

Scenario 1

Comments received from: 1) Dianne McCutcheon
2) Adolfo Tarango
3) John Riemer

1) I had trouble determining the underlying issue(s) behind this scenario. As I read this scenario, I asked myself is this a metadata issue or a collection development issue or both? Are we expecting that Google is acquiring the books themselves, so no library catalog data would exist? I don't see this as likely. But if it turned out that way and libraries chose to subscribe to the Google collection instead of acquiring material themselves, then we would have to count on Google to provide metadata for the materials. If Google continues to digitize materials from some library's collection, then the library would still need to create a bibliographic record for its own records. ~Dianne McCutcheon~

2) Recently, members of my staff have been involved in a project to confirm whether or not Google has correctly rejected items for digitizing. As you know, Google initially took a digitize whatever we can get approach. Now, however, Google is being more selective about what it will digitize and is taking metadata provide by participating libraries, and comparing it to metadata they have about items they have digitized, rejecting those they consider duplicates. While not party to the full analysis of the results of the project, my staff did report to me that Google was rejecting items as duplicates that were not duplicates, or, the Google metadata was insufficient to make a correct assessment. In other words, the Google metadata seems to be of insufficient quality to support even a simple process like correctly identifying duplicates. If Google can't correctly identify materials, doesn't it follow that users will be poorly served by that same metadata which seems to be supported by such news stories as this one on the Google settlement:

"[Geoff] Nunberg doesn't have a problem with the books themselves; they've been accurately scanned and present the text verbatim. His problem lies with the metadata, or information Google publishes about the books: the authors, the publisher, the date of publication, etc. In case after dismaying case, Nunberg found that Google had critically misreported key information about the books they offer users around the world. Tom Wolfe's novel *The Bonfire of the Vanities*, for example, is listed as having been published in 1888, and Raymond Chandler's novel *Killer in the Rain* is listed as hitting the shelves in 1899. Over and over, Google got important details wrong, potentially misleading future researchers and bookworms around the world. [East Bay Express, Oct. 14, 2009: <http://www.eastbayexpress.com/ebx/PrintFriendly?oid=1211860>]"

It therefore would appear that PCC/PCC members could have a very important role to play in providing quality metadata or enhancing metadata already provided to Google. The issue to me isn't whether we have a role, but whether we should and the mechanism, financial and technological, to support doing so. Thinking about the should part, briefly, I'm inclined to believe that we need to

Responses to Scenarios PCC Policy Committee Meeting Nov. 5, 2009

support to the extent possible, user use of Google scanned materials, which to me means doing whatever we can to improve the metadata Google has for items. Regarding the mechanisms, financially, might PCC/PCC members be able to be compensated by Google for such metadata enhancement. Technologically, could the metadata be pulled from OCLC. As we continue to move toward an overall network approach to metadata creation, with ongoing enrichment of records, would it not make more sense to enrich a single database and create mechanism to tap into that database to support external operations. ~Adolfo Tarango~

3) The first impulse might be to ask "With full text available online, who needs metadata?" Summary metadata remains likely to help users determine if they wish to pursue using a resource, rank it in relevance-based search result listings, and to support bibliographic citation management software. Without summary level metadata, every word in a document would constitute a coequal keyword. The need to differentiate among similarly named individuals is likely to become important in a new milieu-making sure the correct author receives his/her royalties. Authority records will likely be mined to support recommender services, such as "more like this" and "did you mean ...?" Google's behavior toward previously-created library metadata is anything but disdainful at the moment. ~John Riemer~

Scenario 2

Comments received from: 1) Dianne McCutcheon
2) Adolfo Tarango
3) John Riemer

1) The trends described in this scenario may reduce the need for individual libraries to create/maintain serial cataloging records. If a library is using pay-per-view, they won't need serial records in the catalog - they have no checkin or journal titles to track anymore and users can find the articles through Google, an A&I database, or a publisher's website and get access via a link resolver to the full-text.

