5-31-2017

What About Classification Bias?: Channeling Sandy Berman

F. Tim Knight
Osgoode Hall Law School of York University, t.knight@osgoode.yorku.ca

Follow this and additional works at: http://digitalcommons.osgoode.yorku.ca/librarians

Part of the Cataloging and Metadata Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 License.

Repository Citation
Knight, F. Tim, "What About Classification Bias?: Channeling Sandy Berman" (2017). Librarian Publications & Presentations. 25.
http://digitalcommons.osgoode.yorku.ca/librarians/25

This Conference Proceeding is brought to you for free and open access by the Law Library at Osgoode Digital Commons. It has been accepted for inclusion in Librarian Publications & Presentations by an authorized administrator of Osgoode Digital Commons.
Today I’m going to talk about three things: classification bias, mostly in terms of the Library of Congress subject headings; Sanford Berman, who fought to address that classification bias; and finally, I’ll touch on what we might do to try and deal with classification bias going forward.
As a profession, and this is especially true in academic libraries, we have come to rely on two interrelated classification systems.
The Library of Congress Classification scheme ...
...and the Library of Congress Subject Headings or LCSH.

Both of these tools have allowed us to engage with our library users and provide an effective way to organize, find and think about our library collections.
And lately I've been thinking about how these classification systems will work in an information ecosystem that exists outside of our so-called “bibliographic universe.”

In their book, “Sorting Things Out”, Geoffrey Bowker and Susan Leigh Star speak about the work of information scientists and make the following observation:
“Information scientists work every day on the design, delegation, and choice of classification systems and standards, yet few see them as artifacts embodying moral and aesthetic choices that in turn craft people’s identities, aspirations, and dignity. Philosophers and statisticians have produced highly formal discussions of classification theory, but few empirical studies of use or impact.”

In other words, we think a lot about how to put a classification system together, but we don’t tend to think about how these classification systems affect the people that use them.
Our collections have usually been formed within the walls of a single building or as part of an institution. The collected resources are intended to support the information needs of our library users and the resulting classification systems help find and discover relevant resources that we’ve acquired.
This principle, known as “literary warrant,” is a method that essentially ensures that our classification system directs users to things that are actually held in the library.
Or, to quote Elaine Svenonius, this means that “the subject vocabulary is empirically derived from the literature it is intended to describe.”

It’s useful to remember that LCSH was never intended to describe or represent everything, be it all published work or the scope of human knowledge. If there are knowledge gaps in our collection, and more specifically in the Library of Congress’s collection, those gaps will also be found in the classification system.
“Literary warrant” does not on ontology make.

And when you take a collection of terms based on literary warrant and drop them into a cultural situation that exists outside of the North American context the intrinsic biases become much more apparent.

One person who came to realize this early in his career was Sanford Berman.
For those of you who may not know much about Sandy Berman here’s a little background. He was born in Chicago in 1933 and relocated with his family to Los Angeles when he was about eight years old. In 1955 he graduated from UCLA with a political science degree and minors in English, sociology and anthropology. He went to library school at the Catholic University of America graduating in 1961 with a library of science degree and a minor in history. In Wikipedia’s list of notable CUA alumni Berman is characterized as a “radical librarian,” Berman characterized the library program at CUA as “dull and irrelevant.” He worked in West Germany, for a couple of years returning to Los Angeles in 1967 to hone his skills as a serials librarian at the UCLA Research Library. In 1968 he went abroad again working as a serials librarian at the University of Zambia Library and the Makerere University Library in Kampala, Uganda. He returned to the U.S. in 1973 where he became the Head of Cataloguing at the Hennipen County Library in Minnesota a position he held until his retirement in 1999.
In 1971 Scarecrow Press published his seminal work “Prejudices and Antipathies” which was later reissued by McFarland and Company in 1993.
In the introduction to the 1971 edition Berman described his experience with LCSH like this:

"... in the realm of headings that deal with people and cultures--in short, with humanity--the LC list can only ‘satisfy’ parochial, jingoistic Europeans and North Americans, white-hued, at least nominally Christian (and preferably Protestant) in faith, comfortably situated in the middle- and higher-income brackets, largely domiciled in suburbia, fundamentally loyal to the Established Order, and heavily imbued with the transcendent, incomparable glory of Western civilization. Further, it reflects a host of untenable--indeed, obsolete and arrogant--assumptions with respect to young people and women. And exudes something less than sympathy or even fairness toward organized labor and the sexually unorthodox or ‘avant-garde.’"

