2,040 in Google Books. This is the related LC authority:

150 __ |a Pléiade

550 __ |w g |a French literature |y 16th century

550 __ |w g |a French poetry |y 16th century

670 __ |a Levi, A. Guide to French literature : beginnings to 1789, 1994: |b p. 610 (hdg.: Pléiade, La; list of 7 French poets active mostly in the third quarter of the 16th cent., drawn up by Ronsard)

670 __ |a New Oxford companion to lit. in French, 1995: |b p. 627 (hdg.: Pléiade, La; constellation of 7 poets active from 1549 to 1589, under leadership of Ronsard)

From the cross references, we learn that this heading is related to sixteenth century French literature and poetry. The non-public notes justifying the entry add scant additional scope.

Contrast this with the first paragraph of the Wikipedia entry in English in December 2005:³

The Pléiade is the name given to a group of 16th-century French Renaissance poets whose principal members were Pierre de Ronsard, Joachim du Bellay and Jean-Antoine de Baïf. The name was a reference to another literary group, the original Alexandrian Pleiad of seven Alexandrian poets and tragedians (3rd century B.C.), corresponding to the seven stars of the Pleiades star cluster. The name "Pléiade" was also adopted in 1323 by a group of fourteen poets (seven men and seven women) in Toulouse.

The Wikipedia entry is much richer in both informational content and links. The entry covers two French entities of the same name in some detail, identifies the star cluster namesake, and provides links to selected individual poets and to an ancient Alexandrian Pleiad. The French Wikipedia includes an entry, Pléiade (XVI^e siècle), which provides even more specifics. Oddly, there is another archived French version, Pléiade, which contains additional useful information not included in the current entry.

In considering context and relationships, it is not difficult to tease out information from these hyperlinked descriptions and, with additional effort, to unearth salient details from Google's 2.24 million entries on the stemmed word, including full text of the *Dictionary of Phrase and Fable* (1894). This exploration anticipates the need for disambiguation and variant access--perhaps as follows:

150 __ |a Pléiade (France : 16th cent.)

450 __ |a French Pleiade, First

450 __ |a Groupe Pléiade

450 __ |a Pléiade française

450 __ |a French Renaissance Pléiade

450 __ |a Pleiad (France : 16th cent.)

450 __ |a Brigade |f 1553-1556

Who “belonged” to this movement? There were initially seven poets, but actually more due to replacements. Potential personal name relationships with their duration lead to those

involved. While sometimes the names encountered did not match entries established in LC authorities, it was not difficult to locate all of them as well as a couple of other related names:

500 1_ |e Leader |a Ronsard, Pierre de, |d 1524-1585.

500 1_ |e Member |a Du Bellay, Joachim, |d 1525 (ca.)-1560.

500 1_ |e Member |a Baïf, Jean-Antoine de, |d 1532-1589.

500 1_ |e Member |a Jodelle, Étienne, |d 1532-1573.

500 1_ |e Member |a Tyard, Pontus de, |d 1521-1605.

500 1_ |e Member |9 1553-1582 |a Peletier, Jacques, |d 1517-1582.

500 1_ |e Member |9 1582- |a Dorat, Jean, |d 1508-1588.

500 1_ |e Member |9 1553-1554 |a La Péruse, Jean de, |d 1529-1554.

500 1_ |e Member |9 1554- |a Belleau, Remy, |d 1527?-1577.

500 1_ |e Associated |a Des Autels, Guillaume, |d 1529-ca. 1581.

500 0_ |e Inspiration |a Francis |b I, |c King of France, |d 1494-1547.

Related headings, perhaps coded to indicate their “semi-established” status, would provide awareness of the six other septets known by the same name, benefiting both catalogers and users:

550 __ |a Pléiade (Alexandria, Egypt : Tragic)

550 __ |a Pléiade (Alexandria, Egypt : Poetic)

550 __ |a Pléiade (Carolingian)

550 __ |a Pléiade (France : 17th cent.)