But I don't see the "journal" as an entity (a "package" of articles) going away that quickly. Even if libraries cancel titles and begin using "pay-per-view" access, some type of serial bibliographic control will still be needed by A&I services and publishers. I foresee that ISSN will still be used for linkages, so at least some basic serial bibliographic record will be needed which records that information for use in databases and with link resolvers. The question is who (and how many of us) will be needed to create these records? ~Dianne McCutcheon~

2) I think we need to research this more thoroughly to come up with the appropriate response. For example, it must be at least three years ago when I attended a NASIG presentation where the issue of the long-term future of the "journal" as entity was raised with a panel of several publishers. I recall that every one of them responded that while it was true that articles were what many users are looking for, they all believe that the "package" of the journal was still very important, especially as

a branding mechanism. For example, being published in the New England Journal of Medicine was still very important and that user would continue to look for the New England Journal of Medicine as an entity. Also, and strongly related to the branding, is that many users still want to browse journals. If I recall correctly, this supported "serendipity" discovery. You can't search for something if you don't know it exists. Browsing through a journal could turn up some very important articles related to your research or stimulate new directions for research. Like I said, we need research to show whether this still is important, if not, then yes, we may be spending too much time describing entities no longer of practical usefulness. ~Adolfo Tarango~

3) Collecting and usage patterns may be changing for serial content. With the shift to an online publishing environment, it is not economically imperative to collect multiple articles before publishing an issue; it is possible to offer and sell serial content at the article level. On the opposite end of the spectrum are aggregations of serial content, either based on articles or by publisher; serial content is also being licensed and usage statistics tracked at that level. The emerging interest in and focus on serial content at new levels, both more and less granular than the serial titles featured in traditional cataloging, does not necessarily diminish the importance of that level. A serial title is a form of identity or context for an article. It is a key part of a citation to an article, as well as a connotation of the type and degree of editorial review a manuscript has undergone. Researchers and their colleagues need to know if the repository document they have read is a preliminary version, a vetted and published version, or a post-publication version. The relationships traditionally tracked in continuing resources cataloging and the identifiers that reflect them will continue to hold importance. What has changed is that some new relationships will also become important. More collaboration among a wider circle of interested parties will be required to make the bibliographic control we need come together. ~John Riemer~

Scenario 3

Comments received from:

- 1) David Miller
- 2) Phil Schreur
- 3) Dianne McCutcheon
- 4) Adolfo Tarango
- 5) John Riemer
- 6) Robert O. Ellett

1) My primary response to this scenario is that we'll best be able to face it if we keep a focus on values, as compared with job descriptions. I respect the MLS degree as much as anyone as (and more than many, no doubt), but I do think we've been hampered by relying on metaphors like "handed down" and "lower levels." What are our preferred outcomes? What do we value? We'll be able to deal with reorganization of types of work among types of workers - including the questions of adopting new roles and discarding old ones - if we focus more on medium- to long-term vision and less on "deprofessionalization" anxieties. That may not be a very specific response, but truly, re-asking the old questions will just bring back the old answers. ~David Miller~

Responses to Scenarios PCC Policy Committee Meeting Nov. 5, 2009

2) The scenario is a depressing one if we attempt to stay within our established paradigm. There is an ever shrinking pool of experienced catalogers and a rapidly expanding pool of material we are expected to control. The proportion of cataloger to items continues to decrease. We should turn the attention of these skilled people away from the creation of individual bibliographic records to issues of authority control and controlled term access at the database level. ~Phil Schreur~

3) Scenario 3 and Scenario 4? My thoughts don't exactly correspond to these scenarios, but they are really about where the cataloging profession might go in a digital world. Can the cataloging community reach out beyond itself and its professional culture of deliberation to ally with the creators of the technological advances to improve bibliographic control? The PCC'S strategic directions 1-4 actually point to this idea, but much of our efforts have been within our own community, in a traditional cataloging/MARC-based environment. Can we find ways to reach out to publishers/authors to create new alliances? For example, there is increasing interest in author disambiguation within the scientific community. What if the "authority" records for authors could be created up front by publishers/authors with author ids that could be used instead of librarians spending huge amounts of time after the fact creating authority records? How might the PCC share its wealth of authority information already developed into something useable at a web-scale? I'm sure we can find all kinds of problems with this particular idea, but it is the kind of collaboration I feel we should be thinking about. ~Dianne McCutcheon~

