An Open Reply to

What is Going on at the Library of Congress? by Thomas Mann

by James Weinheimer
Director of Library and Information Services
American University of Rome
Rome, Italy

07 July 2006

The judgements made in this paper do not represent official views of the
American University of Rome.

References:
What is Going on at the Library of Congress? / by Thomas Mann. June 19, 2006.
http://guild2910.org/AFSCMEWhatIsGoingOn.pdf

The Changing Nature of the Catalog and Its Integration with Other Discovery Tools.
Final Report. Prepared for the Library of Congress by Karen Calhoun : A Critical
Review / by Thomas Mann. April 3, 2006.
http://guild2910.org/AFSCMECalhounReviewREV.pdf

The Changing Nature of the Catalog and its Integration with Other Discovery Tools :
final report / prepared for the Library of Congress by Karen Calhoun. (March 17, 2006)
http://www.loc.gov/catdir/calhoun-report-final.pdf

Mr. Mann has written two very thoughtful essays about the future of cataloging, and there
have been similar essays by other librarians in various places and the AUTOCAT
listserv.[1] I feel that I must make my own reply. I have worked as a cataloger for several
years in different types of institutions, from those dealing with MARC21/AACR2, to
UNIMARC, to international institutions that use non-MARC systems and rules. I am also
the author of a highly technical cataloging manual located on the web,[2] so I believe I
understand the issues of cataloging. I appreciate the support and high value that Mr.
Mann and the others place on the catalog record, and I agree with many things they say,
but the fact that I have worked in an international environment and been a cataloging
manager also makes me somewhat more sympathetic to the points raised by Deanna
Marcum and the demands for increased production.[3] This does not mean that I
necessarily agree with some of the recent decisions by the Library of Congress
(specifically over the elimination of series authority control), but it is very clear to me
that the practice of cataloging is changing, and will continue to change substantially. In
fact, it must change and this is a truth that we all have to learn to live with.

There is nothing wrong with change, although some changes may be disruptive and make
us uncomfortable. In many ways, some of the changes that I foresee should bring needed
relief to overworked librarians, so long as the changes are carried out in full awareness of
their consequences. The complexity of understanding these consequences is illustrated by
LC's decision to get rid of series authority. It is obvious they did this to increase
cataloging productivity. I cannot see any other reason for it. Series authority is not a

---

[1] AUTOCAT, a cataloging listserv. http://ublib.buffalo.edu/libraries/units/cts/autocat/. See primarily
David Bade's excellent email *LC series etc.,* from 31 May 2006, and the associated comments on the list.
LC's series decision has provoked a great deal of comment from various quarters including the Program for
Cooperative Cataloging and OCLC.
[2] *Slavic Cataloging Manual*. Latest version at: http://www.indiana.edu/~libslav/slavcatman/. Original
version at: http://library.princeton.edu/departments/tsd/katmandu/sgman/smtocs.html
[3] Quoted by Mr. Mann in his May 5th, 2006 letter in *What is Going on at the Library of Congress?*

simple task and is considered by many monographic catalogers to be one of the more tedious parts of the job. Many (including myself) find personal/corporate name and subject authority far more interesting.

Still, Mr. Mann, and others, maintain that series authority is important to the users, and this argument seems rather pointless to me. I am sure that the managers at LC who made the decision heard very similar arguments but nevertheless decided that continuing series authority simply isn't worth the effort. In the managers' defense, I can say that when I have trained new catalogers, merely teaching the difference between a traced and untraced series can be quite difficult for otherwise intelligent people to understand. Obviously, it's a complex concept. And let's face it: most users don't even know what a series is, much less an authorized form of a series, so why maintain a time-consuming practice that the vast majority of users do not find useful?

These arguments are irrelevant to the decision of whether to maintain series authority, however. The fact is, series authority is *absolutely vital* for the job of acquisitions and selection. This means that these basic library tasks *cannot be done* without series authority. Selectors and other acquisitions staff must know what a collection has received and is currently receiving within certain series. The same goes for reference work: it simply *cannot be done* without series authority. Consequently, for the task of maintaining the library's collections—which certainly is in the interests of users—series authority *must be done*.

This argument is far too technical for users to understand and they would never phrase it this way, but our users trust us to do a good job of selecting materials, maintaining the collection, and providing excellent reference service, and series authority is an indispensable tool to manage the collection. In light of this, what are the consequences of LC's decision?

Series authority of some sort will continue to be done by individual selectors and/or acquisitions departments. They will use spreadsheets, or (who knows?) reuse some of the old catalog cards lying around to get the access they need. Reference librarians doubtlessly will do the same. If the two departments cooperated, they could save a lot of work, but they probably won't and the library will revert to an earlier time when there were many little personal files scattered everywhere. Knowledge of the existence of these files will be by word of mouth and even for those in the know, to make a comprehensive series search will demand a lot of knowledge and take a great deal of time. These people will not be trained in the intricacies of series authority (which isn't simple) and there will be plenty of misunderstandings and errors.