550 __ |a Pléiade (Toulouse, France : 14th cent. : Men)

550 __ |a Pléiade (Toulouse, France : 14th cent. : Women)

Additional relationships and a public scope note would also help:

550 __ |a Renaissance

551 __ |e Locale |e France

551 __ |e Namesake |a Pleiades

678 __ |a This constellation of roughly seven French poets promoted French as a language of literary expression. The name, derived from the seven visible stars of the Pleiades, references the group of ancient Alexandrian tragic poets.

Catalogers with a knowledge of French literature would be better suited to provide such enhancements and the above “straw” record may contain inaccuracies in addition to the nouveau MARC subfields. This type of effort is a bargain when considering that subsequent explorers can avoid sifting through the haystack repeatedly. The level of effort need not be applied consistently, but perhaps proportionately to the prominence or confusedness of the topic. One can argue what to include, but balanced/neutral assertions can help avoid errors of omission and allow the searcher to decide which relationships to pursue.

Other than tedium and frustration, the difficulty in providing this enrichment was in the delineation/disambiguation of the groups and in identifying/finding the authorized forms of the related names. Both are professional activities, and even drudgery can result in the delight of identifying a missing link. Standards to address form of entry create stability and address

weaknesses in open systems. The Wikipedia is not without problems; recently it had to bar unregistered users from creating new pages due to controversial content. Despite disclaimers of validity, it is interesting that citations to it occur in authority records:

970 __ |**a** Wikipedia home page, via WWW, July 11, 2003 |**b** (Friedrich II., German king, emperor of the Holy Roman Empire; b. Dec. 26, 1194, d. Dec. 13, 1250)

The Wikipedia is a work in progress. While more than one Pléiade is distinguished in its entry, this is not reflected in its page devoted to disambiguation of Pleiades; nor does it reference the separate, relevant “Alexandrian Pleiad” page.

Pleiades (disambiguation)

Pleiades can refer to:

- * Pleiades (star cluster) - an open cluster of stars in the constellation Taurus;
- * Pleiades (mythology) - the seven sisters of Greek mythology;
- * Pleiades (volcano group) - a group of volcanoes in Antarctica;
- * La Pléiade - a group of 16th-century French poets.

The Wikipedia’s limitations can be subtle. In reading one of its biographical sketches of a noted scientist, it was poignant to learn that the luckless researcher had not received any of the leading awards in his field, despite his fundamental contributions that were so obvious to the writer. Only from unrelated references in the text was it possible to discern that the page had been based on content written in the 1970’s and that the subject had been fully honored in the

1980's. Although the entry was a recent creation, it relied on an earlier document that the author had failed to verify as to currency. In this case, the reader fixed the error but forgot the scientist's name and cannot cite it. The importance of such time factors is treated further below.

While the Wikipedia's limited controls leave it susceptible to controversies, its model could be adapted so that catalogers could share efforts in building a more effective authority file focused on anticipating users' needs rather than so strictly relying on literary warrant of collections. In reality, libraries rely heavily on books to establish the need for entries. Books are typically several years in the making; and thus terminology lags rather than being ready when books arrive.

STRUCTURE AND FUNCTION IN INFORMATION SYSTEMS

Bibliographic data elements, their attributes, and their relationships to other records determine to a large extent how flexibly this information can be processed for varied tasks, whether to support discovery, circulation, bibliometrics, or Web displays. While MARC has stretched about as far as possible to accommodate all manner of data, many recognize that it is hampering the fullest utilization of bibliographic data in Web environments. XML has shown much greater flexibility in deploying cataloging information, but schema development has been hampered by blind fidelity to MARC. Exceptionally, Martha Yee has responded formally to questions raised about MARC.⁴ Unfortunately, her response preceded the detailed presentation of deficiencies in AACR and MARC included in *Putting XML to Work in the Library*.⁵ These issues are beyond the current scope of this article; but, regardless of one's stance in this protracted controversy, more open dialogue is needed.

Browse Entry (Sequential Access) vs Keyword Search (Random Access)?

Web search engines focus on keywords/phrases with relevance ranking. Their proprietary algorithms and policies largely shroud the systems' internal workings from scrutiny although much can be ferreted out by the persistent.⁶ Powerful keyword access relies on *random access* to uncontrolled data. Library catalogs usually provide this but also emphasize browsing or *sequential access* to controlled data elements in discrete indexes. Library portals fall back on keyword access even when database driven from cataloging data.