Berman frequently criticised LCSH in letters to the Library Journal and other publications. It was those letters that attracted attention and inspired Scarecrow to have Berman collect and summarize his thoughts and his suggestions for the improvement of LCSH.
In Steven A. Knowlton’s 2005 paper, “Three Decades Since Prejudices and Antipathies” he reports that 39% of Berman’s recommendations were changed almost exactly as he had suggested. An additional 24% were changed in ways that partially reflected Berman’s suggestions and about 37% were left unchanged for reasons that Knowlton outlines in his paper.

However, despite this relative success, it is clear that some problems remain and I thought it would be useful to highlight a few examples to illustrate some of the issues that Berman had identified.
One heading that has since been changed, dealt with social roles and various “classes” of people. It was presented in a form like this: <class of person> as <role>. So using ‘Women’ as an example ...
... there were headings for ‘Women as accountants’, ‘Women as architects’, etc. The implication being, as Berman said, was that women were not considered to be “ordinarily competent or otherwise equipped to work at accountancy, bear arms, or fly to the moon.”

Berman also quotes Joan Marshall who put it more bluntly saying that this approach, “does not merely imply, it states that all segments of society other than white Christian males who achieve (anything) are merely role-playing.”
These headings were modified to acknowledge that women do have these roles in society. So now you’ll see ‘Women accountants’ and ‘Women astronauts’, etc. But you won’t find any equivalent headings for Men like ‘Men accountants’ or ‘Men soldiers’, etc. And maybe it’s just me, but this is so societally ingrained, that I actually find it strange to say something like, ‘Men astronauts.’
So now instead of having ‘Women as accountants’, you have ‘Women accountants,’ as a narrower term, under the main heading ‘Accountants.’ Which unfortunately perpetuates the view that this is still an unusual role for women to occupy or, at the very least, it’s subordinate to the role men play.

And there are still many examples like this in LCSH.
Another area that is still troubling is LC’s use of the term ‘Primitive.’ Berman comments that ‘primitive’ is “heavily overlaid with notions of inferiority, childishness, barbarity, and ‘state of nature’ simplicity, whereas the societies, arts, economic modes, music, and religions it purportedly covers may be extremely complex, ingenious, creative, humane, and ... admirable.”

He also noted that, “most anthropologists have renounced the term as unhelpful and ‘loaded,’ and an American Library Association committee in 1983 made specific, sensible suggestions for either cancelling or replacing nearly every ‘Primitive’ descriptor.”
But despite this, these suggestions were ignored, evident for example, in the current scope note for ‘Primitive societies’, which instructs:

“Here are entered works on nonliterate, nonindustrialized peoples as representatives of an early stage of social evolution based on largely 19th century theories of cultural evolution.”

The reference to social and cultural “evolution” are notable. This mindset supports an argument that considers newer or more “modern” societies as somehow naturally superior.
The related subdivision ‘--Discovery and exploration’, also reinforces this “evolutionary” point of view.

However, as Berman points out: “Cortez no more ‘discovered’ Mexico for the Aztecs than Livingstone did Victoria Falls for the Leya people ... [this term] represents an insult to the many peoples and lands which, so it appears in our library catalogues, didn't really exist until Whites happened to notice them.”

So regrettably, elements of “White chauvinism” continue to inform LCSH.
And there are many examples that Berman loosely groups together under race, ethnicity and religion.
This heading, for example, seems relatively harmless. There’s an instruction that also states that ‘intelligence levels’ can be applied as a subdivision under classes of persons and ethnic groups.