As a result, although there may be some savings for the catalog division, there will be no savings at all for the library as a whole since the work has simply been off-loaded. The other departments will not appreciate having to do work traditionally done by the cataloging staff, and the other departments will see this as further proof that catalogers live in their own world and care little about the rest of the library and the users.

This is why I am against LC's decision to eliminate series authority: it won't increase productivity for the library as a whole, but will decrease it, and it will further alienate cataloging from the rest of the library.

But these considerations are not enough: we must argue in a more positive fashion: if not series authority, what would really increase productivity? Are there any procedures or rules or anything that could be eliminated or streamlined without simply offloading the work onto different departments of the library, or maybe even the users themselves? Or is the arrangement of everything so perfect as it is now, and the materials handled in the best, most efficient way, that if anything is changed or eliminated, the entire edifice will come crashing down?

I believe that there are many areas that can be streamlined and/or eliminated, but from his essays, it seems that Mr. Mann takes the second stance: that any changes will present grave risk to the entire bibliographic apparatus. I want to emphasize that I am sure this is not true, but this is the way I read his essays and those of others who put forward similar arguments. I believe this is how other managers will read his essays as well. Although Mr. Mann may be right in many particulars, this is not a way to convince the decision makers.

In my view, experienced catalogers must get involved as active and positive participants in the search for increased productivity. If all they do is point out how nothing can be changed without offering substantive alternatives for improvements, they will simply be ignored. Cataloging (along with the entire library) *must become more productive*. What can be done to improve things for everyone?

Our collections are no longer isolated. The World Wide Web eliminated that forever. It's not that collections really were isolated before the web, but catalogers could more or less ignore other collections to concentrate on their own. Cooperation in cataloging has been justified primarily by the idea that it helps to manage the local collection. For example, participating in cooperative name authority programs reduces the amount of work needed to maintain a local authority file. Before the web, it was exceedingly painful for users to search outdated printed catalogs of different collections, or to have to ask a librarian to search OCLC to find out if something was available in another library, and then decide whether to make a trip to the other collection, wait for interlibrary loan, or just forget it. Now everything is far easier for the users. No single collection—even the Library of Congress—can answer all questions. Our users need to know this, but I believe it's becoming more and more obvious to them as they explore the treasures (and the sewers) of the World Wide Web. This is the environment that we have now, whether we like it or not.

The question arises: Why does cataloging have to become more productive? The obvious answer is simply that this is the nature of modern management. Perhaps some of us have a romantic notion that research and scholarly institutions are beyond the scope of modern management, but anyone who has been involved in the management processes in one of these institutions knows the intricacies of the budgeting process, SWOT analysis, critical

path optimization, and so on. Higher productivity is unavoidable in modern management, and overall, I personally believe that it represents a positive step for libraries.

A more compelling reason why cataloging must become more productive is that the nature of the library's collection has changed fundamentally. There are many, many, many excellent materials available on the web. A lot of them are free—which doesn't mean they are inferior—and our users like them. With the different projects from Google, Carnegie Mellon, and innumerable institutions around the world, these materials will soon dwarf many of our own local collections, if they have not done so already. Therefore, we are faced with something entirely new: a truly shared collection that can be used and consulted just as easily as the volumes in our own physical collections.

These electronic materials aren't perfect, but people like them a lot anyway. After all, if I can get a free electronic copy of H.G. Wells' "War of the Worlds" and then if I want, I can have a decently printed and bound copy for one or two dollars from a print-on-demand service, why should I go out and spend five or ten dollars for a traditionally published copy? And when some company comes out with a decent hand-held electronic screen that people can read as easily as a printed piece of paper, only an archive will need a paper copy. This may happen very soon, for there will be electronic paper, or do we really believe that it is easier to make humans fly and go to the moon than to make some kind of screen that is easy to read? These things may not be perfect, especially at first, but I read that people don't seem to mind watching television shows on their IPODs.

There are excellent materials being placed into Open Archives every day, available to everyone for free. We should also keep in mind that each time someone downloads a resource from the web, this represents an interlibrary loan that has been avoided.

All these are changes for the better, but they represent fundamental changes for publishing and libraries. We ignore these new types of materials at our peril, especially if in the same breath we insist that users need series authority control. Juxtaposing these arguments would sound very strange to many users.

If libraries resolutely ignore the materials that users have demonstrated they want, users will opt for Google or Yahoo. We essentially force users *not* to use our tools, and we become more and more quaint and obsolete in their eyes. Some of us may not care how we are viewed by users because we feel that the electronic materials are in some ways "inferior," but library managers care a lot about how users feel. For many managers, user feelings are the most important goal of all. After all, it's the users who control the budgets—not us.