Is it easier to find "library resources" on the Web than in the catalog? Every day, more full text and more metadata appear directly on the Web. Sometimes this sketchy metadata is directly associated with available full text--sometimes searchable with only a peek permitted and sometimes entirely separate. The following three examples illustrate the blurring of catalogs and Web searches; the disregard, de-emphasis, or disappearance of controlled vocabulary; and the meager bibliographic information even when full text originates from libraries.

1. Amazon.com's limited metadata is offset by additional features such as sales ranking (popularity); searches of full text (when available and after the book is found) via its "Search inside! the book" feature; CAPs (CApitalized Phrases, e.g. proper nouns, occurring in the full text); SIPs (Statistically Improbable Phrases to identify unique content); and access to teaser full text samples.

French Intellectuals And History. The
Nouvelle Revue Francaise under Jean
Paulhan, 1925-1940. (Faux Titre 93)

(Paperback)

by Martyn Cornick

Product Details

Paperback: 224 pages

Publisher: Rodopi Bv Editions (January, 1995)

Language: English

ISBN: 9051837976

2. Google Book Search allows searching full text of books, which are also searchable in regular Google, but reduces bibliographic information drastically. The inclusion of “Publication Date” is significant in view of disregard for dating Web content, and “Related information” recognizes bibliographic relationships.

Related information

- Web search for reviews of Cambridge History of Classical Literature
- Other web pages related to Cambridge History of Classical Literature

Bibliographic information

Title	Cambridge History of Classical Literature
Author(s)	E.J. Kenney
Publisher	Cambridge University Press
Publication Date	Apr 1, 1982
Format	Hardcover

Dimensions 6.25 x 9.50 x 2.25 in

ISBN 0521210437

Scanned library books have the same links for related information but even less metadata although pagination is sometimes given:

Bibliographic information

Title The History of Civilization

Author(s) Amos Dean

Publisher J. Munsell

Publication Date 1868

Bibliographic information

Title Michigan law review

Publisher Michigan Law Review Association, etc.

To date, it is apparent that Google does not encourage direct linking to full text books since the lack of a control number or Digital Object Identifier does not make this simple and persistent.

3. OCLC's "Find in a Library" maps selected cataloging data, including contents notes, and is searchable in Google or separately. Some codes are burst to text values. The

overall emphasis is on links to library holdings and provision of options for users to add contents and notes.

La Pléiade

by Yvonne Bellenger

Type: French : Book : Non-fiction

Publisher: Paris : Presses universitaires de France, 1978.

ISBN: 2130357733

Subjects: French poetry--History and criticism--16th century | Pléiade. | Pléiade.

The segregation of controlled bibliographic data elements in catalogs limits their usefulness. There should be ways of creating more peaceful coexistence between controlled and uncontrolled access on the Web. It should not be an either/or choice.

Undated Documents? Timeless

Many digital resources parallel published documents and include proper attribution and date of issuance. Too many other digital resources lack normative data elements, thus preventing retrieval by the date of content. Unfortunately, digital documents produced by libraries are not immune. An appeal to a search engine company suggested defining content dates for Web pages and using its clout to promote them. A representative from the company answered that it would not work as creators would falsify values to make content appear newer. Bracketed dates provided in cataloging records may become the only indication of the date of some Web content,

especially in cases where pages are moved and copied without content change since each time they receive new date/time stamps. This is another case of the digital environment reflecting too well the manual one of handwritten manuscripts.

While seemingly a small point, it is just one facet of what eventually leads to historical revisionism. Those concerned primarily with publicity and marketing tend to ignore facts by taking liberties that run the gamut from vague or inaccurate literature citations to what amounts to misrepresentation or falsification of the scholarly record. Some of the most egregious examples of this occur when back files of periodicals are mounted online retrospectively. This example lumps three titles under the current title:

Journal of Cancer Research and Clinical Oncology (Historical Archive)

Publisher: Springer-Verlag GmbH

ISSN: 0171-5216 (Paper) 1432-1335 (Online)

[covers 1904 to 1996; 1997- listed separately with same ISSNs]

The following earlier titles, each having separate ISSNs, are not mentioned.