<table>
<thead>
<tr>
<th>Intelligence levels (May Subd Geog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA subdivision Intelligence levels under classes of persons and ethnic groups</td>
</tr>
<tr>
<td>UF Intelligence quotient</td>
</tr>
<tr>
<td>IQ (Intelligence quotient)</td>
</tr>
<tr>
<td>NT Genius</td>
</tr>
<tr>
<td>Mental age</td>
</tr>
<tr>
<td>BT Educational psychology</td>
</tr>
</tbody>
</table>
But it becomes problematic when you discover that the only class of persons this subdivision has ever been applied to in the LC catalog are for ‘Children’ or ‘African-Americans’.
Here’s a recent example using ‘African Americans--Intelligence levels’ for a book published in 2002. It’s very interesting to compare these subject headings provided by the Library of Congress ...
... with the headings used by the cataloguers at the Lillian Goldman Law Library at Yale.

I haven’t had the opportunity to examine this book first hand, but this does make me wonder whether these headings actually refer to the same book?

The Yale subject analysis represents a much more inclusive description of this book.
Seymour Lubetzky once characterized Berman’s criticism as “emotional” with a penchant to look for “sinister political significance.” I’m not sure if Lubetsky had these two headings in mind when he said that, but it’s certainly notable that even today only headings for ‘Jewish capitalists and financiers’ and ‘Jewish bankers’ are established in LCSH.

Berman suggested LC either add similar headings for the Lebanese, Christian, Anglo-Saxon, or Teutonic equivalents or to drop these headings altogether.

Neither recommendation was taken up and these two headings still stand on their own in LCSH.
The point being, why is a heading like this even necessary? Consider this biography of German-born English banker Siegmund Warburg.
Is anything lost if we don’t include the heading ‘Jewish bankers--Biography’? Surely the heading ‘Bankers--Great Britain--Biography’ is sufficient.
One of Berman’s letters penned in 1969, this time to the Library Journal, got a response from A. C. Foskett who defended the described biases by explaining that LCSH is “designed [to be used] in Western libraries” and so “reflect the historical bias of those libraries and their users.” Foskett did concede however, that there may be terms that some might consider “objectionable” within “the context of a different tradition.”

To this Berman countered with:

“What not ‘objectionable’ anywhere? Just because the scheme germinated, historically, within a Western framework of late Victorianism, rampant industrial expansion, and feverish empire-building (with its "White Man’s Burden" rationale)--just because, in short, we were ‘brought up that way’ is no valid reason for perpetuating, either in our crania or catalogues, the humanity-degrading, intellect-constricting rubbish that litters the LC list.”
In their recently published article about cognitive justice and indigenous knowledge, Heather Moulaison Sandy and Jenny Bossaller, essentially agree with this sentiment. They note that the systems we rely on, for example, “fail to provide access to indigenous or traditional knowledge from the point of view of the people whose ideas are being represented.”

And they feel we should strive to provide “cognitively just subject access.” Such an approach “… tends to reject the language of universal human rights as following an unrealistic and particularly Western notion, and seeks to replace that language with autonomy, dignity, and a ‘commons’ approach to cultural authority … the object is … to promote healing and forgiveness by removing the continued burden of colonialism and legacy thinking.”
Eunice Kua, for example, speaks about her experience setting up a library for rural South African high school students. They discovered that the Dewey Decimal Classification scheme they’d planned to use did not have a place for their local language. She says:

“What does it say to a child, when all of the categories in a system seem to accentuate what is not yours, while all the practices and wisdom of your culture are relegated to a tiny sliver of space? What is it like to grow up in a world where unfamiliar languages are revered, where your mother tongue may be good and useful for everyday life, but is not a vehicle for advancement?”

Eunice Kua, “Non-Western Languages and Literatures in the Dewey Decimal Classification Scheme,” *JLMA* 54, 4 (2004)
Kua also echoes Svenonious when she later points out that a “classification scheme reinforces the social systems that were in place at the time the system was devised, and projects, no matter how subtly or ineffectively, the social, moral and intellectual values of that system.”

And the Library of Congress subject headings are no exception. These subject headings have been “actively maintained since 1898” by generations of librarians, and represent the cumulative influence of the many worldviews that helped develop this controlled vocabulary.
I mentioned earlier that building a subject authority file based on literary warrant does not result in an ontology. And, because LCSH is based on traditionally published materials it also does not provide a “cognitively just access to resources for those outside of the Western mainstream.”
Sandy and Bossaller suggest two alternatives to literary warrant: ‘user warrant’, which is based on the language of the end-user, and ...
... ‘structural warrant’, which addresses hierarchical gaps in the system.
And they also propose a third type of warrant called ‘indigenous warrant’ where the “... terms and potentially classification structures are derived from the worldview of the indigenous peoples themselves, not from the dominant cultures who write about them or who search for information about them.”
Which brings me back to my thoughts about interoperability and whether our legacy subject systems provide a good fit when applied to the greater information ecosystem. We've seen a few examples of what the principle of literary warrant has created. And hopefully you agree that these are problems are something that need to be addressed.