4) We maintain standards by adhering to them and not allow substandard work to enter the system; I don't believe "deprofessionalization" is really a factor, but rather the issue is an institution's commitment to train staff appropriately (I think there is the issue of compensating them appropriately as well, but won't comment on that part here). At our institution, I would estimate no less than 90% of our cataloging is done by non-librarians, all of it meets standards. That doesn't mean full-level cataloging, it means records coded to reflect the applicable standard, e.g. "K" level record. We have added thousands of records to OCLC using batch processing, but have reviewed and coded them to reflect the appropriate standard they met. We invest a significant amount of time in training staff for the level of records we expect them to create. We don't have staff create records they have not been trained to create. I see PCC having an ongoing role in defining/creating appropriate standards (e.g. CSR, BSR) and providing training. ~Adolfo Tarango~

5) Declining numbers of staffing resources are but the latest reason why we need to de-duplicate the work we are doing across institutions and to begin in earnest to work in shared files. All the work we have done over the years in PCC to build consensus on and develop national standards has paved the way for this. Even in days of peak amounts of library staffing, we were not really accomplishing all that needed doing, working as single institutions or in small groups of peers. In order to see the level of bibliographic control we are used to extend to a maximum amount of material, there are some changes we need to make.

"We have to be willing to cooperate in using each other's records (largely) as-is, to agree on standards, to trust each other. Libraries should not always feel that they have to be the ones to provide the metadata. If there is similarity in the uses made of bibliographic descriptions by

Responses to Scenarios PCC Policy Committee Meeting Nov. 5, 2009

publishers, vendors, and libraries for their inventories, and if there is significant overlap among needed data elements, then it makes sense to pursue a single metadata creation effort." [Quoted from guest editorial in LRTS 51(1) Jan. 2007] We need to adopt a vision of "the one good record." The recent OCLC project led by Renee Register entitled "ONIX to MARC and Back Again" [sample presentation, from ALA Chicago <http://vidego.multicastmedia.com/player.php?p=mx2o79t1>] is an example of the type and scale of changes we need to make in creating metadata for mainstream publications, so that we can free up our remaining resources to address newer types of and more unique resources.

In the early days of PCC's formation about 15 years ago, there were grand ambitions to expand BIBCO and CONSER to large numbers of institutions, and to achieve correspondingly large quantities of records. In reality those programs have held steady at around 4 dozen members each for a number of years. Is that a desirable state of affairs? In conjunction with OCLC's Expert Community Experiment earlier this year, I heard disturbing reports of comments at the OpCo meetings such as, "What's the point of having a PCC if just about anybody can edit the records?" Do we need to challenge such thinking and/or reexamine what it means to be in PCC in the future? ~John Riemer~

6) Scenario 3 assumes (I believe incorrectly) that the use of support staff will decrease standardization of cataloging. I do believe that the majority of descriptive cataloging is transitioning to paraprofessionals, but the majority of subject cataloging is still handled by professional staff. I think we need to position ourselves for this separation of duties and emphasize it is not a de-professionalization of duties. ~Robert O. Ellett~

Scenario 4

Comments received from: 1) Phil Schreur
2) Adolfo Tarango
3) John Riemer
4) David Miller

1) This is a difficult scenario to discuss. In many ways, the PCC is a conservative group. The guardianship of standards is by nature a conservative process. Changes to them can have a direct impact on workflows and implications for budgets. The PCC is also most known for its responses. The comment I hear most often is, what is the PCC response to RDA, to the Working Group on the Future of Bibliographic Control, etc. To shift towards a more proactive mode implies action. To be seen as a leader, we would need to not only discuss emerging technologies (e.g., authority control via entity extraction) but sponsor or collaboratively develop prototypes. The shift is a dramatic one and can not be done halfheartedly.