1971-1978: Zeitschrift für Krebsforschung und klinische Onkologie 0084-5353

1903-1971: Zeitschrift für Krebsforschung 0301-1585

For a century, articles in these German titles have been cited by the title *as published*, indexed *as published*, and searched *as published*. This practice creates problems for link resolvers; and, when users reach the “right” place, they are confused in not seeing the title sought. The

separation of a current file from the archive hampers navigation by introducing an unnecessary boundary.

Citation practice is hampered further when online versions of publications drop pagination, withdraw superseded editions of books, etc. Similarly, organizations claim accomplishments under current names; and sometimes appear to try to eradicate their earlier names. Libraries have shown that information retrieval nightmares due to such practices are avoidable by respecting bibliographic history and the scholarly record as does Project Muse. NLM practice regarding retracted publications accurately reflects changes without compromising the original. Inclusion of the Publication Type, “Retracted Publication,” and displaying this relationship prominently at the beginning of the record corrects errors without altering the facts:

Retraction in: Clin Infect Dis. 2004 Jul 1;39(1):152.

Accurate reflection of bibliographic history, including the date of content, will benefit catalogers in the long run. As the Web grows and ages, time’s critical role in delineating information currency and historicity will become more evident to users.

Proliferation or Prevention of Errors?

Once created, deliberate or inadvertent misinformation can proliferate. This is particularly true of often-copied Wikipedia content but also occurs in copy cataloging. The Wikipedia entry for the MacArthur Foundation lists “John Cairns (biologist)” under 1981 fellowships, which used to link in error to a preacher who had died in 1892, until corrected on August 29, 2005. However, as of December 22, 2005, a Google search for “‘john cairns’

macarthur wikipedia” yielded 14 websites on the first two screens of a total 246 results. The results may be classed as correct (3), failed (5), and incorrect (6) as shown in the following sequence retrieved. Of the two screens of highly ranked pages in Google, only two had picked up the Wikipedia correction after several months.

Site	Result
1. en.wikipedia.org	Correct (John Cairns (biologist) with no link)
2. answers.com	Incorrect (link to biography of 19 th century preacher)
3. reference.com	Incorrect
4. searchspaniel.com	Failed (“Not Found” message)
5. biocrawler.org	Incorrect
6. factbug.com	Failed (redirect to search form)
7. thefreedictionary.com	Correct
8. lowide.com	Incorrect
9. oldion.com	Correct
10. peopledump.com	Failed (redirect to pornography)
11. mygoinfo.com	Failed (“Not Found” message)
12. carluvers.com	Incorrect
13. bharathnet.com	Failed (peculiar link to generic Wikipedia news)
14. netipedia.com	Incorrect

The LC Authority files include 10 different John Cairns, including “Cairns, John, 1922-,” the one who received the MacArthur Fellowship in 1981 although this information is not

provided. Another one born in 1923 is also a biologist. The *Biographical and Genealogical Master Index* lists 20 entries with some duplication, but none born in 1922. The MacArthur site lists him as “H. John Cairns,” but this search in Google only yields the same site. The *Boston Globe* announces his award under “Hugh John Cairns.”

Unfortunately, listings of major award winners do not lead directly into the bibliographic apparatus. Conversely, authority files do not identify instances of topical headings. The LC catalog does list works by this author, but nothing about him. A potential relationship:

100 1_ |a Cairns, John, |d 1922-

550 __ |e Awardee |9 1981 |a MacArthur Fellowship

Better integration of library catalogs, authority files, and reference tools are long overdue. If authority information were more apparent on the Web, those creating resources might avail themselves of using preferred forms and become more aware of the value of context and disambiguation. Currently, they are more likely to consult the Wikipedia, which does have disambiguation pages, but not in this case as there is only a biography of the 19th century preacher.

Sometimes there is a lot of noise and not much information. Elements that could lead to clarity regarding people, organizations, and events are sadly dispersed across incompatible systems, unarticulated with their counterparts. Disclaimers about the veracity of information likely go unread. Perhaps herein lies opportunity for the cataloger. Authorities for selected cyclical data, such as new Nobel laureates, could contain triggers to prompt creation of new instances as they are announced.