Today, it is just as easy for our users to search catalogs of other institutions as our own local catalogs. I can search French catalogs[4] just as easily as Italian catalogs[5] just as easily as Russian catalogs[6] just as easily as catalogs for parts of the United Nations,[7] to

---

[4] e.g. http://www.bnf.fr/pages/catalogues.htm
[5] e.g. http://opac.sbn.it/cgi-bin/IccuForm.pl?form=WebFrame
[6] e.g. http://corporate.gpntb.ru/

name just a few. Our users want the materials described by these catalogs, and many resources are available electronically for free. Users (including myself!) want these materials. Our users would use digital materials far more often if they knew what existed and where they are, but users don't know how to find them. And although we are discussing formal catalogs here, not every one uses the same standards, other libraries in the world use different formats, different cataloging rules, different authorized forms ....

Hey! It sounds like we need a librarian to sort this out!

These are just some of the tasks that I believe catalogers will be doing in the future. It certainly doesn't mean reducing or eliminating standards, but getting the different standards to work together somehow. For example, a user should be able to effortlessly find all the works about Mark Twain in US, German and Italian libraries, Open Archives, as well as in other databases that practice personal name authority control. Systems are powerful enough today to allow this type of search, but they don't yet exist and need to be built. If we can show how useful this kind of control is, perhaps our users won't be quite so happy with Google and Yahoo. Maybe we could even get Google and Yahoo to join in. So long as we're imagining, let's imagine all the way!

OK. So There's a Lot of Extra Work. Where are the Savings?
Every institution is facing the same problems, so there are opportunities for cooperation. Computer systems have advanced in some very handy ways to help provide solutions. One example of savings is that while users can search other catalogs very easily, a cataloger can too, so a record created in, e.g. Russia, can be taken and placed into my own catalog. Of course, their format is in RUSMARC, but they may be able to provide MARC21. Even if they don't, systems can be built that could convert the formats on the fly. The Russians mostly follow ISBD so at least most of the description can be shared. Many other databases have their own rules and don't follow ISBD, but I believe there will be tremendous pressure on them to change from their local rules to more internationally accepted rules.

Concerning name headings and title headings, the Russian-language ones naturally would have to change to the forms in the Name Authority File, but can we imagine a way where we could save at least some labor in the conversion process? Is there a way to save 75% of the labor; 50% of the labor? Even saving 10% of the task of assigning headings through sharing and cooperation would represent a lot. What system would we need to build that would allow it? Of course once the system is built, the Russians would be able to convert our forms back into their own as well.

Publishers have their own databases and are willing to share their internal publishing information through a format called ONIX. It includes information important only to publishers, such as the weight of the book. If the publishers would be willing to change just a bit, libraries might get ISBD information from them. Maybe even more. So far as I

---

[7] e.g. http://www.fao.org/documents/

am aware, ONIX records are used only by the Library of Congress right now, but these records could offer many more opportunities.[8]

With some cooperation and changes, suddenly the number of useful copy cataloging records could increase exponentially. This could be a great help to overworked librarians.

Perhaps we could simplify MARC21 radically. After all, it was created in an earlier time, and there are parts that aren't used very often, if at all. I'm not advocating throwing it all out, but streamlining it. For example, what is the use for a lot of those subfields, or are they simply unnecessary baggage complicating cataloger training, inputting conventions and the computerized catalog for no good reason? For example, where is the added value in the subfields b and c below? The heading displays the same and searches the same.

> 100 0_ **|a** Napoleon **|b** I, **|c** Emperor of the French, **|d** 1769-1821
>
>   vs.
>
> 100 0_ **|a** Napoleon I, Emperor of the French, **|d** 1769-1821

Could the following coding (not the heading, itself) be simplified?

> 130 0_ **|a** Bible. **|p** N.T. **|p** Acts and Epistles. **|l** English. **|k** Harmonies. **|s** American revised. **|f** 1906.

What is the utility of some information in the fixed fields? How about that Festschrift code? How many users even know what a Festschrift is? If no one can show precisely where and how these subfields are useful and/or demonstrate some concrete plans to utilize this information, they should be seriously considered for elimination.

Could subject headings be simplified and/or be made more useful to our users? Or can nothing be changed of any substance? According to the LC Authorities page, there are 265,000 subject authority records.[9] That's a lot for users and catalogers to use actively. How well have they been used? Research has shown—consistently—that there has always been a great amount of "interindexer inconsistency."[10] "Interindexer inconsistency" represents how often different catalogers, who are all fully trained, assign the same subjects to the same item. It rarely happens. Some interested people have interpreted this research to mean that subject headings are useless.