I almost deleted my response to this scenario after reading the Strategic Directions document and the goals for SD1; however, I thought my reaction was a telling one. ~Phil Schreur~

Responses to Scenarios PCC Policy Committee Meeting Nov. 5, 2009

2) On the one hand, it seems that in an environment of quickly changing technological advances that a slow deliberative process is actually a good thing. What is the point in quickly implementing a new process dependent on one technology when in a couple of months that technology becomes obsolete. But perhaps technology is a red herring issue. More it seems that PCC should develop a technology independent strategy, outlook. Technology is after all, only the tool we use to accomplish our goals. Should not PCC set a goal of say, creating CONSER authenticate records for all online journals to "X" standard, and let the various PCC members and OCLC pick and choose whatever technological tools they best can use to accomplish the task, allowing them to drop one technology for another as appropriate. On the other hand, perhaps the point of this question is to begin to push out to OCLC, publishers, PAMs, ILS vendors, etc., ideas for how they could improve data mining, data manipulation that would allow us to better augment existing records, or repurpose data. If serving as a "clearinghouse" for channeling technology improvements to OCLC, et al, perhaps that could be a role for PCC. ~Adolfo Tarango~

3) If we take the time to gain community-wide consensus on one more thing, probably it should be that we will not have the luxury of lengthy "deliberation and consensus decision-making" and that we will experiment with faster alternatives. As part of the University of California's collaboration with OCLC to build the next generation Melvyl based on WorldCat Local, a number of small groups have been charged with rapid turnaround time expectations to accomplish certain tasks. Relevant boilerplate statements found in the charge documents at http://libraries.universityofcalifornia.edu/about/uc_oclc.html#task include "... given the tight schedule and UC's wish to embrace the concept of rapid development, we need to develop an organizational structure that is capable of working in parallel and providing the above teams with rapid analysis and feedback in an iterative manner." and "Your team may call on others for help, as needed, provided that the above caveats about rapid turnaround time and flexibility are kept in mind. " In PCC we ought to deliberately experiment with task groups working on something on a short timeframe that falls entirely between OpCo and ALA meetings. Speaking of the "technological advances" mentioned in this Scenario, we do live in an era of collaboration software like Elluminate that actually could be used by a small team seeking a bit of input from a larger community while adhering to a short turnaround time. ~John Riemer~

4) Can we differentiate among standards-development scenarios which actually require "lengthy" processes, and those which are lengthy by default (i.e. because that's the way we're used to functioning)? It would be helpful, I think, if we could get out of the slow/fast dichotomy when we discuss this topic, look at different kinds of outcomes which truly require different time frames, and understand why that is the case. For example, it may be my lack of imagination, but I can't picture any "quick" process for developing the new International Cataloguing Principles - that length of time seemed perfectly appropriate if not essential. By contrast, how would we describe the length of time involved in developing the BIBCO Standard Record? Was it quick, lengthy, slowly deliberative, flexibly responsive, something in between? I agree that our culture "has a bias for lengthy deliberation", but I also wonder to what extent this is a story we continue to tell ourselves. Are there exceptions, and what do those reveal? ~David Miller~

Scenario 5

Comments received from: 1) David Miller
2) Phil Schreur
3) Robert O. Ellett
4) John Riemer

1) First I go to the final question, "Are there roles for the PCC in the decision making process involved in implementing these next generation catalogs?" This would take the PCC beyond the work of developing standards and training, which have been its main points of focus. Does the PCC want to get into issues of user interface design, metadata architecture, etc.? If so, it might be very useful for the PCC to help speed the dissemination of open-source tools that are truly as close to plug-and-play state as possible. Here I can speak to my own situation, which is probably similar to those of many libraries of medium to small size. We are longing to do much more experimentation in this area, but we absolutely need more or less off-the-shelf tools. I realize that a lot lies behind that statement. As a side note, I'd like us again to eschew binary thinking - there's no reason why a good interface can't flexibly incorporate both the Google-like search box and the so-called "traditional" index searches. I have no issue with making the former widely available. I do have an issue with the apparent fixation on only providing that. ~David Miller~

2) Currently at Stanford, we are implementing an instance of Blacklight that we call Searchworks (Searchworks.stanford.edu). As an open source discovery environment, we are truly able to respond to user feedback quickly and experiment with alternative access strategies in a rapid, iterative development process. Features like relevancy ranking and faceting can be tailored to our environment, or even by discipline to provide special views of particular data.

The PCC can have impacts on this development in two main areas. First, our trained PCC catalogers have become the metadata experts for the development of this discovery environment. As the acknowledged discovery metadata experts, their understanding of MARC and other metadata standards is imperative to technical development. However, they must learn to apply the broader principles they are trained in new ways to be affective.