Extensible Relationships?

The importance of relationships in the bibliographic apparatus continues to gain currency. This could be due to the ease of “hotlinking” in online catalogs, the influence of hyperlinking in web documents, buzz in publications advocating this, or other factors. The influence of FRBR on the RDA hopefully bodes well in addressing this strategic aspect of bibliographic information. As rule simplification is a goal, a generic linking mechanism is easily imaginable. Currently, linking entry fields, along with some of their 2nd indicators, provide approximately 30 specific relationships, mostly relating to seriality. Consider the following examples that omit related control numbers, etc.:

780 00 |t Journal of the American Medical Association

760 0_ |t American journal of hygiene

767 0_ |t Finance & development. Spanish. Finanzas y desarrollo

785 07 |t Quarterly review of pediatrics

785 07 |t Clinical pediatrics

785 00 |t New England journal of medicine

In other cases, a linking entry complexity note (580) describes a “nonspecific” relationship:

580 __ |a "Companion to Rohen/Yokochi Color atlas of anatomy."

787 1_ |a Rohen, Johannes W. |q (Johannes Wilhelm). |t Color atlas of anatomy.

Linking entry fields could be reduced to one field (possibly 788) by defining a relator subfield (possibly |e) to explicitly name the relationships. Recommended display constants could serve as an initial set of values and be expanded to cover cases unnecessarily relegated to notes. Consider these entries paralleling the foregoing examples:

788 __ |e Continues |t Journal of the American Medical Association

788 __ |e Subseries of |t American journal of hygiene

788 __ |e Translated as |t Finanzas y desarrollo

788 __ |e Merged with |t Quarterly review of pediatrics

788 __ |e Continued by |t Clinical pediatrics

788 __ |e Continued by |t New England journal of medicine

788 __ |e Companion to |t Color atlas of anatomy (2003)

Other data specific to relationship would be better covered as part of the relationship field. Generally, serials have notes in a separate area, despite MARC linking entry fields including a notes subfield (|n). The following section contains an example of a note embedded in a relationship.

Whatever the reason that series are treated as a distinct ISBD area, they inherently express a relationship to an implied parent work, either via a numbered sequence of another title (indistinguishable from a serial) or via sharing a collective title alone. The simplifying mechanism could be extended by inclusion of a subfield for sequence note.

788 |e Series |t Advances in cardiology |v v. 10

Such practice would simplify a confusing array of fielding and support series display in numeric or alphabetic sequence in particular. Moreover, this would help in interfiling links for analytics and component parts.

▼ Advances in cardiology	[serial]
v. 10 Body surface mapping of cardiac field	[analytic]

Instead of:

Advances in cardiology	[serial]
Advances in cardiology ; v. 10.	[analytic]

Simplicity in structuring bibliographic linking would support more intuitive and advanced approaches to information management and use, e.g. links from maps to geographic entries; from a periodic table of chemical elements to entries for elements and isotopes; graphical displays of time lines including works created, active creators at the time, extant organizations, events of the time, works about the time period, places existing at the time, etc. Further refinements could permit literary maps of periods, chronologies of concepts in a field, and visual displays of relationships as seen in products from Antarctica Systems, KartOO, etc.

Relationships in Controlled Vocabularies

Development and use of controlled vocabularies are often associated with libraries and indexing agencies. These tend to focus on a topic or field of study with the aim of providing consistent terminology, cross-references from variant terms and related terms, and sometimes hierarchical relationships and other features--all focused on facilitating users' access to information. Vibrant schemes continue evolving to keep current with advances relevant to their scope and to associated reinterpretations of terminology. Due to such change, it is inevitable that previously valid terms persist in some databases, despite significant effort to maintain currency. Without constant vigilance and ongoing maintenance, obsolescence can seriously hinder the realization of the benefits of a controlled vocabulary. For example, how many libraries using *Medical Subject Headings* would claim their catalogs' full compliance with 2006 MeSH?

Insubordinate Subheadings?