But how do we fix this? Is there a technology-based solution?
In 1989 Marcia Bates wrote about the potential of the then emerging online systems in an article called, “Rethinking Subject Cataloging in the Online Environment”, where she said the following:

“... the introduction of online catalogs into libraries opens up impressive new possibilities of power of retrieval and ease of use for ourselves and our clients. Our task now is to design the intellectual content and arrangement of catalogs so as to take maximum advantage of these new technical capabilities.”

And I agree with her. And in fact I started my career at about the same time that article was written, when libraries began the transition from card catalogues to online catalogues. My expectations at the time assumed we would expose the rich syndetic, hierarchical structures that form part of LCSH. These are important for the consistent application and interpretation of the subject headings during the subject analysis process and they can be equally important when connecting a user’s information needs with the information strategies needed to discover relevant resources in the catalogue.
However, unless you’re a cataloguer you will probably rarely, if ever, see anything like this in your library catalogue. This subject authority record tells you for example that ‘Capital punishment’ and not ‘Death penalty’ is the term you should use when searching for resources on this topic. This is also where you get to see any related or narrower terms that can inform your search strategy. And, as we saw in the earlier examples, uncover any evidence of bias that you might need to take into account.
In another great collection of writings by Sanford Berman from 1981, he wrote about the potential benefits of automation:

“With automation ... we can now use a subject approach ... that allows the naive user, unconscious of and uninterested in the complexities of synonyms and vocabulary control, to blunder on to desired subjects, to be guided, without realizing it, by a redundant but carefully controlled subject access system.”

Exactly! But it didn’t seem to work out that way. This should have long ago been integrated and self evident in our library catalogues. It’s not, but it should be.
A more contemporary view, and something I’ve been thinking about, is that maybe linked data could provide a solution. Sandy and Bossaller also considered this possibility when they said:

“The rigidity of information systems, which was necessary in the card catalog and even in electronic surrogates for the card catalog, could be reconsidered in light of both the recognized needs for cultural autonomy for indigenous people and the flexibility that is granted by newer web technologies, such as linked data.”
In a way this is similar to the research I’m currently doing using the KF Modified classification scheme as linked data. Only I’m thinking in terms of using the classification as a way to connect and enhance information about the resources we have in our library collection or in our institutional repository.

But can we use linked data technology to somehow correct or adjust headings in a subject classification list? I’m not sure.
I come back again and again to this thought from David Weinberger’s great book, “Everything is Miscellaneous.”

“What you really want is a [classification] tree that arranges itself according to your way of thinking, letting you sort first by expertise and then by experience, and then tomorrow lets you just as easily sort first by language and then by cost, location, and expertise. You want a faceted classification system that dynamically constructs a browsable, branching tree that exactly meets your immediate needs.”

And although I don’t think he was necessarily talking about linked data when he wrote this, this idea certainly fed my original interest in that technology. And it seems to me that this is the way you would want a classification system to behave if you had the choice.
Maybe we could create some sort of an ontology that takes existing LCSH terminology and maps it on to a fuller, more complete ontology. Something that attempts to present an unbiased list of terms according to a combination of ‘user warrant’ based on the language of the end-user and ‘structural warrant’ to fill in the gaps left by ‘literary warrant’?

But how would this work exactly? And who would maintain it?
Another possibility might be to use an attribute similar to the way that multilingual labels can be managed using SKOS in a linked data application. Here, if you were to access the information from a German domain, your application would know to look for the ‘@de’ language attribute and display the second label for the German name of this collection.

Can we create a similar ‘user warrant’ attribute?
I can’t offer any definitive answers today but I’d be very interested to hear any thoughts you might have on these or other possibilities.

Thank you for your attention!