The second sphere is a potential area of development. These new discovery environments will include a mixture of metadata following differing standards. Should the PCC expand its focus to look beyond the creation of individual bibliographic records to principles and methods of controlled access in heterogeneous data environments? Faceting is one of the most popular features of these new environments and can be made ineffective by uncontrolled data points. What can we do to improve access? ~Phil Schreur~

3) Scenario 5 supports the role of mash-ups in the library catalog environment. Open source environment often appear more flexible, but they increasingly require more technical support. The PCC has made quality and standardization tenets of value. The PCC if it forays in the open source environments should maintain their promotion of these values. ~Robert O. Ellett~

4) Several reactions:

A) "The downside of these new and improved public access interfaces is that they encourage us to continue thinking about how to improve local bibliographic databases, rather than thinking about how to manage without them." [This is a line in a manuscript that I anonymously reviewed for Rebecca Mugridge for an upcoming CCQ theme issue on cooperative cataloging had a powerful resonance for me.] The same author bemoans the lack of true record sharing that long ago should have taken the place of our practice of record copying. Look how long we were held back from being able to close out death dates in person name headings, as users logically would expect us to! I say the real reason was fear of imposing bib file maintenance workloads on libraries that hold all those copies of records; it is trivial to make such changes in an environment where there is but a single copy of the records we all use.

B) "An intuitive interface is not by definition "dumbed down" or anti-scholarly" [from a Powerpoint presentation "Pathway to the Future: Library Bibliographic Services for the 21st Century" made about the BSTF report at Living the Future 6 conference, Apr. 7, 2006, Tucson]

C) Of course, in any effort to get more mileage out of metadata, the expertise and imagination of PCC catalogers should be called on. ~John Riemer~

Scenario 6

New Scenario from John Riemer

In the "Collections Grid" (<http://www.oclc.org/reports/escan/appendices/collectiongrid.htm>) that is a prominent feature in many OCLC presentations, it is often observed that libraries spend a very high percentage of their metadata energies and acquisitions dollars on the upper left quadrant, i.e. on published non-unique materials. Libraries need to "spread out" these energies to effectively cover the other three quarters of the "information space" that our users want to be able to search across.

A lot of the developments discussed above in other scenarios are geared toward or pave the way for catalogers to tackle more unique materials, to take on newer roles and new formats.

It has always struck me that PCC and its standing committees have a key role to play (A) in helping libraries establish strategies and best practices for organizing digital objects and (B) in helping catalogers develop excellent expertise and leadership in the metadata realm to match the quality of what they have historically provided in the traditional cataloging arena.

Shouldn't libraries and cataloging units seeking to take on these new roles be able to find a lot of help from the PCC? (Taking the Cataloging for the 21st Century series of courses is only a good

first step.) Where is the equivalent to the "core level" record for Dublin Core? What should a cataloging unit be thinking about when they get that call for advice on data element usage for a digital library project? What type of authority work is it practical to do for a project? How do you go about selecting a controlled vocabulary to use? How do we strategize to do things algorithmically, to improve metadata in batch actions versus one record at a time? What practical measures are there for freeing up staff resources to take on this kind of work, without getting behind on other responsibilities? Can PCC become a clearinghouse of ideas and advice for these types of things? ~John Riemer~

Scenario 7

New Scenario from Phil Schreur

New Scenario: Cataloging of Monographs is no longer effective. As with journals and the article economy, most users are looking for information in a monograph at the chapter level or below. To focus on the description of the monograph as a whole and to provide broad, although controlled subject access to the item as a whole, does not serve these users well. They will continue to migrate to other discovery environments that will allow them to find the information they seek. The PCC should allow publishers and vendors to become the main creators of descriptive records for monographs as a whole. The PCC should focus on principles and strategies for providing controlled access of names and subjects at the chapter level and below. The effort should be twofold, the first to provide methods through entity extraction and semantic analysis of extracting names and topics from digital items. And the second to develop and provide authoritative forms of names and easily usable taxonomic subject trees for terms that any metadata creator would find attractive to use. ~Phil Schreur~