While the National Library of Medicine's use of MeSH is exemplary, limitations other than obsolescence are more challenging, even for national libraries. A long-standing feature of both MeSH and LCSH (*Library of Congress Subject Headings*) is the use of subheadings. How many integrated library systems effectively use subheadings to eliminate clutter in alphabetical subject indexes (displaying the substructure on demand) or to break a large topic into organized subsections? Would it really be so difficult to display subject indexes to reflect the careful coding that went into them?

A subject browse in Lane's Voyager catalog recently retrieved 208 topics beginning with the word "heart." The following subheadings are widely dispersed amongst intervening terms, many of which also have subheadings, e.g. "Heart Valve Diseases" with 10 subheadings. One

may argue that a combined alphabetical list is useful, but what about the 50 entries beginning with “cardiac”? Is this vocabulary out of control?

Subheading	Position in Retrieval
abnormalities	2 nd (xref to precombined term)
anatomy & histology	3 rd
blood supply	37 th
drug effects	95 th
embryology	96 th
growth & development	117 rd
injuries	(xref suppressed; matches Heart Injuries)
innervation	124 th
metabolism	131 st
physiology	139 th
pathophysiology	140 th
radiography	141 st
radionuclide imaging	142 nd
surgery	162 nd (xref to precombined terms)
transplantation	(xref suppressed; matches Heart Transplantation)
ultrasonography	171 st

Hierarchical Relationships

Hierarchical structures can effectively reveal relationships as seen in the MeSH Browser⁷ and in PubMed.⁸ However, some of the content is not displayed there, and a great deal of it disappears in the context of the catalog. In 2005, MeSH was restructured via an XML schema into three levels: Descriptor, Concept, and Term.⁹ The influence of NLM's Unified Medical Language System (UMLS) is evident in provision for a richer array of relationships.

In trying to grasp this major reorganization, mapping the XML structure to the relatively flat MARC structure proves an interesting exercise. The quasi-MARC below distinguishes the typical inclusion of synonyms under a preferred term from the subsumption of terms with narrower meanings. Both are most often intermixed but distinguished in XML MeSH data.

035 __ |a (DNLN)D002397

150 __ |a Catfishes |3 M0003649 |4 T006951 |9 1989-11-17

450 __ |1 -- |a Catfish |3 M0003649 |4 T006952 |9 1999-01-01

450 __ |1 -- |a Siluriformes |3 M0003649 |4 T421188 |9 2000-08-09

450 __ |1 -- |a Cat Fish |h [local] |3 M0003649 |9 2006-01-02

450 __ |e Includes |a Arius |2 Genus of the family Ariidae (sea catfishes).

|3 M0367470 |4 T421189 |9 2000-08-09

450 __ |e Includes |a Eremophilus mutisii |2 A fish species of the order

SILURIFORMES, family Trichomycteridae. |3 M0438709 |4 T519813

|9 2002-09-18

450 __ |1 -- |e Eponym |a Colombian Catfish |3 M0438709 |4 T519814 |9 2002-09-18

450 __ |e Includes |a Heteropneustes |3 M0003651 |4 T006958 |9 1986-04-15

450 __ |e Includes |a Plotosus |3 M0003652 |4 T006959 |9 1986-04-15

550 __ |e Broader |a Fishes |w (DNLM)D005399

550 __ |e Narrower |a Ictaluridae |w (DNLM)D007059

680 __ |i Common name of the order Siluriformes. This order contains many families and over 2,000 species, including venomous species. Heteropneustes and Plotosus genera have dangerous stings and are aggressive. Most species are passive stingers.

This coding of a MeSH record for Catfishes has a unique Descriptor ID in field 035. The first group of fields includes the base Concept in field 150 and three variants in field 450. One of these, Cat Fish, illustrates with “[h [local]]” one way of distinguishing this local enhancement from NLM’s data. Note that each synonym shares the same Concept ID, given in |3. Other subfields are explained in the next group.