To a casual observer, this may appear to be a logical deduction, but it is not. Yet, we need to explain why. We should not ignore this type of research out of embarrassment, but rather deal with it honestly and seriously to discover ways for improvement or change. Could subject analysis from a foreign library be useful in some way for a US library? What systems would be needed to allow this? Again, even a savings of 10% would be substantial. I want to emphasize that I am not at all advocating eliminating subject headings, because with new types of search mechanisms, I believe they can be utilized to

---

[8] For some sample records, see http://dali.cdlib.org:8080/onix/onix.cgi
[9] http://authorities.loc.gov/help/contents.htm
[10] For only one study, see: *Indexing Consistency and its Implications for Information Architecture: A Pilot Study* / Hope A. Olson & Dietmar Wolfram, specifically the "Previous research" section. http://www.iasummit.org/2006/files/175_Presentation_Desc.pdf

greater advantage than ever before, although it does appear that we need modified standards.

Mr. Mann likes browsing. I do too, but users need to be aware that browsing is not the best way to find information. The example that he provides of "browsing" to find information relevant to the Dreyfus case is interesting, but ignores the fact that most libraries have no choice except to store books offsite and that many libraries have multiple locations, so that a reliance on browsing the shelves guarantees that users will miss many materials. I have met some users who believe that if they browse the shelves "well enough," they don't have to use the catalog at all!

The Library of Congress Classification is only one way to arrange information, and there are many other classifications, such as Dewey, Universal Decimal, Colon, BBK and so on. Another system of classification might well classify the original documents of a trial far away from the subject books on the person(s) involved, and the example Mr. Mann gives would fail. Still, perhaps something else equally relevant and "serendipitous" could have been found in another arrangement. As an aside, I have found that serendipity works quite well in Google—or, perhaps it works too well.

Does this mean that there is no need for classification? For physical materials, they certainly need to be arranged somehow, and it was established long ago in the US that the arrangement should be by subject, and the basic arrangement of the subjects was determined in the late 19th and early 20th centuries. Other countries have come up with different arrangements and they seem to have survived pretty well. For electronic materials, perhaps there is no need for classification so long as there are good subject headings. But another idea would be to cooperate with libraries around the world to add multiple classification numbers. The catalog records (although not the items themselves, of course) could be sorted by Colon, Dewey, LCC, Bliss or other ways. It seems that this could only improve serendipity.

Finally, there have been several complaints among catalogers themselves about the quality of the catalog records being produced. When catalogers take records created by another library for their own catalogs, they often have to correct practically everything. Much of the problem is that there is little or no control over the people who are allowed to create records for inclusion into the shared databases. The person who made a record can range anywhere from a student with two or three days' training, to professionals who spend their entire careers creating records. Perhaps it is time to consider introducing cataloger certification by the national agencies, similar to the American Bar Exams for attorneys. This would raise the prestige of catalogers and the practice of cataloging, while improving the quality of catalog records.

If libraries are to play an important role in the World Wide Web, which I believe they should, it will be through their catalogs and the methods they choose to merge their catalog records with the wider world of metadata. Records meeting high standards and quality are the only products we have to offer, since records of no standards and no quality can be made automatically for next to nothing.

There is a movement among the World Wide Web gurus called "The Semantic Web." It is the "hot topic" right now. In my own opinion, the Semantic Web is exactly what we are discussing: standards and quality, interoperability with all kinds of systems, and various ways of achieving these goals. Librarians and catalogers have dealt with the practical aspects of these issues as much or more than anyone else.[11]

These are just some of the tasks that I believe will be done in the future, and these tasks will be done either with or without the help of experienced catalogers. I'm sure that some managers would absolutely love to eliminate MARC21 and use simple Dublin Core, or eliminate formal subject headings and go with folksonomies. At first glance, it all seems to be a lot cheaper and productivity would "rise." We need experienced catalogers not only to explain why these ideas would be bad, but to advance their own, more realistic suggestions for improvement.

It should be clear from my comments that I believe catalogs and cataloging are just as necessary today as at any time in history, if not more so, but the world has changed almost overnight. The World Wide Web became a major force only about 10 or 15 years ago, and the changes it has brought have been disorienting for many, including myself. There are tremendous opportunities now, but no matter what else happens: things will continue to change. I'm sure many managers would agree with the words of Eldridge Cleaver, "if you're not part of the solution, you're part of the problem." It's a new world, and we need new ideas to more fully utilize the bibliographic edifice that has been built by our predecessors for such a long time, and with such great care.

---

[11] One of the best introductions is: The Semantic Web In Breadth / by Aaron Schwartz. http://logicerror.com/semanticWeb-long. Or, there is a speech given April 5, 2006 at Princeton University by Tim Berners-Lee, "The Future of the Web," which is available for viewing at: http://www.princeton.edu/WebMedia/lectures/