In the second group, four Concepts each have a relator “[e Includes” to indicate more clearly their subsumption under Catfishes. Note that each of these subsumed Concepts has a separate Concept ID, again in |3, implying that they could be established separately. Like the base Concept, synonyms of a particular subsumed Concept share its Concept ID. Each synonym for the base Concept or for a subsumed Concept is “indented” by use of a pair of hyphens in |1 to enforce the hierarchy. Similarly, “[e Eponym” uses a relator for type of subsumed Concept. Each term, regardless of type, has a unique Term ID appearing in |4. Notes specific to a subsumed Concept appear with the Concept in |2, and the date a specific Term was added appears in |9.

The third group includes references to related Descriptors and a public note. The mapping is fabricated but fairly literal. Additional fields not of interest were omitted, and the example of a local enhancement added.

The following group of fields could be included to subsume a new local Concept in the MeSH data. The only difference is an arbitrary local Concept ID to group the terms within this record.

450 __ |a Includes |a Malapteruridae |h [local] |3 L1 |9 2004-12-18

450 __ |1 -- |a Electric Catfishes |h [local] |3 L1 |9 2004-12-18

450 __ |1 -- |a Catfishes, Electric |h [local] |3 L1 |9 2004-12-18

Potentially, the data thus far could be displayed more transparently. Base concepts are in bold and subsumed concepts in italics:

Topic: **Catfishes**

Scope: Common name of the order Siluriformes. This order contains many families and over 2,000 species, including venomous species. *Heteropneustes* and *Plotosus* genera have dangerous stings and are aggressive. Most species are passive stingers.

Varia: -- Catfish
 -- *Siluriformes*
 -- Cat Fish [local]

Includes:

Arius Genus of the family Ariidae (sea catfishes).

Eremophilus mutisii A fish species of the order SILURIFORMES,
family Trichomycteridae.

-- Eponym: Colombian Catfish

Heteropneustes

Plotosus

Malapteruridae [local]

-- Electric Catfishes [local]

-- Catfishes, Electric [local]

Related: Broader:

Fishes

Narrower:

Ictaluridae

To further stretch the example, consider adding more local subsumption:

450 __ |a Includes: |a Ariidae |h [local] |3 L2 |9 2005-12-18

450 __ |1 -- |a Sea Catfishes |h [local] |3 L2 |9 2005-12-18

450 __ |1 -- |a Catfishes, Sea |h [local] |3 L2 |9 2005-12-18

450 __ |1 -- |e Includes: |a Arius |2 Genus of the family Ariidae (sea catfishes). |3 L3

|3 M0367470 |4 T421189 |9 2000-08-09

450 __ |1 -- -- |a Arius seemani |h [local] |3 L4 |9 2005-12-29

It is possible to introduce a local intermediate level subsumed Concept, Ariidae, in this record. The existing MeSH subsumed Concept, Arius, is logically subordinate to the new local one. Making the Concept ID subfield |3 repeatable could introduce processing problems. Adding another subordinate term, Arius seemani, implies the need for an awkward second level of indention.

The example in this rather tortured mapping exercise illustrates the limitations of the flat MARC structure. Elsewhere, the XML MeSH data exhibits two other types of relationship coding that this author has dubbed as ‘suprasumption’ (subordinate inclusion of a broader term) and ‘circumsumption’ (subordinate inclusion of a related term). These examples are not just curiosities, but illustrate the broader structural issues of maintaining clarity of relationships of an evolving topic. A preferred term becomes a sort of holding pen under which to lump other terms while maintaining the actual relationships among the subsumed terms until such time that these might merit separate headings.

This example further illustrates the challenges of localization when using a centralized data source. Another example is editing of bibliographic records, where local changes are not always reconciled when the host record changes. This blurred shift of locus of control to the local catalog may be offset by coding conventions to keep track of local practice. At a broader level, MARC’s |5 (Institution to which field applies) and RLIN’s clustering partially support these distinctions.

This lumping phenomenon is fairly common. Similar bibliographic challenges are found in latest entry serial cataloging, the single-record approach to print/digital versions with multiple holdings records, structured contents notes, etc. Currently, NLM’s MeSH in MARC format avoids problems by omitting the troublesome data and its substructure. Oddly, the MeSH

Browser and PubMed also do not display the added information found in the XML data. The advantages of cooperation are diminished by informal solutions. Libraries need more robust shared mechanisms to encourage inclusion and maintenance of complementary local data.

My Silo is Better Than Your Silo vs. Distributed Cooperation?

There are many interesting and successful examples of library-created websites, portals, repositories, etc. However, these do not seem much different than the plethora of competing commercial products in their striving to be *the* information destination for a target audience. Efforts to federate resources recognize the value of integration but do not fair particularly well when disparate data structures are involved. Better solutions, such as NLM's products and Elsevier's Scopus, have melded content and context by creative treatment of relationships and categorizations. However, these work well precisely because they largely *control* the creation and deployment of the data involved.

Silos or silo-complexes, including catalogs, remain islands of organization in a generally chaotic Web environment. The boundaries have blurred further, e.g. Google's crawling the controlled content of PubMed. Thus, PubMed gems sometimes pop up in Google searches. However, such controlled metadata often occupies the bottom of the Google heap since few Web pages link directly to them. These hits also represent an ill-defined, unevenly-updated, partial subset of the host data. The lack of transparency in Google's coverage is unfortunate. The tendency of search engines to obscure their workings and make retrieval appear to be magic is questionable.

Is control or chaos the only choice? Could we compete to find better ways to cooperate? Could we create generalized ways to add specialized metadata? Developing robust mechanisms

aimed at coordinated retrieval in a distributed digital environment could take advantage of both useful controls and widespread Web technologies. Libraries have a history of cooperation, but are union silos enough? Structured metadata intended for retrieval on the Web could add coherence without being prescriptive. Emphasis on structured relationships could facilitate retrieval in search engines if citations of a work represented by its metadata could be made equivalent in relevance to Web page links.

Taken together, libraries' scarce catalogers could accomplish more by coordinating efforts and upholding higher standards in the open Web environment rather than by duplicating efforts. By seeking ways to address both local and shared needs concomitantly, cataloging could be recast and potentially enjoy a renaissance. The distributed creation and maintenance of quality metadata in a virtual structure directly on the Web could consolidate our singular diluted efforts and garner recognition of our unique contributions. Instead of increased polarization, we need to learn from cooperative efforts such as the Wikipedia and seek to balance control and chaos. Even Google recently has recognized that they and librarians share the same mission.¹⁰ Considering what has happened since Dr. Lane's observation, we can expect great things in the future; but the wisdom of balancing past lessons, present exigencies, and future prospects remains as true today as in 1886.

REFERENCES

¹ Levi Cooper Lane, "Things old and new, with a chapter from Caelius Aurelianus," *Pacific Medical and Surgical Journal and Western Lancet* v. 29, no. 6 (July 1886):398-9.

² Joint Steering Committee for Revision of Anglo-American Cataloging Rules, *RDA: Resource Description and Access*, (Dec. 20, 2005), <http://www.collectionscanada.ca/jsc/rda.html>.

³ La Pléiade (Jan. 6, 2006). In: Wikipedia, <http://en.wikipedia.org/wiki/Pleiade>.

⁴ Martha M. Yee, "New Perspectives on the Shared Cataloging Environment and a MARC21 Shopping List," *Library Resources & Technical Services* v. 48, no. 3 (2004):165-78.

⁵ Dick R. Miller and Kevin S. Clarke. *Putting XML to Work in the Library* (Chicago: American Library Association, 2004), p. 101-136.

⁶ Nancy Blachman, *Google Guide*, <http://www.googleguide.com/>. Although Google cooperated in this endeavor, they advocate that the system should be intuitive and not need extensive documentation.

⁷ National Library of Medicine (U.S.), *MeSH Browser*, <http://www.nlm.nih.gov/mesh/MBrowser.html>.

⁸ National Library of Medicine (U.S.), *PubMed*, <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>. Select MeSH on pulldown menu.

⁹ National Library of Medicine (U.S.), *Introduction to MeSH in XML Format*, (Oct. 4, 2004), <http://www.nlm.nih.gov/mesh/xmlmesh.html>.

¹⁰ Google's Newsletter for Librarians, (Dec. 2005), http://www.google.com/newsletter/librarian/librarian_2005_12/newsletter.html. This is also called Google Librarian Newsletter, 1st ed.